Lithuanian discourse markers na and nu ‘well’: a glimpse at parallel corpus data

Audronė Šolienė
Department of English Philology
Vilnius University
Universiteto st. 5
LT-01513 Vilnius, Lithuania
Email: audrone.soliene@gmail.com

Abstract

The focus of the present corpus-based study is on the quantitative and qualitative distribution of the Lithuanian discourse markers na and nu ‘well’. The aim of the paper is to establish the translational correspondences of na and nu in English and to reveal their functional versatility in fiction and spoken discourse. The research method is a quantitative and qualitative contrastive analysis based on the data extracted from a self-compiled bidirectional parallel corpus of fiction texts and the spoken sub-corpus of the Corpus of the Contemporary Lithuanian Language. The quantitative results show that nu, due to its greater degree of informality than na, is extremely frequent in spoken discourse. The qualitative findings witness an array of different functions the two discourse markers can perform both in responsive and non-responsive environments as well as inner dialogue.

Key words: discourse marker(s), multifunctionality, corpus-based analysis, frequency, contrastive analysis, translational correspondence

1. Introduction

Research on discourse markers (DMs) has been substantially increasing over the past few decades and resulted in a considerable body of studies (see Fraser 1999; Schiffrin 2008; Urgelles-Coll 2010; Amador-Moreno et al. 2015; Auer & Maschler 2016; Brinton 2017; Fedriani & Sansó 2017, inter alia). DMs have also achieved a great deal of scholarly attention in a contrastive perspective (Aijmer & Simon-Vandenbergen 2003; Lewis 2006; Johansson 2007; Degand 2009; Beeching & Detges 2014; Furkó 2014). In Lithuanian, DMs (sometimes referred to as ‘discourse particles’) have been sporadically analysed in terms of their functional classes (Ambrazas 2006b), lexical sources and categorial status (Holvoet & Pajėdienė 2005) as well as diachronic development (Ambrazas 2006a; Nau & Ostrowski 2010). However, contrastive corpus-based Lithuanian-English studies of DMs drawing on large-scale empirical data from parallel corpora are rather innovative and rare.

This cross-linguistic corpus-based study sets out to describe the quantitative and qualitative distribution of the Lithuanian DMs na and nu ‘well’, to determine their translational correspondences (TCs) in English, as well as to reveal their functional diversity in responsive and non-responsive contexts as well as narratives and inner monologues, e.g.:

– Na matai, – kiek atlyžo Vaitkus.
EN-trans: "Well, all right. You have told me that. Twice."
"Well, you see," Vaitkus softened a bit.

(2) LT-orig: – Ukrainon einu, meilut, Ukrainon, – atsako ji jam švelniai.
– Nu, nu, prasmek... – nusikeikė riebiai ukrainietis ir nusisuko į šalį, savo sąmoji patenkintas.
EN-trans: "I’m going to the Ukraine, dearest, to the Ukraine," she tenderly tells him.
"Sure, sure, now fade away..." growls the Ukranian. He struts off, pleased with his wit.

1 All the examples provided in the paper are taken from ParaCorpEN→LT→EN.
2. Discourse markers in Lithuanian

Though discourse markers are notorious for their elusive nature and a number of definitions given in the literature, this paper utilises the operational definition proposed by Crible (2017), which combines two groups of criteria for an item to qualify as a DM – syntactic (integration and scope) and pragmatic (multifunctionality):

DMs are a grammatically heterogeneous, multifunctional type of pragmatic markers, hence constraining the inferential mechanisms of interpretation. Their specificity as part of the PM category is to function on a metadiscursive level as procedural cues to situate the host unit in a co-built representation of on-going discourse. (Crible 2017: 106)

In Lithuanian, which is a language that still has many uninvestigated linguistic issues, studies on DMs are scarce. In Lithuanian linguistics, the term ‘discourse maker’ is used only in several studies (Masaitienė 2003; Bielinskienė 2009; Smetona & Usioniene 2012). Most tokens that seem to qualify as DMs have been investigated as parenthetical phenomena or as sentence connectives. Few attempts have been made to identify Lithuanian linguistic expressions that could possibly be ascribed to the category of discourse markers, to compile an exhaustive list of the inventory or to analyze DMs in a cross-linguistic perspective.

This paper seeks to contribute to the cross-linguistic description of the family of discourse markers na and nå (see Auer & Maschler 2016), their Lithuanian cognates being na and nu. Auer and Maschler (2016: 2) claim that “[t]hey exist in all modern Germanic and almost all Slavic languages (cf. English now, German nu(n)/na, Dutch nu/nou, Norwegian nå, Danish and Swedish nå, nu, Icelandic nú, Yiddish nu, etc.; Serbian, Croatian, Slovenian, Czech, Slovak, Serbian, Polish, Ukrainian, Belarusian, Russian nu/no), but by no means do they have the same meaning in each of these languages”. The Lithuanian na and nå seem to have escaped the description, but for Sawicki’s (2012) pioneering paper on responsive DMs in Lithuanian focussing on the turn-opening particles na ‘well’ and kąd ‘but’; however, her research is not based on corpus-derived data but on the analysis of one novel; this by no means diminishes the linguistic insights of the paper.

The etymology of the Lithuanian DMs na and nå needs a more thorough investigation, but it seems that the origin of nå is explained through its ties to the adverbs nū and nūnai ‘now/today’, still used in present-day Lithuanian, though, according to the corpus data, non-existent in spoken Lithuanian but present in non-fiction texts. So there seems to be a relation between the temporal usage (or even resultative) of the items under study and their later development into DMs (Auer & Maschler 2016: 6). Smoczynski (2007: 429) takes up this line of explanation and derives nå from the verb nū ‘Polish teraz = now’ as in nū diena or nūdien ‘Polish dotąd, do teraz, do dzisiaj = till this time, till now, till today’. Some scholars tend to link the DM nå to the pronominal stem *eno- ‘that one’ (Auer & Maschler 2016: 7). A similar view as regards the origin of the Lithuanian na is expressed in Fraenkel (1962: 477). In his dictionary he gives an entry for na presenting it as an interjection and relating it to the demonstrative *ônû-, *ênû- or anas ‘German jener = that’ or an(a) ‘German siehe da = look there’. Though Auer and Maschler (2016: 7) maintain that it is rather unlikely that this interpretation may gain ground, since it might be the case that na is “a plausible outcome of a grammaticalization chain which starts from a temporal adverb NU” and that it split off from nå as a result of the lowering of the back vowel /u/, more investigation is needed.

Another thing to consider is the multifunctionality of the DMs in question. Usually they occur in conversational situations in responsive or non-responsive environments. They may refer to the previous turns of the interlocutors, elucidate or prompt upcoming turns, or serve as cohesive devices within longer passages.
produced by the same speaker. To sum up, they may perform the following more fine-grained functions (based on Sawicki (2012, 2016); Auer & Maschler (2016: 11-35)):

a. acknowledgement of the content of the previous turn;
b. affirmative response to a question;
c. dialogue opener or a topic-change device;
d. (reluctant) agreement;
e. impatience to an utterance or an action;
f. hesitant reply to a question;
g. expressing surprise, amazement or mirativity;
h. self-urgency/correction;
i. reinforcing commands;
j. expressing threat, indifference and other emotions in the affect dimension;
k. cohesive device used in narratives.

Moreover, *na* and *nu* can also be found in passages of inner monologue. But even in these cases the colloquial tenor remains (Sawicki 2012: 172). By no means are these functions distinct on their own; there is always some degree of the overlap, e.g., between urging the interlocutor to add information and a note of impatience or between reinforcing a command and impatience. These functions are discussed in greater detail in Section 4.2.2.

3. Data and methods

The corpus-based approach adopted in this study helps to reveal patterns and meanings of DMs which would be difficult to pin down by studying introspective data. The research method is a quantitative and qualitative contrastive analysis based on the data extracted from a self-compiled bidirectional parallel corpus – ParaCorp<sub>EN→LT→EN</sub> (Soliéné 2013). The corpus is designed following the model of the English-Norwegian Parallel Corpus (Johansson 2007). The ParaCorp<sub>EN→LT→EN</sub> was compiled from original English fiction texts and their translations into Lithuanian and original Lithuanian fiction texts and their translations into English. The advantage of such a corpus design is that it allows different directions of comparison and can serve both as a parallel corpus and a comparable corpus (Johansson 2007: 11). The size of the corpus is about 5M words (see Table 1):

<table>
<thead>
<tr>
<th>Corpus</th>
<th>Original</th>
<th>Translation</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>ParaCorp&lt;sub&gt;EN→LT&lt;/sub&gt;</td>
<td>1,983,266</td>
<td>1,541,038</td>
<td>3,524,304</td>
</tr>
<tr>
<td>ParaCorp&lt;sub&gt;LT→EN&lt;/sub&gt;</td>
<td>608,426</td>
<td>788,897</td>
<td>1,397,323</td>
</tr>
</tbody>
</table>

Table 1. Size of the two sub-corpora ParaCorp<sub>EN→LT</sub> and ParaCorp<sub>LT→EN</sub>

A reference has also been made to the Corpus of the Contemporary Lithuanian Language (CCLL) (http://donelaitis.vdu.lt), namely the sub-corpus of spoken register (447, 396 tokens).

Frequencies of particular patterns are of paramount importance to this paper, since frequency may be an important factor in specification of meaning. Since the sub-corpora are of different size, the raw frequency numbers have been normalized per 10,000 words. Moreover, in order to verify whether the similarities and differences in frequency are statistically significant, I have also performed the log-likelihood (LL) test, which is commonly considered to be a more statistically reliable tool than the chi-square test. The higher the LL test value, the more significant is the difference between two frequency scores. A difference in frequency is considered to be statistically significant if the LL test value is 3.84 or higher at the level of p < 0.05.

4. Findings

This section presents the quantitative and qualitative findings of the analysis performed. It starts with the analysis of the quantitative distribution of *na* and *nu* in fiction and spoken registers (Subsection 4.1). Subsection 4.2 deals with a qualitative view of the functional diversity of the DMs under study and their TCs.
4.1. A quantitative view of na and nu in fiction and spoken registers

The first step in the analysis was to look at the frequencies and distribution of the two Lithuanian DMs in fiction and spoken registers. Table 2 presents the quantitative findings in CCLL (spoken sub-corpus) and ParaCorp\textsubscript{EN→LT→EN} (fiction).

<table>
<thead>
<tr>
<th></th>
<th>na</th>
<th>nu</th>
</tr>
</thead>
<tbody>
<tr>
<td>raw</td>
<td>f/10,000</td>
<td>raw</td>
</tr>
<tr>
<td>Fiction</td>
<td>212</td>
<td>10</td>
</tr>
<tr>
<td>Spoken</td>
<td>878</td>
<td>7,986</td>
</tr>
<tr>
<td>Total</td>
<td>1,090</td>
<td>7,990</td>
</tr>
</tbody>
</table>

Table 2. Frequencies of na and nu in CCLL and ParaCorp\textsubscript{EN→LT→EN}

The quantitative results show that nu is strikingly frequent in spoken Lithuanian – \( f = 178.5 \) (f per 10,000 words), in contrast to only 0.16 in fiction. The reason for a rather low frequency of nu in fiction must be the fact that nu is considered to be sub-standard in Lithuanian and fiction texts are normally edited. The normalized frequency of na is 3.48 in fiction and 19.62 in spoken register, respectively. Both DMs are scarcely used in fiction; their prototypical use is in spoken language (Brinton 2017: 3-4). The table below gives the log-likelihood test values of na and nu in fiction and spoken discourse.

<table>
<thead>
<tr>
<th></th>
<th>Fiction</th>
<th>Spoken</th>
<th>LL value</th>
</tr>
</thead>
<tbody>
<tr>
<td>raw</td>
<td>212</td>
<td>878</td>
<td>-667.43</td>
</tr>
<tr>
<td>nu</td>
<td>10</td>
<td>7,986</td>
<td>-13571.41</td>
</tr>
</tbody>
</table>

Table 3. The log-likelihood test values of na and nu in CCLL and ParaCorp\textsubscript{EN→LT→EN}

The log-likelihood scores (–667.43 for na and –13571.41 for nu) indicate a statistically significant underuse of na and nu in fiction texts as compared to their frequency in spoken register.

4.2. Functional distribution of na and nu and their translational paradigms

This section gives an overview of the different types of the TCs of na and nu as well as draws parallels between the function performed by the DMs and their TCs.

4.2.1. Overview of the translational profile

Translational paradigms, as was indicated in Aijmer and Simon-Vandenbergen (2003), Johansson (2007), and Usonienė, Šolienė and Šinkūnienė (2015) (just to name a few), are a useful means for the investigation of multifunctional expressions. Aijmer and Simon-Vandenbergen (2006: 5) suggest that employing a parallel corpus allows the establishment of cross-linguistic paradigms by showing how an element in Language A is translated into a variant of Language B. Degand (2009: 174) similarly suggests that translational data allow gaining “insight into the precise meaning of the linguistic items under study”. Thus, the analysis of translational data can contribute to disclosing the functionally-versatile nature of DMs. The analysis of the translations of na and nu (the latter is very infrequent in fiction (n=10), so the TCs of both DMs are taken together, all in all 222 instances) into English exhibited a great range of their translational correspondences, which may be indicative of their multifunctional nature. Table 4 presents the results obtained from the analysis of the TCs of na and nu in translated English:
The analysis of the translational paradigms of the DMs in question has proved that, due to their extreme multifunctionality, non-propositionality, context-dependence and non-referential (interpersonal and textual) function, they exhibit a wide array of different TCs. The DMs na/nu are extremely functionally and translationally versatile: their translational profile comprises as many as 34 different TCs. As can be seen from Table 4, the prototypical translational correspondence of na/nu in English is well (47.3 %). It takes almost a half of all the TCs; however, the other half of the correspondences shows a great versatility in linguistic expression. As is claimed in Aijmer (2007: 34), singleton and infrequent translations may reveal new or emerging developments of a DM.

### 4.2.2. Correlation between translational correspondences and functions

As regards the correlation between the translational correspondences and the functional diversity of the two Lithuanian DMs under study, their prima facie TC well appears with the majority of the functional variants. When na/nu function as acknowledgement markers of the previous turn, they are usually translated as well or so into English e.g.:

(4) **LT-orig:** – Darosi vėlu. Ir galva įsiskaudo...
– Na, tai aš jau eisiu.

**EN-trans:** "It’s getting late. And I have a headache....
"Well, I’m off then."

In such cases, na/nu (usually followed by tai ‘so’) convey justifications, explanations, reactions or reasons pertaining to the information conveyed by the previous speaker.

In addition, besides well, when na/nu are used to encode an affirmative response or agreement, they are translated as ok, okey, aha, yes, yeah, fine, all right, certainly, etc.:
“Ahh...” Nora heaved a sigh of relief.

“It seemed to me that you wanted to say something else... like, like this dress really grew on her..."

"Yeah, that's exactly what I wanted to say..."

Na/nu may invite the interlocutor to add information and thus hasten the flow of discourse. In this case, the range of TCs encompasses the following items: ok, go ahead, well, etc. The urging or prompting a participant in further developing an upcoming action (be it verbal or non-verbal) results in a number of different TCs (e.g.: come on, go on, ok, so, come now, etc.). Moreover, sometimes urging converges with a note of impatience on the speaker’s side, especially when na and nu are used to reinforce commands or requests (Auer & Maschler 2016: 14), e.g.:

Here the teacher assistant’s eagerness to get the right answer is highlighted by the repetition of na; however, her effort is also tinted with a shade of impatience and pressure on the addressees.

When na/nu figure as tokens marking reluctant agreement or a hesitant reply, their prototypical equivalents in English are well, well anyway, well yes, in a way you're right, etc.:

In (7) na renders the answer as somewhat reluctant or it seems that the interlocutor wants to withhold the information or hopes to be implored for a more precise response.

The DMs in question may be employed as dialogue openers, usually in sentences containing direct address forms prompting the interlocutor to enter the conversation, or may indicate a change of topic. Again the English TCs are well, well now and so, e.g.:

In (9) Na renders the answer as somewhat reluctant or it seems that the interlocutor wants to withhold the information or hopes to be implored for a more precise response.

The DMs in question may be employed as dialogue openers, usually in sentences containing direct address forms prompting the interlocutor to enter the conversation, or may indicate a change of topic. Again the English TCs are well, well now and so, e.g.:
“Every trade has its secrets, Miss Nora,” Joseph took the compliment as if it went without saying. "So, where are our seams?"

In sentence (8), the speaker begins a conversation with na and encourages the addressee to discuss things, whereas in (9), using na, the speaker seems to end up the previous topic of the conversation and switches to a new one.

When the DMs na/nu express amazement, surprise and mirativity, they are always accompanied by ir ‘and’ and are rendered into English by exclamatory structures with what, e.g.:

(10) LT-orig: – Nu ir peilis! – žavisi Roza. – Aš jį pasiimsiu, gerai?
– Jokių būdų, – pamokomai dėsto Žilvinas.

EN-trans: "Wow, what a knife," Roza is charmed. "I’ll take it, okay?"
"No way," Žilvinas lectures.

The speaker’s surprise is even reinforced by the interjection wow in the translation.

Besides the listed functions, it seems that the Lithuanian DMs, like Hebrew nu (see Maschler & Dori-Hacohen (2012: 432)), colour the utterances they precede with other different shades of emotion ranging from humorous mockery, indifference, reproach and contempt to even threat (in this case usually combined with a pejorative direct address):

(11) LT-orig: – Na, na, maitos vaikai, nesivaipykite! Greičiau, driskų išvažos!
– putoja padūkęs tirolietis Toni Fabro,
– nėr čia ko!

EN-trans: "Uh-uh, you sons of bitches. Don’t make faces! Why bother? – In a month you’ll be the same! Faster, you dogshit!" foams the wild Tyrolian.

For example, often the speaker’s indifference is conveyed by a sequence na ir kas or nu ir kas meaning ‘it’s not important; I don’t care’, whose English correspondences usually are so what constructions (so what, and what of it, well so what):

(12) LT-orig: – Tą kalendorių padovanojo mano būsimoji žmona.
– Na ir kas, – pasakė mergaitė, vėl įsikniaubdamas knygon <…>.

EN-trans: "That calendar was given to me by my future wife."
"So what," said the girl, burying her head into her book again <…>.

When the DMs in question are used non-responsively, they can be employed in self-repair or correction contexts such as searching for the right word, giving a summary of what was said, explaining or rewording:

(13) LT-orig: – Viena mano pažįstama buvo Paryžiuj, – įsitępia Beta. – Na, žinot, kuriuos vyras dirba CK. Tai pasakijo, kad visos merginos vaikščioja rudai lakuotais nagais.

EN-trans: "A friend of mine was in Paris." Beta interjects, "Well, you know, the one whose husband works at the Central Committee. She said all the girls were walking around with brown nail polish."

Moreover, in longer passages produced by the same speaker the non-responsive na/nu are often used as “non-specific, strictly cohesive intra-textual” devices (Sawicki 2016: 99) introducing narratives of personal experience. They also retain the same function when they are employed in non-dialogic environments, i.e. in passages reproducing inner thought. The author then seems to conduct an inner conversation with himself/herself: elaborating his/her arguments, explaining or adding additional information, retelling events, accounting for consecutive events, e.g.:

(14) LT-orig: Papasakodavo ne įskundimo forma, ne pakenkti kam norėdamas, bet iš neišpasakomos savo naivumo, – na, žinoma ir todėl, kad komendantas buvo jam toks didelis autoritetas, tokia įkūnyta kilybė, kad jis pas jį lyg išpažintę atlikdavo.
EN-trans: He wasn’t giving him all this information as an informer, or out of malice, but simply because he was incredibly naive – and, of course, because in his eyes the Commandant was such a looming figure of authority, an embodiment of such nobility, that visiting the Commandant was for Klawan like going to confession.

The TCs of the non-responsive and non-dialogic uses of *na/nu* are especially versatile: *well, and, yes, despite all, in fact, in other words,* etc.

### 4.2.3. Zero correspondence

Another interesting observation regarding the analysis of the translational paradigms is zero correspondence, or cases of omission. The cases of zero correspondence of *na/nu* in translated English amount to 22.5% of the whole concordance, e.g.:

(15) LT-orig: – Blogai, – sako Konstantinas, – dabar aš mirsiu.
– Na, Konstantinai, kad tu taip neprotingai juokauji!
EN-trans: “It’s bad,” says Konstantins. “Now I’m going to die.”
“Ø Konstantins, don’t be ridiculous!”

Other parallel-corpus based cross-linguistic studies of DMs show similar results. Aijmer and Simon-Vandenbergen’s study (2003: 1153) demonstrated that in the English-Swedish translations *well,* which is a *prima facie* equivalent of *na/nu,* is not translated in 21% of all cases, and in the English-Dutch translations – 7% respectively. Similarly, in the study by Bazzanella and Morra (2000), the omission rate of *well* in the English-Italian texts was as high as 39%. Johansson’s (2007: 289) investigation of *well* and its translations into Norwegian accounted for 16% of zero correspondence.

There might be several reasons why DMs are so frequently not translated. The phenomenon of zero correspondence may speak in favour of the statement made in Aijmer and Altenberg (1996: 32) that the more grammaticalized an item, is the more frequently it is omitted in translation, so zero correspondence may be a result of language-specific conventions or different degree of grammaticalization of DMs. Also, “[e]xpressions that do not contribute to the propositional content are often untranslatable in the sense that an exact equivalent cannot be found in another language” (Aijmer & Altenberg 2001: 32). Another reason can be the notorious multifunctionality of discourse markers and their doubling function, i.e. “cases where the meaning of *well* is supported by other means” (Johansson 2007: 289). Such instances are found in the data of the present study too, e.g.:

(16) LT-orig: – Ak, ką tu, ką tu! Aš juk atėjau čia siūlydama savo meilę...
– O aš spijoviau tau į veidą ir parodžiau duris!
EN-trans: “But, but, I came here, after all, offering my love…”
“'And I spat in your face and showed you the door.’”
“Ø Don’t you see how mean you are? Everyone is like that. Everyone.”

Here we have two DMs at work – *na* and *mатаi* ‘you see’, and the meaning of the intended reproach is successfully conveyed in the translation by the negative imperative *don’t you see,* so *na* can be dispensed with in the translation.

### 5. Conclusions

The quantitative findings demonstrate that *nu,* due to its greater degree of informality than *na,* is overwhelmingly prevalent in spoken discourse. The reason for a rather low frequency of *nu* in fiction could be the fact that *nu* is considered to be a feature of colloquial Lithuanian and fiction texts are usually proofread. Both DMs are less frequent in fiction; their prototypical use is in oral discourse. The qualitative results reveal an array of different functions the two discourse markers can perform both in dialogic responsive and non-responsive environments as well as non-dialogic contexts. Their primary functions
encompass responding to a previous turn or facilitating cohesion and the flow of discourse in conversations and narratives.

The functional & semantic potential of the DMs can be fully reflected by their TCs. Though the prototypical correspondence of *na* and *nu* in their functional repertoire is the English *well*, there is a correlation between different functions performed by the DMs in question and their other various TCs. In addition, zero correspondence of the DMs is indicative of their extreme multifunctionality, non-propositionality, context-dependence and non-referential (interpersonal and textual) function.

**Acknowledgements**

Thanks are due to the Research Council of Lithuania, which funded this research within the framework of project No S-MIP-17-44 (*Discourse markers in Lithuanian: A synchronic and diachronic study*).

**List of abbreviations**

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>DM(s)</td>
<td>discourse marker(s)</td>
</tr>
<tr>
<td>EN</td>
<td>English</td>
</tr>
<tr>
<td>f</td>
<td>normalized frequency</td>
</tr>
<tr>
<td>LL</td>
<td>log likelihood test value</td>
</tr>
<tr>
<td>LT</td>
<td>Lithuanian</td>
</tr>
<tr>
<td>n</td>
<td>number of tokens analysed</td>
</tr>
<tr>
<td>Ø</td>
<td>zero correspondence</td>
</tr>
<tr>
<td>orig</td>
<td>original texts</td>
</tr>
<tr>
<td>TC(s)</td>
<td>translational correspondence(s)</td>
</tr>
<tr>
<td>trans</td>
<td>translated texts</td>
</tr>
</tbody>
</table>

**Data sources**

- CCLL: Corpus of Contemporary Lithuanian Language. Available at [http://tekstynas.vdu.lt/](http://tekstynas.vdu.lt/)
- ParaCorpEN→LT→EN: Bidirectional Parallel Corpus of English and Lithuanian (Šolienė 2013)

**References**


