COGNITIVE STUDIES | ÉTUDES COGNITIVES, 18
Warsaw 2018

Article No.: 1680
DOI: 10.11649/cs.1680

Citation: Czoska, A. (2018). The usage and meaning of *mentioned* (*wspomniany*) and *described* (*opisany*) metatext markers in Polish scientific texts. *Cognitive Studies | Études cognitives*, *2018*(18). https://doi.org/10.11649/cs.1680

AGNIESZKA CZOSKA

Adam Mickiewicz University, Poznań, Poland
agaczoska@gmail.com

THE USAGE AND MEANING OF *MENTIONED* (*WSPOMNIANY*) AND *DESCRIBED* (*OPISANY*) METATEXT MARKERS IN POLISH SCIENTIFIC TEXTS

Abstract

Metatext markers (MMs) are defined as expressions used in a text to inform readers of its structure or relations between its elements (Fraser, 1999; Goldman & Rakestraw, 2000; Schiffrin, 1988). In this paper the usage of two MMs *mentioned* (*wspomniany*) and *described* (*opisany*) in scientific texts is described and the relation between their meaning and usage is analysed. In Study 1, the frequency, scope and direction of the MM is analysed in a corpus of peer-reviewed scientific texts in Polish. In Experiment 2, participants decided whether to use the aforementioned MMs to fill gaps in short scientific texts. The results of both experiments suggest that while the meaning of an MM may influence its usage, this relation may be affected by the size of the whole text.
Keywords: metatext; metatext markers; scientific text; cloze-test; Polish

1 Metatext and metatext markers

In every text there may be various metatextual elements or signalling devices (a broader category covering both lexical and visual elements), that inform the reader of the structure of the text, the hierarchy of the information presented or the relations between pieces of information or text units (Goldman & Rakestraw, 2000; Lemarié, Lorch, Eyrolle, & Virbel, 2008; Schiffrin, 1988). Many signalling devices are realised lexically within a text — these are called discourse markers (DMs) or, when they refer to other elements of the text in which they are used, metatex markers (MMs). MMs are a subcategory of metadiscourse or discourse markers (DMs). These broad terms cover all the lexical means of guiding readers through a message and to manage communication, realised as short phrases or singular words (DMs) or as longer passages (metadiscourse). However, this differentiation is not necessarily used by all the scholars contributing to the topic.

Distinguishing metatext from metadiscourse, Hyland (e. g. 2005) divides discourse markers into two groups: textual (here: MMs) and personal (he also uses the terms *interactive* and *interactional*). In his recent paper (Hyland, 2016) metadiscourse was defined as all the means of

engaging the audience in the text, whether by guiding their interpretation (MMs) or creating "an appropriate relationship with readers" (p. 1475). MMs function as guidelines for the reader, but their usage is affected not only by the possible effect they may have on the reader, but also by the genre of the text and the practices of the community (e. g. Hyland, 2005; Mur-Dueñas, 2011), and by the features of the MMs, for example their lexical form (Lemarié, Lorch, Eyrolle, & Virbel, 2008). The aim of the study and experiment described in this paper was to analyse to what extent the meaning of the acts of describing or mentioning something in a text may influence the usage of the two MMs (later also referred to as metatextual repetition cues, MRCs) *opisany* (*described*) and *wspomniany* (*mentioned*). The term MRC will be used to refer to the cases of using the MMs to signal a repetition (lexical or semantic) but, as will be seen later, *wspomniany* and *opisany* may also be used to announce that something *will be described later*. The results of Study 1 support the hypothesis that this meaning may still affect the usage of the MMs in scientific texts in Polish. The second study, Experiment 1, attempts to ascertain whether the size of the text modifies this effect of meaning. The results of the experiment indicated that the MRCs may be used in very short scientific texts and are also deemed necessary by readers when the whole text consists of only a few sentences and the two occurrences of the repeated word are close to each other.

There are many corpus or case studies on metadiscourse in Slavic languages, including Polish (see for example: Grochowski, 2008; Ożóg, 1991; Wiemer, 2006; Winiarska, 2001). For example, Grochowski (2008) analysed MMs in the form of the superlative of adverbs that expressed certainty or inference: *najpewniej* (*most surely*), *najprawdopodobniej* (*most probably*), *najwidoczniej* (*apparently*), compared their syntactic features to those of the adverb counterparts and concluded that the MMs are grammatical homonyms. However, he does not elaborate on the differences in their procedural meaning or communicative function. Ożóg (1991) analysed MMs containing the semantic component *mówię* (*I say*, *I tell*) in dialogues and polylogues. Among the MMs he chose for analysis were those containing different forms of the verb *mówić* (*as I say*, *to tell you the truth*, *so to say*), as well as those semantically connected to it (such as *mention*, *aforementioned*, *to put it simply*, *to conclude*, *by the way*, *and so on*). He concludes that MMs such as these indicate an informative function (e. g. referential rather than phatic) of the phrase they announce. They may signal its information status as something new, but may also rephrase something already given. Ożóg also mentions that more personal or interactional uses of expressions containing *mówię*, those that refer to the utterances of other polylogue participants, may be more emotionally expressive. Wiemer (2006) analysed examples of evidentiality markers in Polish such as *rzekomo* (*allegedly*), *podobno* (*supposedly*), *jakoby* (*allegedly*, *supposedly*) and other similar markers that express a speaker's doubts and uncertainty about the evidence quoted. He describes the semantic and syntactic characteristics of the DMs, the purpose of their usage and the author's involvement, or the extent to which the author's stance is expressed by the DMs. Last but not least, Winiarska (2001) collected a corpus of television broadcasts from 1977–78 and 1996 that covered 7 hours of dialogues, consisting of discussions and live interviews. She analysed the usage of 17 DMs such as *na przykład* (*for example*), *po pierwsze* (*first*), *czyli* (*thus*, *which means*), *ale* (*but*). The aim of her study was to capture how DMs are used in spoken language to "support the creation of a spoken text showing explicitly (...) the way it is organised"[1] (p. 152) and to describe, on the basis of the data, which DMs are used in this genre to mark changes of topic, rephrasing or defining, and speech acts disambiguating the speaker's communicative intentions.

1.1 The meaning of metatext markers

Common to all the theories describing DMs, or MMs in particular, is the claim that the role of DMs in a text is to guide the audience, facilitating their orientation in the text, foregrounding its connections to other texts or the field, and constructing a relationship between the reader and the author. Discourse markers (DMs), are thus analysed as having procedural meaning (Blakemore,

[1] Originally in Polish: "wspomagają tworzenie tekstu mówionego, ujawniając wprost (...) sposób jego organizacji".

2002; Fraser, 1999; Roulet, 2006). They "function as instructions from speaker to hearer on how to integrate their host unit into a coherent mental representation of the discourse" (Hansen, 1998, after Waltereit, 2006, p. 64). They explicitly present the structure of text information, and indicate its elements (phrases, words) as salient, foreground or background information. However, "Discourse markers are not devoid of semantic content, if by that we understand conventional or coded meaning." (Lewis, 2006, p. 44). DMs may be seen as expressions that have undergone some form of grammaticalisation process or, more generally, have become a part of formulaic language (for a discussion see the papers collected in Fischer, 2006). Because of this, the meaning and role of an MM (or a discourse marker, a DM, in general) may be vague and only loosely connected to the semantics of the terms they are derived from and they may change substantially between different contexts and, from an historical perspective, even lose connection with the "original" meaning (e.g. Gonen, Livnat, & Amir, 2015; Lewis, 2006; Tanghe, 2016).

According to Lemarié et al. (2008) the information function of an MM (the guidelines they provide) seems to be connected most directly to the marker's lexical realisation. MMs may be used, for example, in order to signal the organisation or hierarchy of the text (*first*, *next*), to disambiguate relations between text units (*on the other hand*, *moreover*, *although*) or to explicitly signal the information status of a given text unit (*as mentioned before*). Other features that should be taken into consideration while analysing the role of MM are scope and direction (or localisation) (Ädel, 2006; Lemarié et al., 2008). Both have to be attributed according to the context in which the MMs occur. Scope describes how large, how many or how distant the text units the MM refers to are: from the whole text (global scope) to an adjacent word (local scope). Direction is connected to the information status of the referred unit and describes where it can be found in the text: either it has already been given (backward direction, a review) or it will be presented later (forward direction, a preview). Both scope and direction can only be applied to an MM used in a text. There may be some correlation between scope, direction and realisation, these features are not always independent. However, the correlation may differ from one type of MM to another. For example *in this text* will probably always have a global scope, while *here* may be used to refer to the whole text or to a single paragraph. Lemarié et al. (2008) also claim that "the scope of a signalling device may indirectly influence the accessibility of (some of) its information functions" (2008, p. 44). For example, the reference units of an MM with a local scope are easily accessible, even if the reader ignored them on the first reading. Text units signalled by an MM with a global scope have to be remembered or searched for to be accessed.

1.2 Described (opisany) and mentioned (wspomniany) as examples of metatext markers

This paper focuses on metatextual repetition cues. In Hyland's classification (Hyland, 2005) they are called text connectives (endophoric markers in an earlier version of the classification — Hyland, 1998). Their function in the text is to explicitly refer to relevant text fragments, to name relations between text parts, and to present the structure of the text. The MMs described here are passive participles or passive voice forms of the verbs *describe* and *mention* and as such may be used to refer to notions or elements that are either elaborated on or occur in the current text briefly. Moreover, *opisany* and *wspomniany* have a perfective aspect, like the verbs *opisać* and *wspomnieć*, while *opisywany* and *wspominany* (*opisywać* and *wspominać*) — imperfective. This may further influence their usage. Conversely, the difference between mentioning and describing may be irrelevant for the metatextual usage of the expressions, and thus the MMs may only inform readers that something has appeared (or will appear) in the current text.

Opisany (*described*) and *wspomniany* (*mentioned*) are most frequently used with a backward scope functioning as reviews (Czoska, 2011) and in these cases indicate that a lexeme or a concept already given in the text is to appear again. They may also be used as previews announcing a concept that will appear in the text, but in these cases they appear in a phrase, for example with the verb *zostania* (*will be*). Thus, they belong to the same category of MMs: evidentials that may

be both previews and reviews, but are mostly used as the latter, or as metatextual repetition cues (MRCs). However, as mentioned above, the verbs they are associated with differ in their meaning which may further influence the usage of the MMs.

Opisywany and *wspomniany* should be associated with different communication acts if their lexical meaning is taken into account. This may influence their scope. In this respect, the MMs may be similar to the demonstrative pronouns *this* and *that* (Gray, 2010). If the meaning, here defined as the meaning of the verbs or the difference in the text acts associated with the verbs, of the two MMs evokes expectations concerning the text units referred to, these expectations should influence the usage of the MMs. Moreover, since global MMs refer easily accessible information (such as the topic of the whole text) or to text objects far apart from the MM, the frequency of using global MMs should be lower than local MMs. Local MMs, on the other hand, may be used as tools for increasing coherence and persuasiveness (Dafouz-Milne, 2008; Waltereit, 2006), thus they should be used more often in more persuasive texts or text fragments (for example, summaries and discussions).

Both expressions may also be used as content words, not necessarily as part of a metatext. The term *content words* is used to describe all the expressions or words that are part of text content (rather than metadiscourse) and which refer to external reality or the mental processes of the characters described, and not to the text in which they are used. For example *described* in *described in Chapter 3* is an MM, while *symptoms described by the patient* would be classified as a content word. Moreover, citations and references to other texts will not be classified as MMs but as *evidentials* (after Hyland, 2005) — for example: *as described by Hyland (2005)* — even though Hyland treats them as part of textual metadiscourse.

Experiments 1 and 2 were designed to analyse the usage of the MMs *opisany and wspomniany*. The main research question was whether their usage is influenced by the meaning of the verbs the MMs are derived from. Experiment 1 is a corpus study and enables an analysis of the correlation between the MMs' verb-derived meaning and their scope in scientific texts. Experiment 2 is a fill-the-gap study which aims to determine whether the results of Experiment 1 are also applicable to very short texts.

2 Study 1: MMs meaning and scope — a corpus study

Study 1 was a corpus study aiming to analyse the frequency of usage of 3 expressions in Polish scientific texts:

a) *opisywany*, *opisany* (*described* in the imperfective and perfective aspect);

b) *wspomniany* (*mentioned*).

The main hypothesis of this study assumes that (according to the meaning of the verbs they are derived from and their association with the acts of describing, elaborating on, and mentioning) when used as an MM the first expression (*opisany*) will tend to have a wider (global) scope and the second expression (*wspomniany*) will tend to have a narrower (local) scope.

2.1 Corpus and method

The corpus consisted of 85 peer-reviewed papers from three issues of the proceedings of *Poznańskie Forum Kognitywistyczne* (*Poznan Cognitive Sciences Forum*, PFK) conference (editions 5–7, years 2009–2011; a shorter version of the corpus, including only editions 5–6, was described in Czoska, 2011). The texts were 1532 — 5672 word-long (mean=2835,39; SD=831,991), and sometimes the same participant authored or co-authored more than one text.

AntConc (Anthony, 2014) was used to search the corpus. Since *opisany* and *wspomniany* may end with different inflectional morphemes, the search terms were *opisan**, *opisywan** and *wspomnian**.

The occurrences of the expressions *opisany*, *opisywany*, *wspomniany* and *wspominany* were classified according to their function: a) *MMs*;

b) *Evidentials* (intertextual, referring to other papers);

c) *Content words*.

The scope of each MM was marked:

a) *Global* — if the text unit referred to was the topic of the whole text;

b) *Medium* — if the text unit referred to was in another paragraph or section;

c) *Local* — if the text unit referred to was in the same paragraph or sentence as the MM.

The scope was always marked according to the location of the last occurrence of the word or notion marked by the MM and whether it was the topic of the whole text or one of the chapters. Additional direction markings (such as *here*, *above*, *previously*, *in this paper* — see Table 1) were sometimes taken into account as bearing information on scope.

Additionally, the direction of the MM was annotated, along with its occurrence in complex MMs with explicit direction marking.

2.2 Results

In the corpus 215 occurrences of the expressions were found, including 152 MMs, 33 evidentials and 30 content words. Only *opisywany*/*opisany* were used as evidentials, whereas *wspomniany* was used almost exclusively as an MM (see Figure 1). The most frequently used form was *opisany* (68 occurrences). A chi-squared test performed over the three groups indicates that there are significant differences in the usage of the expressions (chi-squared = 37.284, $p < 0.001$) when form and function are taken into account (see Figure 1).

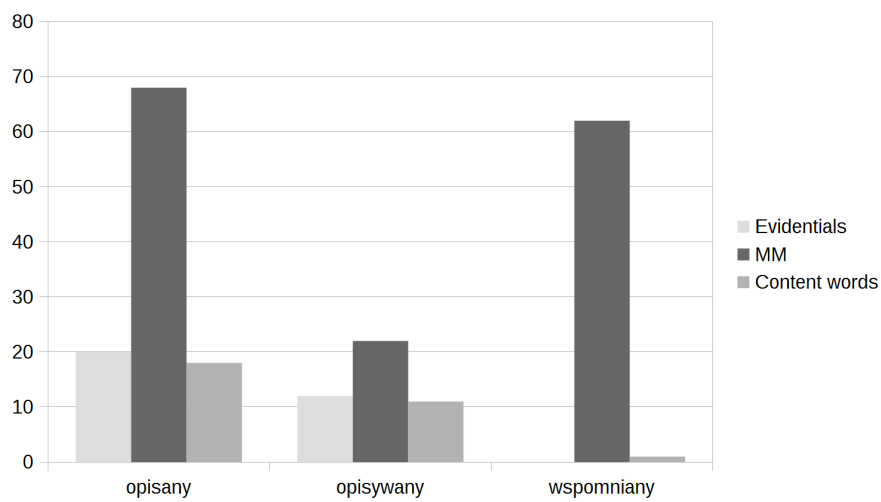

Figure 1: Functions of *opisany*/*opisywany* and *wspomniany* used in.

A medium scope was the most frequently attributed (68 times). There were 48 local and 36 global MMs in total. Both *opisany* and *wspomniany* may occur as an MM with all the scopes, but the distribution of scopes differs between them (see Figure 2). *Wspomniany* is used with local and medium scope rather than global (4 occurrences in 3 different texts). A chi-squared test performed over the three groups indicates that there are significant differences in the distribution of scope between the groups (chi-squared = 22.811, $p < 0.001$): *opisany* occurs with a broader scope (global and medium), while it is the least frequent form with regards to local scope.

Since the *opisywany* form does not enrich the information, as it is evenly distributed between global, medium and local usages, the two categories *opisywany* and *opisany* were collapsed into

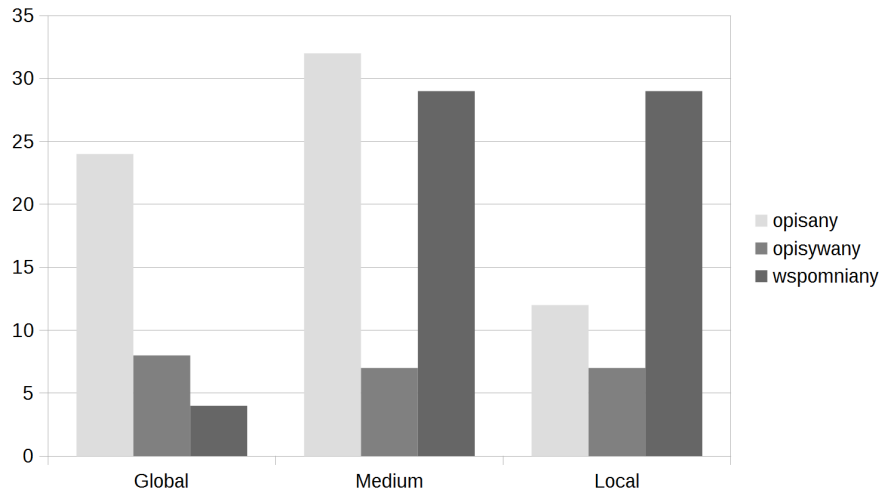

Figure 2: MM scopes.

one (see Figure 3), thus only the meaning of the verb, not its aspect, was to be taken into account. The tendency remains for *wspomniany* to be used more often with a narrower scope and *opisany* to be used when a reference with a global scope is required (chi-squared = 20.880, $p < 0.001$).

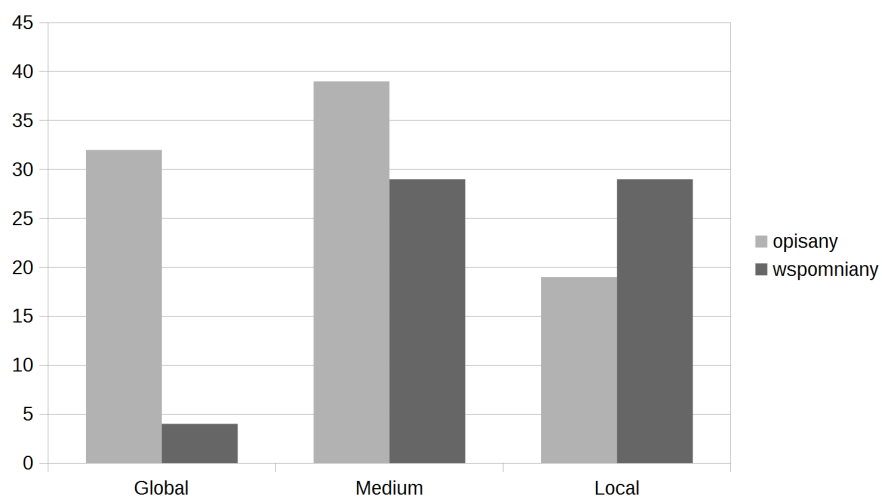

Figure 3: MM scopes (aspect ignored).

As Figure 3 illustrates, there were many examples of both types of MM with all the scopes. Some examples of each kind are presented below:

- Global scope (from text abstract):

  (1) *Wspomniana zostanie krytyka samej koncepcji hiperobliczalności.* (txt24)

  (1) [*Wspomniany*] *will be criticism of the conception of hipercomputability.*

- Medium scope (a preview announcing the topic of the next section):

(2) *W opisywanym niżej eksperymencie podjęto próbę manipulacji* (...). (txt4)

(2) *In the experiment* [*opisywany*] *below an attempt was made to manipulate* (...).

- Local scope (referring to a notion introduced in the first sentence of the current paragraph):

(3) *Innymi słowy wszelkie heurystyki powodujące wyżej wspomniane odchylenia stanowią* (...). (txt2)

(3) *In other words all heuristics that cause above* [*wspomniany*] *deviations are* (...).

As shown in the examples, *wspomniany* and *opisany* often occur in complex expressions such as *mentioned below* and the whole expression may be treated as one complex MM with an explicit direction marking or, in some cases, as a combination of two DMs (Fraser, 2015), since expressions like *tu* (*here*) or *powyżej* (*above*) may also occur as an MM. 67 of the MMs analysed occurred with an explicit direction marking (see Table 1). The majority (51) of the complex MMs were reviews and had a backward direction marked with *wyżej*, *powyżej*, *wcześniej*, *tu* or *już*. *Wyżej* occurred most frequently as an explicit direction marking element. Of the 85 unmarked MMs, only 13 were previews.

Table 1: Frequency of occurrence in a complex MM.

MM	no direction marking	wyżej (above)	powyżej (above)	wcześniej (before)	poniżej (below)	tutaj (here)	dalej (further)	już (already)
opisany	33	10	11	6	2	2	1	1
opisywany	18	1			2			
wspomniany	34	11	3	7		1		6
sum	85	22	14	13	4	3	1	7

There was a correlation between an MM's scope and its explicit direction marking (chi-squared = 16.16, $p = 0.003$; see Figure 4). Forward direction was marked only in MMs with a medium scope, and backward direction was marked only once in MMs with a global scope. If only the presence of an additional marker is taken into account, the results remain significant but weaker (chi-squared = 6.655, $p = 0.036$) and the main difference is between *opisywany* and *opisany* — the first marker was used without explicit direction marking.

As was mentioned in the introduction to the current section, *opisany* and *wspomniany* may also be used as part of a passive voice construction (e.g. *zostanie opisany — will be described*). In the corpus only 11 occurrences of this were found, 6 with a forward and 5 with a backward direction.

2.3 A brief interpretation of the results

The results of Experiment 1 suggest that the meaning of an MM, or the communication act of describing or mentioning the marker may be associated with, seems to influence the scope of the MM, and therefore its usage. However, this seems to be more visible in the usage of *wspomniany* than *opisany* (e.g. Figure 3). It may be the case that *opisany* (especially in its imperfective form *opisywany*) is less semantically marked, or that it seems to be a more prototypical repetition cue (Hyland, 2005) and therefore is used more often and in different contexts. Moreover, the choice of the exact expression to be used may be influenced not only by the distance from the text unit being referred to, but also by the direction of the MM: whether the text unit has already been given or is new (as especially in the case of the only instance of *wspomniany* with a global scope).

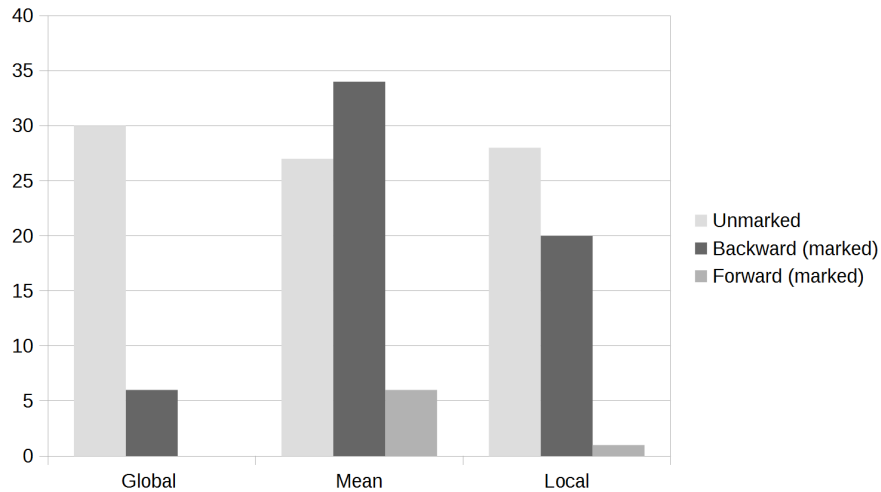

Figure 4: Frequency of the occurrence of explicit direction marking by MM scope.

Both expressions may be used as reviews (MRCs) and previews (see examples and Fig. 4) and thus refer to text units either given or new. The markers may have their direction explicitly marked, especially when they have a global scope.

3 Experiment 2: filling gaps with MMs

The method used in Experiment 2 is an example of the so called rational cloze-test. The category of cloze-test covers all the instances of gap filling or text completion. The first cloze-tests were created in the 1950s (Sadeghi, 2014) and applied a random elimination of every n-th word. Next, a rational cloze-test was developed, in which the elimination was restricted to words of a given linguistic category or motivated by the task. Intended by Taylor (1953, after Sadeghi, 2014) as a measure of readability, the cloze-test became one of the most popular methods for testing foreign language proficiency. Moreover, it may be used, more in keeping with the original idea, to measure the appropriateness of an expression in the context of a given text. It is claimed (Sadeghi, 2014; Storey, 1997) to be a method sensitive to contextual influences on the interpretation of the gap — what is missing according to the rest of the text provided.

The aforementioned appropriateness testing is on the one hand an example of the rational cloze-test and, on the other hand, it is in many cases a "selected deletion gap-filling test" during which participants are presented with a list of expressions that may be used to fill the gap (Bensoussan & Mauranen, 1989). The list may include options like "none expression at all" (there is no need of an additional expression), "other" (where participants may suggest alternative choices) and inappropriate options to control readers text understanding.

Rohde et al. (2016) used a single-choice cloze-test in their study on the interpretation of discourse relations via the usage of discourse connectives which signal the relations explicitly, for example *so*, *because*, *otherwise*, *instead*. In Rohde et al. (2016; as well as their previous study: Rohde, Dickinson, Clark, Louis, & Webber, 2015) the participants read passages from the NY Times Annotated Corpus that included two spans of text (in most cases: two phrases) that were originally connected by a discourse relation signalled with an adverbial like *instead*, *indeed*, *after all* and (in half of the cases) conjunctions such as *because* or *so* that disambiguated the relation signalled with the adverbial. For the cloze-test all the conjunctions were removed, and gaps were also inserted into the passages that did not contain any conjunction preceding the adverbial. The participants had to decide whether a conjunction should be placed in the gap and choose

one that "best reflects the meaning of the connection" (Rohde et al., 2015, p. 24) from the list of 6 possible conjunctions with an additional "none" and an "other" option. The results were used to create profiles of the adverbial's usage. Moreover, the results of the first study provide information on the appropriateness of the occurrence of complex DMs or, in more details, of conjunctions disambiguating the relation marked with the adverbials in short passages. Only in approximately 6% of the cases (half of which did not contain a conjunction in the original text) did the participants chose "none" instead of a conjunction. Taking into account the high proportion of agreement between the participants and the authors in the explicit passages (those originally containing a complex DM), the results show that in short passages readers may find explicit, precise relation signalling necessary and appropriate, even when the relations are easy to infer.

Storey (1997) also used a single-choice cloze-test to analyse the usage of coherence devices as an indicator of discourse processing ability. The participants in his study first read an extract from an educational textbook (a discussion of a problem with a solution implied) and were then presented with the cloze-test: a summary of the extract with "discourse markers, anaphoric pro-nouns and lexical substitutes acting cohesively" removed (Storey, 1997, p. 218). The DMs removed were, for example, *moreover*, *however* and *even though*. The aim of the study was to analyse the process of global text comprehension, the readers' perception of text structure and the relations between the excerpt and its summary, not only with the cloze-test but also with the think-aloud method. The results of the self-reports analysis indicated that "Items consisting of deleted discourse markers encouraged subjects to decompose the associated arguments and analyse the rhetorical structure of the text in some depth." (Storey, 1997, p. 226). Analysis of these results also showed that the participants judged the appropriateness of the available DMs in the local, direct context, even when they could use a more global context of a broader, original text.

The studies cited above show that a method known recently mostly to language teachers as a language competency test, may be used successfully to analyse not only readability on the level of text content, but also on the metatextual level: the understanding of text structure and relations between its elements. Used alongside corpus methods, rational, discourse level cloze-tests may be another source of data on the appropriateness and the naturalness of MMs or DMs in different contexts and registers.

In the aforementioned studies by Storey (1997) and Rohde et al. (2015, 2016) the cloze-tests required from the readers an interpretation of the local context only: two text spans, mostly phrases, adjacent and related by a relation relatively easy to infer. In the current study a single-choice cloze-test was also used, but the span of the MM that might be used to fill the gap surpassed the adjacent phrases: a term introduced in the first sentence of a text reappeared in the last (third) sentence, preceded by the gap that could be filled with the MMs *wspomniany* or *opisany*. Only filler items included gaps that could be filled with a rhetorical MM explicitly naming a relation between the adjacent phrases.

The aim of the study was to assess whether MRCs like *opisany* and *wspomniany* are judged as appropriate in very short scientific texts (3 or 4 sentences long) and whether the verb-derived meaning of an MM influences its choice in such a text (see Experiment 1).

Participants were asked to read texts and mark which MMs they would chose to fill a gap preceding a repetition of a term introduced in the first sentence of the text. Since all the gaps were put in the same place (introducing a repetition) and the texts were 3–4 sentences long, all the MMs had a backward direction and a narrow scope. Moreover, only participles were used as MMs.

3.1 Stimuli

For Experiment 2, 50 short summaries of scientific papers were prepared, 32 experiment stimuli (ES) and 18 fillers (F). Both types of texts consisted of a similar number of words (including function words): mean(ES) = 86.5, mean(F) = 85. In all the texts there was a gap to be filled by the participants.

In all the ES a term was introduced in the first sentence and then repeated in the last one (see Figure 5). The terms repeated may have had different levels of salience in the text (Kehler, 2004; Trnavac & Taboada, 2016): there were both topic-terms related to the main concept of a text (19 texts) and subtopic-terms related to a notion connected to the main term (13 texts). More generally, the topic-term was a notion defined or described throughout the text, and the subtopic-term was mentioned and important only within the first sentence. The correctness of all the texts was analysed by 3 independent experts.

In Figure 5 an example of the cloze-test task is presented (as seen by a participant). The text that may be filled with the MRC *aforementioned* describes a working memory text. It can be translated into the following paragraph (with the lexical repetition, marked with red circles in Fig. 5, marked with boldface):

> There are many tests for the working memory **functions** assessment. It has been indicated in many studies that improvement in those tests results correlates with the improvement in the general cognitive performance. A comparison of behavioral and neuroimaging studies outcomes done in Wisconsin shows that the reason of the improvement may be better communication between the brain areas active during those tasks. The areas are responsible for the ... [aforementioned]working memory **functions**, however some of them are active also during other cognitive tasks.

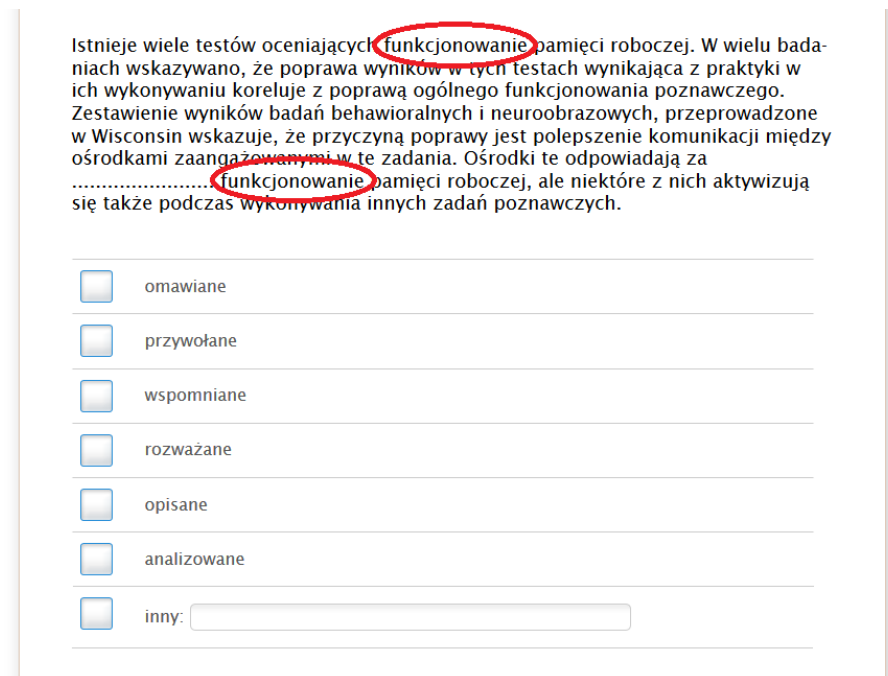

Figure 5: An example of a ES question (first appearance and repetition of the target word marked red). The translation of the text (target is marked with boldface, and the gap that could be filled with an MM — with the marker aforementioned).

In the ES condition the gap preceded the repetition and the list consisted of 6 metatextual repetition cues (MRCs):

a) *omawiany* (*discussed*);

b) *rozważany* (*considered*);

c) *opisany* (*described*);

d) *analizowany* (*analysed*);

e) *przywołany* (*recalled*);

f) *wspomniany* (*mentioned*).

The first four (a–d) and the last two MRCs (e–f) may be regarded as examples of two different types, of which are *opisany* and *wspomniany* are also examples (see Experiment 1). The first (a–d), according to the results of Experiment 1 (all refer to something being *described*), may be chosen more often when an MM with a wider scope is needed, whereas e and f have a narrower scope, since they refer to something only *mentioned* in the text.

In the F condition the participants were asked to choose a logical marker (F), also from a given list:

a) *między innymi* (*inter alia*);

b) *dla przykładu* (*for example*);

c) *na przykład* (*for example*);

d) *z kolei* (*in turn*);

e) *z drugiej strony* (*on the other hand*);

f) *ponadto* (*moreover*).

The logical markers in the above list differ in their information function: a–c indicate a list relation, d–e indicate contrast and f indicates an additive relation or elaboration. While in the ES condition there were no correct answers, in the F condition the participants had to choose the correct relation. Therefore, the F texts served as a control condition for monitoring text understanding.

3.2 Procedure and participants

18 participants, undergraduates and doctoral students from the Institute of Linguistics, took part in the experiment. There were two different lists of texts, so each participant read only half of the texts(16 ES and 9 F). The order of the texts for each list was semi-randomised before creating the questionnaire, but was the same for all participants reading the same list. An F text always appeared first. The participants were asked to choose a marker from multiple-choice lists to fill the gap in the text. In both ES and F conditions the ES and F lists (presented in the previous subsection) were presented in 4 different orders. An interactive, on-line questionnaire was created with Instantly Survey Tool (Instantly, 2016).

3.3 Results

70 people started the questionnaire but only 18 participants finished it. 16 of these participants took part in the experiment individually and their behaviour or surroundings were not controlled. 2 participants completed the questionnaire in controlled conditions while visiting a university laboratory. Time spent on the questionnaire was measured with Instantly Survey Tool as the time from clicking on the survey link to sending the results (with a link provided in the survey). In 4 cases the participants spent more than 1.5 hour completing the questionnaire, and the average time spent in the other 14 cases was approximately 21 minutes, which is closer to the result of the 2 participants whose behaviour was controlled (approximately 42 and 22 minutes).

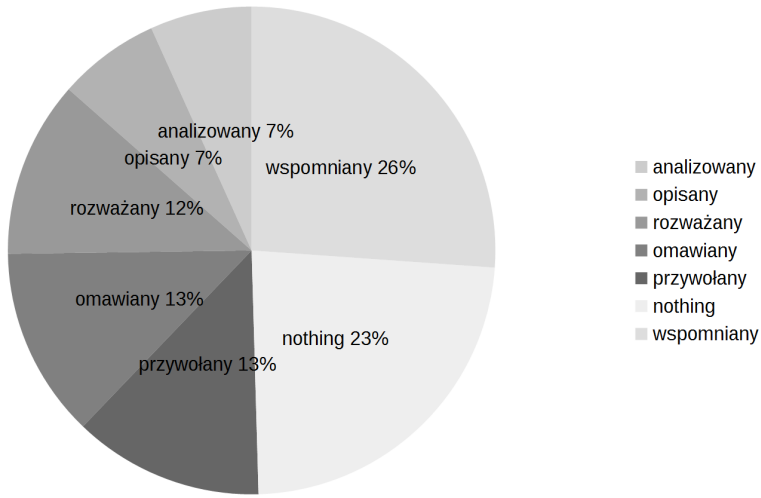

Figure 6: MMs chosen.

There were no errors in the F condition, which suggests that the participants understood the texts and chose all the markers according to their understanding.

In the ES condition (see Figure 6) the participants decided to insert an MM into the gap more often than to leave it empty (*nothing*, 23%) and the most frequently chosen MM was *wspomniany* (26%). MMs of a wide-scope type were chosen in 49% of cases when an MM was assumed to be appropriate, and MMs of a narrow-scope type were selected in 51% of cases.

An analysis of correspondence was performed to check whether there was any difference in the choice of MMs (from the options a–f) between the texts in which a word connected with topic and subtopic was repeated. The results were significant only if the F condition was also taken into account (chi-squared = 471.217, $p < 0.001$) and indicated that *wspomniany* was chosen more often to mark the repetition of subtopic terms and *opisany* and *omawiany* were selected more often to mark the repetition of topic terms. When only the ES condition was considered, the difference between topic and subtopic was non-significant (chi-squared = 5.5, $p = 0.482$). Thus, there was no difference in terms of which MMs were chosen to mark a repetition of a topic or a subtopic term.

Only when MMs a–d and e–f were treated as two groups (wide and narrow scope), did there seem to be a difference between the metatextual repetition cues types. MMs were chosen more frequently to fill a gap preceding a subtopic-term repetition than a topic-term repetition. However, this is only a tendency: while MMs of wider scope-types were chosen almost as frequently in those texts, narrow-scope MMs were chosen slightly more frequently to indicate a topic-term repetition (chi-squared = 15.5367, $p < 0.001$). However, when only the presence or absence of an MM is taken into account (wide-scope and narrow-scope MMs are counted together) the difference between topic- and subtopic-term repetition announcing becomes more significant (chi-squared = 14.4777, $p < 0.001$).

3.4 A brief interpretation of the results

MRCs seem to be considered necessary or natural not only in longer texts, but also in short summaries. In the context of a very short text the meaning of a MRC, whether it is connected to describing or mentioning, does not seem to influence its choice. However, when filling gaps in short texts the participants preferred *wspomniany* over *opisany*.

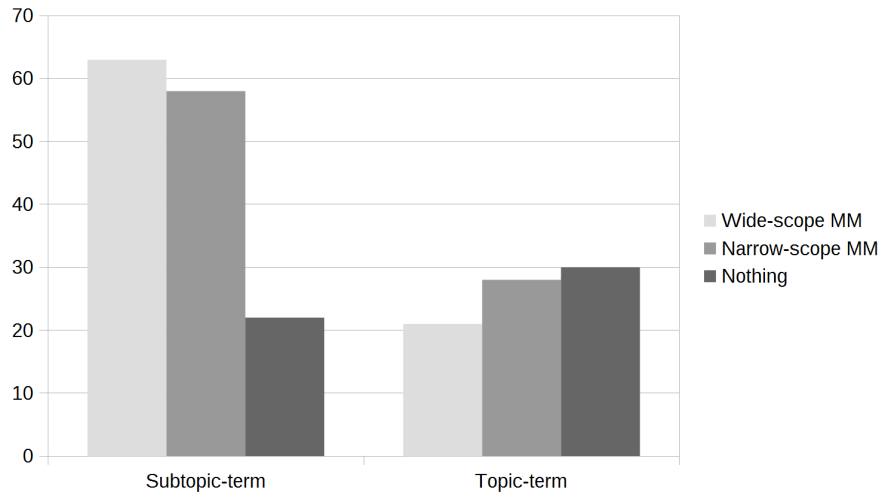

Figure 7: MMs chosen.

The repetition of a topic or a subtopic term did not influence the choice of a MM, but it did influence the frequency of choosing an MM rather than to leave the repetition "bare", with the announcement of a subtopic-term repetition being more frequent. Since MMs may be viewed as text processing instructions, an instruction to re-activate concepts that are not the main topic (and thus may not be active throughout the whole text) may be more relevant to a reader than to further activate a concept already active and accessible (which would to some extent replicate the results from Experiment 1, in which global markers where less frequent than others).

4 Conclusions and general discussion

The meaning of an MM influences the way it is used in a text and empirical predictions about MM usage may also be based on its linguistic form where the difference between two MMs is a question of quantity, such as between *describing* and *mentioning*, and not quality, such as between different logical markers like *consequently* and *in contrast*. In Experiment 1, evidence was found for a correlation between the MM meaning (defined as the textual act it is connected to) and its scope. Moreover, the scope and accessibility of the text unit a marker refers to correlated in the scientific texts written in Polish with the usage of additional explicit direction marking.

The results of Experiment 2 indicate that the aforementioned correlation between the marker's meaning and scope may be significant in texts whose length makes the difference between local and global scope large enough to influence the accessibility of the information provided by the marker. Moreover, the size of the whole text may also influence the choice and the appropriateness of an MRC to be used in it. This adds further weight to the idea that the meaning and function of an MM should be analysed in the context of a text. Moreover, this correlation may be mediated by the importance or role of the term or notion which is to be repeated: in short texts, subtopic-terms are more likely to be marked than topic-terms which, remain in focus throughout the whole text. These results are in line with propositions by Lemarié et al. (2008), who claim that the accessibility of text units which MMs refer to, or the informativeness of the MMs, may be one of the crucial features influencing a marker's usage and processing. Across a long text differences in accessibility between topic-terms and terms relevant in only one section or paragraph will be more prominent than across a text that consists of one paragraph only.

The notion of informativeness may further explain the results of both experiments. For example, in Experiment 1 a minority of the MMs from the corpus had a global scope (they referred to a

concept that was supposed to be active at every stage of the text processing; this replicates the results obtained on a shorter corpus in Czoska, 2011). Moreover, direction was rarely explicitly marked when an MM had a global scope and referred to a text unit which should be held active throughout the text processing. In Experiment 2 the majority of the MMs were used to fill gaps before subtopic-term repetitions. All the target terms occurred in each text only once but the topic should be activated and accessible to the reader throughout the whole text. Consequently, it seems to be more relevant to instruct the reader to reactivate a concept that may be inactive and thus inaccessible (see Lemarié et al., 2008 and Sanders & Gernsbacher, 2004 for a further discussion on accessibility). Another notable finding of the studies may be the fact that in Experiment 2 the participants tended to chose an MM to fill the gap rather than to leave it empty. This indicates that MMs are considered appropriate not only in long texts, but also in short texts. Consequently, both long and short texts may be used in empirical studies to test MMs processing, although the results may not be fully generalisable.

The results obtained in the current study may apply mostly to scientific and popular science texts, which are formal, present structured argumentation, and are characterised by a specialised jargon. Additionally, the structure of such texts is strongly conventionalised, which may make it relatively easy for an experienced reader to update what has already been given and what is expected at a given moment in a text (Dahl, 2004; Graesser, McNamara, & Louwerse, 2003; Mur-Dueñas, 2011). Conversely, these texts are written to be persuasive and their content should be presented as clearly and systematically as possible (Dafouz-Milne, 2008; Hyland, 1998). This may increase the usage of MMs and other means of instructing a reader on how to process the text.

References

Ädel, A. (2006). *Metadiscourse in L1 and L2 English*. Amsterdam: John Benjamins Publishing. https://doi.org/10.1075/scl.24

Anthony, L. (2014). AntConc (Version 3.4.3). Retrieved May 4, 2015, from http://www.laurenceanthony.net

Bensoussan, M., & Mauranen, A. (1989). SeDelGap tests of macro-level reading comprehension: An exercise in international collaboration. Retrieved May 4, 2016, from https://eric.ed.gov/?id=ED309644

Blakemore, D. (2002). *Relevance and linguistic meaning: The semantics and pragmatics of discourse markers*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511486456

Czoska, A. (2011). Klasyfkacja operatorów metatekstowych i częstość ich występowania w krótkich tekstach naukowych w języku polskim. *Investigationes Linguisticae*, *23*, 1–33. https://doi.org/10.14746/il.2011.23.1

Dafouz-Milne, E. (2008). The pragmatic role of textual and interpersonal metadiscourse markers in the construction and attainment of persuasion: A cross-linguistic study of newspaper discourse. *Journal of Pragmatics*, *40*(1), 95–113. https://doi.org/10.1016/j.pragma.2007.10.003

Dahl, T. (2004). Textual metadiscourse in research articles: A marker of national culture or of academic discipline? *Journal of Pragmatics*, *36*(10), 1807–1825. https://doi.org/10.1016/j.pragma.2004.05.004

Fischer, K. (Ed.). (2006). *Approaches to discourse particles*. Amsterdam: Elsevier.

Fraser, B. (1999). What are discourse markers? *Journal of Pragmatics*, *31*(7), 931–952. https://doi.org/10.1016/S0378-2166(98)00101-5

Fraser, B. (2015). The combining of discourse markers — A beginning. *Journal of Pragmatics*, *86*, 48–53. https://doi.org/10.1016/j.pragma.2015.06.007

Goldman, S. R., & Rakestraw, J. A. Jr. (2000). Structural aspects of constructing meaning from text. In M. L. Kamil, P. B. Mosenthal, P. D. Pearson, & R. Barr (Eds.), *Handbook of reading research* (Vol. 3, pp. 311–335). Mahwah, NJ: Lawrence Erlbaum.

Gonen, E., Livnat, Z., & Amir, N. (2015). The discourse marker axshav ('now') in spontaneous spoken Hebrew: Discursive and prosodic features. *Journal of Pragmatics*, *89*, 69–84. https://doi.org/10.1016/j.pragma.2015.09.005

Graesser, A. C., McNamara, D. S., & Louwerse, M. M. (2003). What do readers need to learn in order to process coherence relations in narrative and expository text. In A. P. Sweet & C. E. Snow (Eds.), *Rethinking reading comprehension* (pp. 82–98). New York, NY: Guilford Press.

Gray, B. (2010). On the use of demonstrative pronouns and determiners as cohesive devices: A focus on sentence-initial this/these in academic prose. *Journal of English for Academic Purposes*, *9*, 167–183. https://doi.org/10.1016/j.jeap.2009.11.003

Grochowski, M. (2008). Operatory metatekstowe o kształcie superlatiwu przysłówka. *Juznoslovenski filolog*, *64*, 61–72.

Hyland, K. (1998). Persuasion and context: The pragmatics of academic metadiscourse. *Journal of Pragmatics*, *30*(4), 437–455. https://doi.org/10.1016/S0378-2166(98)00009-5

Hyland, K. (2015). Metadiscourse. In K. Tracy, T. Sandel, & C. Ilie (Eds.), *The International Encyclopedia of Language and Social Interaction* (pp. 1–11). Hoboken, NJ: John Wiley & Sons, Inc. https://doi.org/10.1002/9781118611463.wbielsi003

Hyland, K. (2016). Academic publishing and the myth of linguistic injustice. *Journal of Second Language Writing*, *31*, 58–69. https://doi.org/10.1016/j.jslw.2016.01.005

Instantly. (2016). Retrieved January 17, 2016, from http://www.instant.ly/

Kehler, A. (2004). Discourse topics, sentence topics, and coherence. *Theoretical Linguistics*, *30*(2–3), 227–240. https://doi.org/10.1515/thli.2004.30.2-3.227

Lemarié, J., Lorch, R. F. Jr., Eyrolle, H., & Virbel, J. (2008). SARA: A text-based and reader-based theory of signaling. *Educational Psychologist*, *43*(1), 27–48. https://doi.org/10.1080/00461520701756321

Lewis, D. (2006). Discourse markers in English: A discourse-pragmatic view. In K. Fischer (Ed.), *Approaches to discourse particles* (pp. 43–60). Amsterdam: Elsevier.

Mur-Dueñas, P. (2011). An intercultural analysis of metadiscourse features in research articles written in English and in Spanish. *Journal of Pragmatics*, *43*(12), 3068–3079. https://doi.org/10.1016/j.pragma.2011.05.002

Ożóg, K. (1991). Elementy metatekstowe ze składnikiem "mówię" w polszczyźnie mówionej. In J. Bartmiński & R. Grzegorczykowa (Eds.), *Język a kultura* (Vol. 4) Wrocław: "Wiedza o Kulturze".

Rohde, H., Dickinson, A., Clark, C., Louis, A., & Webber, B. (2015). Recovering discourse relations: Varying influence of discourse adverbials. In *Proceedings of the EMNLP 2015 Workshop on Linking Models of Lexical, Sentential and Discourse-level Semantics* (pp. 22–31). https://doi.org/10.18653/v1/W15-2703

Rohde, H., Dickinson, A., Schneider, N., Clark, C. N. L., Louis, A., & Webber, B. (2016). Filling in the blanks in understanding discourse adverbials: Consistency, conflict, and context-dependence in a crowdsourced elicitation task. In *Proceedings of LAW X — The 10th Linguistic Annotation Workshop* (pp. 49–58). https://doi.org/10.18653/v1/W16-1707

Roulet, E. (2006). The description of text relation markers in the Geneva model of discourse organization. In K. Fischer (Ed.), *Approaches to discourse particles* (pp. 115–131). Amsterdam: Elsevier.

Sadeghi, K. (2014). Phrase cloze: A better measure of reading. *The Reading Matrix*, *14*(1).

Sanders, T., & Gernsbacher, M. A. (2004). Accessibility in text and discourse processing. *Discourse Processes*, *37*(2), 79–89. https://doi.org/10.1207/s15326950dp3702_1

Schiffrin, D. (1988). *Discourse markers* (No. 5). Cambridge University Press.

Storey, P. (1997). Examining the test-taking process: A cognitive perspective on the discourse cloze test. *Language Testing*, *14*(2), 214–231. https://doi.org/10.1177/026553229701400205

Tanghe, S. (2016). Position and polyfunctionality of discourse markers: The case of Spanish markers derived from motion verbs. *Journal of Pragmatics*, *93*, 16–31. https://doi.org/10.1016/j.pragma.2015.12.002

Trnavac, R., & Taboada, M. (2016). Cataphora, backgrounding and accessibility in discourse. *Journal of Pragmatics*, *93*, 68–84. https://doi.org/10.1016/j.pragma.2015.12.008

Waltereit, R. (2006). The rise of discourse markers in Italian: A specific type of language change. In K. Fischer (Ed.), *Approaches to discourse particles* (pp. 61–67). Amsterdam: Elsevier.

Wiemer, B. (2006). Particles, parentheticals, conjunctions and prepositions as evidentiality markers in contemporary Polish (A first exploratory study). *Studies in Polish Linguistics*, *3*(1), 5–67.

Winiarska, J. (2001). *Operatory metatekstowe w dialogu telewizyjnym.* Kraków: Wydawnictwo Universitas.

This work was financed by the Polish Ministry of Science and Higher Education.

The author declares that she has no competing interests.

Publisher: Institute of Slavic Studies, Polish Academy of Sciences, University of Silesia & The Slavic Foundation