

Jarosław Wiliński

Uniwersytet w Siedlcach

ORCID: <https://orcid.org/0000-0002-3136-6529>

e-mail: [jaroslaw.wilinski@uph.edu.pl](mailto:jaroslaw.wilinski@uph.edu.pl)

## ***Speaking-AC vs. AC-speaking in Polish:*** **A distinctive collexeme analysis**

**Konstrukcje z quasi-imiesłowem *mówiąc* w języku polskim:  
Analiza dystynktywnych koleksemów**

### **Abstract**

Utilizing the concept of *construction* and the corpus-based technique known as *Distinctive-Collexeme Analysis*, this research aims to identify adverbial complements (ACs) that exhibit a distinct inclination toward the *mówiąc*-AC-construction as opposed to the AC-*mówiąc*-construction. Drawing upon data extracted from NKJP, this study reveals that these constructions convey nuanced functional distinctions and consistently co-occur with specific ACs.

**Keywords:** quasi-participle *mówiąc*, NKJP, construction grammar, distinctive collexeme analysis, Fisher's exact test

### **Abstrakt**

Niniejsze badanie wykorzystuje pojęcie *konstrukcji* oraz metodę korpusową, znaną jako *Analiza Dystynktywnych Koleksemów*, aby zidentyfikować uzupełnienia okolicznikowe (ACs), które wykazują wyraźną tendencję do występowania w konstrukcji *mówiąc*-AC w przeciwieństwie do konstrukcji AC-*mówiąc*. Na podstawie danych pozyskanych z NKJP badanie to ukazuje, że obie konstrukcje charakteryzują się subtelnymi różnicami funkcjonalnymi, a także konsekwentnie współwystępują z określonymi przysłówkami i wyrażeniami przysłówkowymi.

**Słowa kluczowe:** quasi-imiesłów *mówiąc*, NKJP, gramatyka konstrukcyjna, analiza dystynktywnych koleksemów, test Fishera

## **1. Introduction**

The availability of computerized corpora and the increased interest in functional vocabulary have brought attention back to the phenomenon of Polish metatextual expressions. These expressions take the form of the

quasi-participle *mówiąc* ‘speaking’ preceded or followed by adverbs, nouns, or adverbial phrases. This linguistic phenomenon has been observed by various researchers (e.g., Ożóg 1991; Stępień 2014; Birzer 2017; Kubicka 2017; Żabowska 2020). Some researchers have traced the origins of these constructions (Birzer 2017) and addressed questions about their lexical status (Weiss 2005; Kisiel and Żabowska 2011; Stępień 2014). They have also distinguished between *[Adv] speaking* and *speaking [Adv]* (Żabowska 2020). Other scholars have briefly discussed the nature of specific expressions (Wróbel 1970; Weiss 2005; Grochowski 2002; Żabowska 2009; Bojałkowska 2010; Kisiel and Żabowska 2011), explained their functions and prosodic characteristics (Stępień 2014), and examined various formal and functional properties of adverbs (Kubicka 2017; Moroz 2007 for their formal characteristics). Additionally, hypotheses have been formulated regarding the circumstances under which adverbs can operate independently (see Kleszczowa 2015; Kubicka 2017).

To date, there has been limited research comparing and contrasting the usage of adverbs, nouns, and adverbial phrases in both the *mówiąc*-AC construction and the AC-*mówiąc* construction, not to mention their functional classification and detailed quantification and evaluation in terms of statistical significance. Previous research (Żabowska 2020) has focused exclusively on making observations regarding the status of both constructions and identifying several collocates within these structures: i.e., words that occur frequently and repeatedly with these specific constructions. Żabowska noted that these constructions serve distinct functions, with these differences corresponding to variations in word order and prosodic features. Furthermore, they exhibit variances in the lexical and formal attributes of their adverbial components, which subsequently impact the possibility of negation and the conditions for the independent use of adverbs. The *[ADV] mówiąc* pattern functions as a metatextual linguistic unit. In contrast, the *mówiąc [ADV]* pattern serves as a spoken commentary (Żabowska 2020: 33). It emerges as a result of an operation on the action verb *mówić* and represents the outcome of the quasi-participial transformation (Bogusławski and Danielewicz 2005: 347).

While Żabowska’s study has contributed to our understanding of both constructions’ status, it falls short in explaining the nuanced differences in their function and specific collocational preferences. Consequently, there is still a need to investigate the distribution of ACs and their frequency of occurrence in these patterns, as variations in distribution may exist. This paper, therefore, seeks to discern subtle but consequential distinctions between the two patterns. It aims to support the hypothesis that, despite

their apparent synonymy at first glance, these constructions exhibit nuanced differences in usage and demonstrate distinct preferences for specific categories of adverbs, nouns, and adverbial expressions.

The rest of this paper is structured into three sections. Section 2 discusses the theoretical and methodological foundations that form the basis of the quantitative examination. Section 3 combines the results of the quantitative analysis with a functional portrayal of ACs, shedding light on subtle differences in distribution between the two constructions. In Section 4, we undertake an assessment of the findings.

## 2. Theoretical and methodological frameworks

The study is based on the notion of *construction* (Goldberg 2006). This concept assumes that grammar consists of pairings of linguistic form and meaning across varying degrees of specificity and schematicity, ranging from morphemes and individual words to syntactic patterns. They are accompanied by associated rules governing their pragmatic and discourse-functional elucidation. For example, *mówiąc*-AC is recognized as a construction because it has a particular form—a quasi-participle combined with a specific AC—and a meaning/function. Specifically, it expresses the idea of a speaker communicating a message to a recipient in a particular manner. Notably, all linguistic units may be represented as constructions, provided they manifest with the requisite frequency<sup>1</sup>. This frequency, in turn, governs their level of entrenchment within a speaker's or hearer's grammatical system (Croft and Cruse 2004: 292–293).

The *Distinctive-Collexeme Analysis* method (Gries and Stefanowitsch 2004) was employed to identify ACs that exhibit distinctive associations with either the *mówiąc*-AC-construction or the AC-*mówiąc*-construction: i.e., those words that best distinguish between both constructions. This method involved a series of four steps, which can be exemplified using the adverb *wprost* 'bluntly'. Table 1 presents the distribution of this adverb in both patterns, along with the other frequencies necessary for conducting a distinctive-collexeme analysis.

---

<sup>1</sup> In this study, a specific threshold value for "requisite frequency" is set based on the p-value derived from the Fisher's exact test, which indicates the likelihood of the observed data occurring under the null hypothesis (i.e., no association between the words and constructions). A typical threshold for significance is a p-value of 0.05. If the p-value is below this threshold, you reject the null hypothesis and conclude that the words and constructions are significantly associated.

**Table 1.** Contingency table cross-tabulating frequency scores of the constructions under study

Constructions	ACs ( <i>wprost</i> )	All other ACs	Total
<b><i>mówiąc</i>-AC</b>	Frequency of AC ( <i>wprost</i> ) in the <i>mówiąc</i> -AC-construction A = 324 (93.99)	Frequency of all other ACs in the <i>mówiąc</i> -AC-construction b = 4110	Total frequency of the <i>mówiąc</i> -AC-construction x = 4434
<b>AC-<i>mówiąc</i></b>	Frequency of AC ( <i>wprost</i> ) in the AC- <i>mówiąc</i> -construction c = 3 (233.01)	Frequency of all other ACs in the AC- <i>mówiąc</i> -construction d = 10989	Total frequency of the AC- <i>mówiąc</i> -construction y = 10992
<b>Total</b>	Total frequency of AC ( <i>wprost</i> ) e = 327	Total frequency of all other ACs f = 15099	Total frequency of both constructions z = 15426

The first step involved searching for these patterns in the balanced version of the National Corpus of Polish (NKJP), extracting their occurrences in concordance lines, and calculating the observed frequencies. This version of the corpus encompasses over 250 million textual words, comprising samples from a variety of written and spoken sources in Polish, including classic literature, daily newspapers, specialist periodicals, and several others (see Przepiórkowski et al. 2012). The PELCRA corpus search engine (Pęzik 2012) was employed to search for all instances of the participle *mówiąc* within a context of 5 words and a sample size of 10,000 words<sup>2</sup>. Manual inspection of concordance lines allowed for identifying potential collocates of the two patterns, such as *wprost mówiąc* and *mówiąc wprost*. These combinations were then entered into the search engine once again, and their total frequencies were automatically extracted from the corpus within a maximum span of five words to the left and right of the participle.

Initially, the corpus search uncovered 17 potential combinations of *wprost mówiąc* and 346 combinations of *mówiąc wprost*. Subsequently, these instances of both constructions were manually examined to identify genuine combinations, resulting in 3 occurrences of *wprost mówiąc* and 324 occurrences of *mówiąc wprost*. Any occurrences that did not conform to the pattern or indicated co-reference between the subject of the participle and the first argument of the matrix verb were discarded from further

<sup>2</sup> The module of automatic extraction of collocations implemented in the search engine PELCRA, called *Kolokator* 'Collocator', was not used in this study due to its technical limitations. First, *Kolokator* only allows for extracting collocates within a maximum distance of two words to the left and right of the node. Secondly, the module only searches for words that co-occur at least five times in the balanced version of the corpus.

analysis<sup>3</sup>. This procedure was repeated for each combination, thereby allowing for calculating the observed frequencies of the patterns and the remaining instances of ACs. The figures (a, x, c, y) in Table 1 were obtained directly from the corpus, while the remaining values were derived through addition and subtraction.

Next, the observed frequencies were employed to ascertain the expected frequencies of the AC (*wprost*) in both constructions. This calculation was executed in Microsoft Excel by means of the following formula:  $(e \times x) / z$ . For the sake of clarity, the expected frequencies for the adverb *wprost* in each construction are enclosed in parentheses (see Table 1). Should the observed frequency of the AC (*wprost*) in the *mówiąc*-AC-construction significantly deviate from the expected value, it signifies a strong association between the AC *wprost* and this construction, characterized by either attraction or repulsion. In essence, the AC *wprost* is consequently classified as a collexeme significantly drawn towards or repelled from the *mówiąc*-AC-construction, respectively. Similarly, if the observed frequency of the AC (*wprost*) in the AC-*mówiąc*-construction significantly diverges from the expected frequency, then the adverb *wprost* is deemed a significantly attracted or repelled collexeme of this construction.

In the third step, the strength of association between the AC (*wprost*) and the constructions under examination was estimated. This was achieved by utilizing the figures (a, b, c, d) from Table 1, which were then input into a 2-by-2 table and subjected to the Fisher exact test. The p-value obtained from this test served as the metric for evaluating the collocation strength of each adverbial complement. In essence, it quantified the extent to which the AC (e.g., *wprost*) exhibited an affinity for the *mówiąc*-AC-construction as opposed to the AC-*mówiąc*-construction. A smaller p-value indicated a higher probability that the observed distribution was not a result of chance, thereby signifying a stronger association between an AC and one of the constructions in question. The computation of statistical significance was executed using an online calculator designed for the Fisher's exact test applicable to two-by-two contingency tables.

Following this, each p-value (e.g., 1.40e-173 for *wprost*) was logarithmically transformed to base 10 (e.g., 172.85). This transformation provided a more readable score compared to the often-cumbersome p-values expressed

---

<sup>3</sup> For example, the occurrence of *wprost* and *mówiąc* in *Gage odpowiada brutalnie, choć nie całkiem wprost, mówiąc, że Hamlet [...]* (Gage responds brutally, though not entirely directly, speaking of how Hamlet [...]) was considered to be a standard use of the participle *mówiąc*. Thus, this instance was omitted from the statistical calculations to preserve the accuracy of the data set.

in powers of ten. Collostruction strength values exceeding 1.301 indicated a robust association between the AC and the construction, while values below this threshold indicated a significant deviation from the construction. The number 1.301 is approximately equal to  $-\log_{10}(0.05)$ , which is the negative logarithmic transformation of the commonly used significance threshold  $p=0.05$ .

This means that any value above 1.301 on the log scale corresponds to p-values less than 0.05, indicating statistical significance. Importantly, this analysis identified the adverb *wprost* as a lexeme of substantial significance within one of the two constructions, albeit without specifying which one. To ascertain this distinction, the observed frequencies of *wprost* were compared to the expected frequencies. This comparison revealed that *wprost* occurred more frequently than expected in the *mówiąc*-AC-construction and less frequently than expected in the AC-*mówiąc*-construction. Thus, it is a lexeme more strongly associated with the former than with the latter.

Finally, the outcomes of the quantitative research were arranged in accordance with the strength of attraction and subsequently subjected to qualitative assessment. Specifically, the quantitative findings were integrated with a functional classification of ACs. The classification was guided by the observed functions and patterns of usage of the ACs within both constructions as they appear in the corpus data. These functions were identified based on empirical evidence.

### 3. Results and discussion

A thorough examination of concordance lines has revealed that the *mówiąc*-AC-construction and the AC-*mówiąc*-construction occur 4,434 and 10,992 times, respectively, within the corpus. In simpler terms, the latter construction appears to be more than twice as prevalent as the former, approximately 2.48 times more frequent, with each of the first five lexemes occurring over one thousand times in that construction (see Table 4 below). The *mówiąc*-AC-construction shows associations with 208 different types of ACs, with 78 of them occurring only once in this specific construction. Conversely, the AC-*mówiąc*-construction is associated with 139 different ACs, 49 of which are exclusively used in this construction when compared to the AC-*mówiąc*-construction.

These results suggest that many ACs appear in both constructions, though with varying frequencies. For example, words such as *najprościej* ‘the simplest way’, *ogłędnie* ‘cautiously’, *nieładnie* ‘improperly’, *konkretnie* ‘specifically’,

*krótko* ‘briefly’, *szczerze* ‘sincerely’, *inaczej* ‘differently’, and others occur in both contexts but are more strongly associated with one of the constructions when we consider their expected frequencies and low p-values. Additionally, words such as *językiem* ‘in language’, *wprost* ‘directly’, *serio* ‘seriously’, *słowami* ‘in the words of’, *bez ogródek* ‘bluntly’, *żargonem* ‘in jargon’, *w przenośni* ‘figuratively’, and *banalnie* ‘trivially’ are more unique to the *mówiąc*-AC (left) context, while *prawdę* ‘truth’, *nawiasem* ‘incidentally’, *właściwie* ‘actually’, and *lekko* ‘lightly’ are uniquely associated with the AC-*mówiąc* (right) context. Their frequencies differ significantly: for instance, *językiem* is much more frequent in the *mówiąc*-AC-construction (350 occurrences) compared to the AC-*mówiąc*-construction (4 occurrences). Conversely, *prawdę* shows the opposite pattern, with 24 occurrences in *mówiąc*-AC and 1789 occurrences in AC-*mówiąc*.

### 3.1. Findings for the *mówiąc*-AC-construction

Tables 2 and 3 below show the sixty most distinctive collexemes associated with the *mówiąc*-AC-construction, including the observed frequencies of ACs used to determine the expected frequencies, the direction of association (attraction or repulsion), and the strength of association (distinctiveness). Additionally, the tables present the results of the distinctive-collexeme analysis, including p-values and collostructional strength based on the Fisher exact test. The application of this test can be easily justified statistically. In statistics, it is particularly used when at least one expected value within the table falls below the threshold of 5 or when data distribution is characterized by unevenness and infrequency (Gries and Stefanowitsch 2004: 10)

The findings suggest that five particular ACs, namely *językiem*, *wprost*, *obrazowo*, *w skrócie*, and *prościej*, exhibit significant distinctiveness. The log transformations, which serve as indicators of their distinct character, are remarkably high: 185.66, 172.85, 48.24, 46.83, and 41.25, respectively. These values represent the strength of association between particular ACs and the *mówiąc*-AC-construction, where a higher value (above 1.301) indicates a stronger, more statistically significant association. A comparison between the observed and expected frequencies of each of these ACs in relation to the two constructions reveals that these ACs occur more frequently than expected in the *mówiąc*-AC-construction and less frequently than expected in the AC-*mówiąc*-construction. In essence, they demonstrate a very strong degree of distinctiveness when comparing the former to the latter. Noteworthy, *językiem* emerges as the most robust collexeme for the



**Table 2.** The thirty most distinctive collexemes of the *mówiąc-AC*-construction

rank	AC	a	c	e	f	x	y	z	b	d	(a)	(c)	p-values	coll. strength
1.	językiem	350	4	354	15072	4434	10992	15426	4084	10988	101.75	252.25	2.18e-186	185.66
2.	wprost	324	3	327	15099	4434	10992	15426	4110	10989	93.99	233.01	1.40e-173	172.85
3.	obrazowo	135	25	160	15266	4434	10992	15426	4299	10967	45.99	114.01	5.78e-49	48.24
4.	w skrócie	125	20	145	15281	4434	10992	15426	4309	10972	41.68	103.32	1.47e-47	46.83
5.	prościej	103	13	116	15310	4434	10992	15426	4331	10979	33.34	82.66	5.57e-42	41.25
6.	poważnie	134	38	172	15254	4434	10992	15426	4300	10954	49.44	122.56	5.25e-41	40.28
7.	serio	74	3	77	15349	4434	10992	15426	4360	10989	22.13	54.87	1.51e-36	35.82
8.	słowami	61	0	61	15365	4434	10992	15426	4373	10992	17.53	43.47	6.96e-34	33.16
9.	kolokwialnie	74	20	94	15332	4434	10992	15426	4360	10972	27.02	66.98	1.04e-23	22.98
10.	w uproszczeniu	54	6	60	15366	4434	10992	15426	4380	10986	17.25	42.75	3.20e-23	22.49
11.	po prostu	53	6	59	15367	4434	10992	15426	4381	10986	16.96	42.04	1.01e-22	21.99
12.	potocznie	60	17	77	15349	4434	10992	15426	4374	10975	22.13	54.87	4.52e-19	18.35
13.	brutalnie	46	13	59	15367	4434	10992	15426	4388	10979	16.96	42.04	6.03e-15	14.22
14.	bez ogródek	25	0	25	15401	4434	10992	15426	4409	10992	7.19	17.81	2.77e-14	13.56
15.	najprościej	110	97	207	15219	4434	10992	15426	4324	10895	59.50	147.50	1.60e-13	12.80



16.	po ludzku	41	12	53	15373	4434	10992	15426	4393	10980	15.23	37.77	2.94e-13	12.53
17.	ogłędnie	84	76	160	15266	4434	10992	15426	4350	10916	45.99	114.01	2.99e-10	9.52
18.	precyzyjnie	30	9	39	15387	4434	10992	15426	4404	10983	11.21	27.79	6.13e-10	9.21
19.	żargonem	16	0	16	15410	4434	10992	15426	4418	10992	4.60	11.40	2.13e-09	8.67
20.	prosto	24	5	29	15397	4434	10992	15426	4410	10987	8.34	20.66	2.31e-09	8.64
21.	skrótowo	21	4	25	15401	4434	10992	15426	4413	10988	7.19	17.81	1.45e-08	7.84
22.	fachowo	16	1	17	15409	4434	10992	15426	4418	10991	4.89	12.11	2.64e-08	7.58
23.	żartobliwie	25	9	34	15392	4434	10992	15426	4409	10983	9.77	24.23	8.07e-08	7.09
24.	w przenośni	12	0	12	15414	4434	10992	15426	4422	10992	3.45	8.55	3.15e-07	6.50
25.	otwarcie	19	5	24	15402	4434	10992	15426	4415	10987	6.90	17.10	4.37e-07	6.36
26.	banalnie	11	0	11	15415	4434	10992	15426	4423	10992	3.16	7.84	1.10e-06	5.96
27.	dosadnie	14	2	16	15410	4434	10992	15426	4420	10990	4.60	11.40	1.67e-06	5.78
28.	żartem	14	2	16	15410	4434	10992	15426	4420	10990	4.60	11.40	1.67e-06	5.78
29.	ogólniej	16	4	20	15406	4434	10992	15426	4418	10988	5.75	14.25	2.94e-06	5.53
30.	patetycznie	15	4	19	15407	4434	10992	15426	4419	10988	5.46	13.54	8.25e-06	5.08

Explanations: **a** = Observed frequency of AC (e.g. *wprost*) in the *mówiąc*-AC-construction; **b** = Frequency of all other ACs in the *mówiąc*-AC-construction; **c** = Observed frequency of ACs (e.g. *wprost*) in the AC-*mówiąc*-construction; **d** = Frequency of all other ACs in the AC-*mówiąc*-construction; **e** = Total frequency of AC (e.g. *wprost*); **f** = Total frequency of all other ACs; **x** = Total frequency of the *mówiąc*-AC-construction; **y** = Total frequency of the AC-*mówiąc*-construction; **z** = Total frequency of both constructions; **(a)** = Expected frequency of AC (e.g. *wprost*) in the *mówiąc*-AC-construction; **(c)** = Expected frequency of AC (e.g. *wprost*) in the AC-*mówiąc*-construction; **p-values** and **collostructional strength** = indexes of statistical significance

*mówiąc*-AC-construction, with an exceptionally high collostructional strength (185.66). Furthermore, the expected frequency of *językiem* is lower than the observed frequency within the pattern under examination.

The ACs in both tables can be categorized into various functional groups. At the top position is *językiem* ‘in language’, while at rank 19 is *żargonem* ‘in jargon,’ both of which are associated with domain-specific communication. More specifically, phrases such as *mówiąc językiem Kanta* ‘speaking in the language of Kant’ or *mówiąc żargonem piłkarskim* ‘speaking in the language of football’ are used by speakers to suggest a shift towards specialized or technical discourse related to specific fields (philosophy or football), which may not necessarily involve simplification but rather precision and accuracy within those specialized domains.

Another prominent category involves words that serve the purpose of enhancing clarity and simplification in discourse. They are used to state that the following information will be presented in a concise, simple, and easily digestible manner, effectively removing unnecessary complexity. The ACs within this category include *wprost* ‘bluntly’, *w skrócie*, ‘in short’ *prościej* ‘more simply’, *w uproszczeniu* ‘in simplified terms’, *po prostu* ‘simply’, *najprościej* ‘most simply’, *precyzyjnie* ‘precisely’, *prosto* ‘plainly/simply’, *skrótowo* ‘concisely’, *jaśniej* ‘more clearly’, *trywialnie* ‘trivially’, *lapidarnie* ‘succinctly’, *konkretnie* ‘more concretely’, *precyzyjniej* ‘more precisely’, and *czytelniej* ‘more legibly’. These ACs, ranked at positions 2, 4, 5, 10, 11, 15, 18, 20, 21, 35, 37, 40, 43, 45, and 58 respectively, serve multiple functions in discourse when used alongside the participle *mówiąc*. For example, they introduce statements that enhance clarity (*jaśniej* or *czytelniej*), simplify complex ideas (*prościej*, *najprościej*, and *w uproszczeniu*), provide conciseness (*w skrócie* or *skrótowo*), ensure directness (*wprost* and *po prostu*), or improve accessibility of information (*skrótowo*, *jaśniej*, or *precyzyjniej*).

The third category is associated with figurative language or style. This category encompasses a wide array of linguistic expressions, figures of speech, rhetorical devices, stylistic elements, and expressive choices that extend beyond straightforward, literal communication, thus enhancing the depth, nuance, and vividness of the discourse. In the hierarchy of the ACs belonging to this category, *obrazowo* ‘figuratively’ occupies the third rank and stands as the foremost representative in this category. It is accompanied by *fachowo* ‘professionally’, *żartobliwie* ‘jokingly’, *w przenośni* ‘metaphorically’, *banalnie* ‘banally’, *żartem* ‘as a joke’, *patetycznie* ‘grandiosely’, *metaforycznie* ‘metaphorically’, *górnolotnie* ‘loftily’, *pół żartem pół serio* ‘half-jokingly, half-seriously’, *w (jakiś) sposób* (np. *obrazowy*) ‘in a (certain) manner (e.g., figurative)’, *eufemistycznie* ‘euphemistically’, *elegancko* ‘elegantly’,

*przenośnie* ‘figuratively’, *uczenie* ‘eruditely’, *cynicznie* ‘cynically’, *symbolicznie* ‘symbolically’, *oficjalnie* ‘officially’, and *naukowo* ‘scientifically’. These expressions are distributed across ranks 22 to 57.

These figurative and stylistic expressions serve a range of essential functions in discourse: they are used to signal the speaker’s intent to engage audiences by vividly painting mental images and evoking emotions (*obrazowo*, *górnolotnie*, or *żartobliwie*), simplify and clarify complex ideas (*w przenośni*), introduce humor or wit (*żartobliwie* or *żartem*), and establish various tones and styles (*oficjalnie*, *elegancko*, or *patetycznie*). Additionally, they comment on and modify the following text by fostering creativity and artistic expression (*obrazowo*, *przenośnie*, or *symbolicznie*) and by encouraging active interpretation, inviting readers or listeners to think critically and draw connections (*obrazowo* or *metaforycznie*). Finally, they can play a pivotal role in effective communication, introducing nuanced meanings (*obrazowo* or *symbolicznie*) and establishing the speaker’s authority and expertise in a subject (*uczenie* or *naukowo*).

In the following category, a variety of ACs are employed to emphasize seriousness, directness, or candor in communication, ensuring that the speaker’s upcoming message is conveyed clearly and with a sense of gravity. The term *poważnie*, which means ‘seriously’ and ranks 6th in significance, stands out as the most prominent lexeme within this group. Additionally, we find *serio* ‘seriously’, *brutalnie* ‘brutally’, *bez ogródek* ‘without beating around the bush’, *otwarcie* ‘openly’, *dosadnie* ‘bluntly’, *dosadniej* ‘more bluntly’, and *bez osłonek* ‘without covering up’ ranked at positions 7, 13, 14, 25, 27, 42 and 54, respectively. While all of these phrases can convey a sense of seriousness and directness to some extent, they may vary in the degree to which they emphasize one function over the others.

The next group of the most significant lexemes comprises a variety of ACs used for introducing statements that provide additional explanations, clarifications, or summaries. *Słowami* ‘in the words of’, ranked 7th, occupies the top position among the most distinctive ACs in this category<sup>4</sup>. The phrase is employed to summarize or clarify points by referencing the ideas or terminology of scholars or philosophers like Einstein or Kant, thus rendering complex concepts more accessible and enriching the discussion with additional context. Other words falling into this category include *ogólniej*

<sup>4</sup> Note that expressions such as *mówiąc językiem*, *słowami*, or *żargonem* (‘speaking in language, words, or jargon’) do not constitute complete structures. The necessity to complete them determines the placement of the phrase that complements *mówiąc* and affects the proximity or lack thereof in the analyzed combinations.

**Table 3.** Other distinctive collexemes of the *mówiąc-AC*-construction

rank	AC	a	c	e	f	x	y	z	b	d	(a)	(c)	p-values	coll. strength
31.	metaforycznie	15	4	19	15407	4434	10992	15426	4419	10988	5.46	13.54	8.25e-06	5.08
32.	górnolotnie	15	4	19	15407	4434	10992	15426	4419	10988	5.46	13.54	8.25e-06	5.08
33.	szerzej	20	10	30	15396	4434	10992	15426	4414	10982	8.62	21.38	1.81e-05	4.74
34.	pół żartem pół serio	8	0	8	15418	4434	10992	15426	4426	10992	2.30	5.70	4.64e-05	4.33
35.	jaśniej	8	0	8	15418	4434	10992	15426	4426	10992	2.30	5.70	4.64e-05	4.33
36.	w cudzysłowie	11	2	13	15413	4434	10992	15426	4423	10990	3.74	9.26	4.65e-05	4.33
37.	trywialnie	12	3	15	15411	4434	10992	15426	4422	10989	4.31	10.69	5.70e-05	4.24
38.	zwyczajnie	9	1	10	15416	4434	10992	15426	4425	10991	2.87	7.13	9.88e-05	4.01
39.	po naszemu	9	1	10	15416	4434	10992	15426	4425	10991	2.87	7.13	9.88e-05	4.01
40.	lapidarnie	10	2	12	15414	4434	10992	15426	4424	10990	3.45	8.55	1.38e-04	3.86
41.	nieładnie	22	15	37	15389	4434	10992	15426	4412	10977	10.64	26.36	1.50e-04	3.82
42.	dosadniej	11	4	15	15411	4434	10992	15426	4423	10988	4.31	10.69	4.44e-04	3.35
43.	konkretnie	23	19	42	15384	4434	10992	15426	4411	10973	12.07	2.93	4.59e-04	3.34
44.	po polsku	16	10	26	15400	4434	10992	15426	4418	10982	7.47	18.53	6.46e-04	3.19

45.	precyzyjniej		18	14	32	15394	4434	10992	15426	4416	10978	9.20	22.80	1.28e-03	2.89
46.	w sposób		5	0	5	15421	4434	10992	15426	4429	10992	1.44	3.56	1.96e-03	2.71
47.	szczegółowo		5	0	5	15421	4434	10992	15426	4429	10992	1.44	3.56	1.96e-03	2.71
48.	eufemistycznie		16	12	28	15398	4434	10992	15426	4418	10980	8.05	19.95	2.46e-03	2.61
49.	elegancko		9	4	13	15413	4434	10992	15426	4425	10988	3.74	9.26	2.90e-03	2.54
50.	przenośnie		6	1	7	15419	4434	10992	15426	4428	10991	2.01	4.99	2.97e-03	2.53
51.	uczenie		6	1	7	15419	4434	10992	15426	4428	10991	2.01	4.99	2.97e-03	2.53
52.	cynicznie		8	3	11	15415	4434	10992	15426	4426	10989	3.16	7.84	3.18e-03	2.50
53.	symbolicznie		7	2	9	15417	4434	10992	15426	4427	10990	2.59	6.41	3.27e-03	2.49
54.	bez osłonek		4	0	4	15422	4434	10992	15426	4430	10992	1.15	2.85	6.82e-03	2.17
55.	oficjalnie		4	0	4	15422	4434	10992	15426	4430	10992	1.15	2.85	6.82e-03	2.17
56.	nieelegancko		4	0	4	15422	4434	10992	15426	4430	10992	1.15	2.85	6.82e-03	2.17
57.	naukowo		4	0	4	15422	4434	10992	15426	4430	10992	1.15	2.85	6.82e-03	2.17
58.	czytelniej		4	0	4	15422	4434	10992	15426	4430	10992	1.15	2.85	6.82e-03	2.17
59.	ostrożnie		5	1	6	15420	4434	10992	15426	4429	10991	1.72	4.28	8.94e-03	2.05
60.	innymi słowy		11	8	19	15407	4434	10992	15426	4423	10984	5.46	13.54	9.00e-03	2.04

Explanations: see tab. 2.

‘more broadly’, *szerszej* ‘more broadly’, *w cudzysłowie* ‘in quotation marks’, *szczegółowo* ‘in detail’, and *innymi słowy* ‘in other words’. Depending on the context of their usage and the specific communicative goal, these words and phrases can serve dual functions. *Słowami*, *innymi słowy*, or *w cudzysłowie* tend to introduce varying degrees of clarity or precision in the upcoming message, while *ogólniej* and *szerszej* often impose a broader or summarized perspective on the following information.

The ranking list also includes ACs such as *kolokwialnie* ‘colloquially’, *potocznie* ‘informally’, *po ludzku* ‘in a human manner’, *zwyczajnie* ‘commonly’, *po naszemu* ‘in our way’, *nieładnie* ‘rudely’, *po polsku* ‘in Polish’, and *nieelegancko* ‘inelegantly’. The primary function of these words and phrases is to signal that the upcoming message will be delivered in a relaxed, approachable, or colloquial atmosphere, thereby creating a sense of closeness and approachability between the speaker and the audience.

Finally, among the distinctive collexemes of this construction, there are *ogłędnie* ‘cautiously’ and *ostrożnie* ‘cautiously’ at ranks 17 and 59. These words are used when the speaker wishes to address sensitive topics with care, aiming to avoid explicitness, directness, or confrontation. They play a crucial role in maintaining a polite and considerate tone during communication, thereby minimizing the risk of causing offense or misunderstanding.

### 3.2. Findings for the AC-*mówiąc*-construction

Table 4 displays the results of the distinctive collexeme analysis for the most strongly attracted collexemes of the AC-*mówiąc*-construction, compared directly with the *mówiąc*-AC-construction. Surprisingly, these are the only distinctive collexemes for this construction, as all other ACs occur more frequently than expected in the *mówiąc*-AC-construction. Notably, *prawdę* ‘truth’ emerges as the most distinctive collexeme for the AC-*mówiąc*-construction. This is supported by the exceptionally high log transformation of the p-value (2.59e-241), resulting from the Fisher exact test for this noun, which stands at 240.59. Furthermore, upon comparing the observed values with the expected ones, it becomes evident that *prawdę* occurs more frequently than expected by chance in the AC-*mówiąc*-construction compared to the *mówiąc*-AC-construction<sup>5</sup>.

<sup>5</sup> Note that the p-values for *generalnie*, *ściśle*, and *najładniej* are greater than 0.05, and their collocation strength is below 1.301, meaning they are not statistically significant under the typical threshold of  $p < 0.05$ , where values above 1.301 indicate statistical significance. However, they remain distinctive in the AC-*mówiąc* construction

**Table 4.** The twelve most distinctive collexemes of the AC-*mówić*-construction

rank	AC	a	c	e	f	x	y	z	b	d	(a)	(c)	p-values	coll. strength
1.	prawdę	24	1789	1813	13613	4434	10992	15426	4410	9203	521.12	1291.88	2.59e-241	240.59
2.	nawiasem	49	1525	1574	13852	4434	10992	15426	4385	9467	452.43	1121.57	3.16e-170	169.50
3.	krótko	344	1826	2170	13256	4434	10992	15426	4090	9166	623.74	1546.26	5.78e-51	50.24
4.	szczerze	384	1786	2170	13256	4434	10992	15426	4050	9206	623.74	1546.26	3.49e-37	36.46
5.	inaczej	301	1251	1552	13874	4434	10992	15426	4133	9741	446.10	1 105.90	8.78e-19	18.06
6.	właściwie	1	129	130	15296	4434	10992	15426	4433	10863	37.37	92.63	7.01e-18	17.15
7.	ściśle	72	320	392	15034	4434	10992	15426	4362	10672	112.68	279.32	1.88e-06	5.73
8.	delikatnie	160	588	748	14678	4434	10992	15426	4274	10404	215.00	533.00	3.34e-06	5.48
9.	lekko	0	11	11	15415	4434	10992	15426	4434	10981	3.16	7.84	4.10e-02	1.39
10.	generalnie	7	32	39	15387	4434	10992	15426	4427	10960	11.21	27.79	1.58e-01	0.80
11.	ściśle	28	93	121	15305	4434	10992	15426	4406	10899	34.78	86.22	1.90e-01	0.72
12.	najagodniej	4	11	15	15411	4434	10992	15426	4430	10981	4.31	10.69	6.65e-01	0.18

Explanations: see tab. 2.



*Prawdę mówiąc* ‘to tell the truth’ and *szczerze mówiąc* ‘honestly speaking’, ranked 1 and 4, are employed to introduce statements characterized by veracity and integrity, emphasizing the speaker’s commitment to truthfulness and sincerity in their forthcoming discourse. *Nawiasem mówiąc* ‘incidentally’, ranked 2, in turn introduces a parenthetical remark somewhat related to the main topic<sup>6</sup>. *Krótko mówiąc* ‘to put it briefly’, ranked third, is used to indicate the speaker’s intention to summarize or provide a concise version of a longer explanation. The next lexeme is *inaczej* ‘differently’ at rank 5. When employing this phrase, a speaker signals the intent to provide an alternative explanation, primarily with the aim of fostering a clearer understanding of the subject matter.

The ranking list of distinctive collexemes for the construction in question includes three ACs that denote different degrees of precision in communication. Ranked 6th is *właściwie* ‘actually’, followed by *ściśle* ‘strictly’ at 7th and *ściślej* ‘more strictly’ at 11th. These lexemes, combined with the participle *mówiąc*, are used to introduce more precise statements, stressing the importance of accuracy and strict interpretation. There is also one lexeme, *generalnie* ‘generally’ at rank 10, which can be used to make broad or sweeping statements that apply to a wide range of situations or instances.

Lastly, the lower portion of Table 3 encompasses ACs that are typically employed when the speaker wishes to temper their language, mitigate potentially harsh or critical statements, or delicately approach a topic. *Delikatnie* ‘delicately’, ranked 8th, holds the highest position among the lexemes within this functional category, followed by *lekko* ‘lightly’ at rank 9 and *najłagodniej* ‘most gently’ at rank 12.

#### 4. Concluding remarks

This study unveils substantial dissimilarities in the functional and structural constraints imposed on the ACs permissible within each construction. The *mówiąc*-AC-construction does not collocate with *lekko*

---

when compared to the *mówiąc*-AC construction. Therefore, they are presented in Table 4 to highlight the differences between the two constructions.

<sup>6</sup> In this study, constructs such as *nawiasem mówiąc* or *krótko mówiąc* can generally be treated as parentheticals, as they introduce supplementary comments that are related to, but not essential for, the main sentence. These parenthetical remarks can be marked by punctuation or shifts in tone and serve to provide additional context or side information. However, defining parenthesis is complex, as many expressions can function in similar ways, making it important to distinguish them from other structures like parataxis or embedded clauses.

and is relatively rarely used with *prawdę*, *nawiasem*, and *właściwie*. On the other hand, the construction does manifest a strong tendency to occur with *językiem*, *żargonem*, *wprost*, *obrazowo*, *w skrócie*, *prościej*, *poważnie*, *serio*, *słowami*, *kolokwialnie*, *w uproszczeniu*, *po prostu*, *bez ogródek*, *ostrożnie*, and many others. Thus, its primary function in discourse is to emphasize the use of domain-specific jargon or terminology, clarify and simplify complex ideas, and enrich and embellish the discourse by adding figurative language, rhetorical devices, and stylistic elements. Additionally, it is used to emphasize seriousness, directness, or candor in communication, provide additional explanations, clarifications, or summaries, establish a relaxed, approachable, or colloquial tone, or address delicate or sensitive topics with care and caution.

In contrast, the AC-*mówiąc*-construction is employed to emphasize truthfulness, introduce parenthetical remarks, summarize points, provide alternatives, enhance precision, make broad statements, or temper language. This construction is notably selective in its choice of ACs. It does not occur with terms such as *słowami*, *bez ogródek*, *żargonem*, *w przenośni*, *banalnie*, *pół żartem* *pół serio*, *jaśniej*, and several others. Furthermore, it is infrequently used with expressions like *językiem*, *wprost*, *serio*, *w uproszczeniu*, *po prostu*, *precyzyjnie*, *prosto*, *skrótowo*, *fachowo*, *dosadnie*, *żartem*, and several others. However, the construction does exhibit a strong preference for specific ACs, including *prawdę*, *nawiasem*, *krótko*, *szczerze*, *inaczej*, *właściwie*, *ściślej*, *delikatnie*, *lekko*, *generalnie*, *ściśle*, and *najłagodniej*. Notably, the first five lexemes (*prawdę*, *nawiasem*, *krótko*, *szczerze*, and *inaczej*) account for the bulk of ACs in the AC-*mówiąc*-construction, making up a total of 8,177 occurrences within this specific construction. Consequently, this underscores the rationale for considering these expressions not merely as constructs, instances of the construction in question, but as distinct constructions in their own right. Their exceptional frequency and highly conventionalized usage have established them as the customary choice for a specific communicative context through repeated usage, thereby confirming their status as established linguistic units.

Both constructions can be seen as specific subtypes of the *mówiąc* construction, which in turn instantiates a metatextual construction, a construction that comments on or modifies the text itself, giving the speaker a way to signal how the message should be interpreted. The two distinct meanings of these constructions can be inferred from their typical uses. AC-*mówiąc* typically introduces a statement that reflects the attitude or perspective the speaker adopts in delivering the message; hence, it often signals the speaker's intent to be honest, concise, or candid in their communication. In contrast, *mówiąc*-AC specifies how the speech is delivered,

often emphasizing the style, mode, or manner of speaking. Thus, it relates to the form, clarity, or rhetorical style of the communication. These specialized meanings are grounded in the general concept of metatextual commentary, where the speaker's language itself is being used to guide how the listener should interpret the following discourse.

### Literature

- Birzer S. (2017): *Assessing the role of pattern and matter replication in the development of Polish discourse structuring elements based on non-finite 'verba dicendi'*. "Zeitschrift für Slavische Philologie" 73(1), pp. 159–185.
- Bogusławski A., Danielewiczowa M. (2005): *Verba polona abscondita. Sonda słownikowa III*. Warszawa.
- Bojałkowska K. (2010): *Opis składniowy imiesłowów przysłówkowych we współczesnym języku polskim*. Toruń.
- Croft W., Cruse D. (2004): *Cognitive Linguistics*. Cambridge.
- Goldberg A. (2006): *Constructions at work: The nature of generalization in language*. Oxford.
- Gries S.Th., Stefanowitsch A. (2004): *Extending collocation analysis: A corpus-based perspective on alternations*. "International Journal of Corpus Linguistics" 9, pp. 97–129.
- Grochowski M. (2002): *Wielowyzrazowe jednostki funkcyjne. Wprowadzenia do problematyki*. [In:] *Problemy frazeologii europejskiej* V. A.M. Lewicki (ed.). Lublin, pp. 35–50.
- Kisiel A., Żabowska M. (2011): *O zakresie klasy partykuł w słowniku gniazdowym partykuł polskich*. „Polonica” 31, pp. 113–132.
- Kleszczowa K. (2015): *U źródeł polskich partykuł. Derywacja funkcjonalna, przemiany, zaniki*. Katowice.
- Kubicka E. (2017): *Jak mówimy jakoś mówiąc? Formalne i semantyczne właściwości adverbialnych uzupełnień quasi-imiesłowowego mówiąc*. "LingVaria" 12(23), pp. 99–113.
- Moroz A. (2007): *Uwagi o ciągach parentetycznych z segmentem mówiąc*. [In:] *Studia nad słownictwem dawnym i współczesnym języków słowiańskich*. J. Kamper-Warejko, I. Kaproń-Charzyńska (eds). Toruń, pp. 187–194.
- Ożóg K. (1990): *Leksykon metatekstowy współczesnej polszczyzny mówionej. Wybrane zagadnienia*. Kraków.
- Pęzik P. (2012): *Wyszukiwarka PELCRA dla danych NKJP*. [In:] *Narodowy Korpus Języka Polskiego*. A. Przepiórkowski, M. Bańko, R. Górski, B. Lewandowska-Tomaszczyk (eds). Warszawa, pp. 253–279.
- Przepiórkowski A., Bańko M., Górski R.L., Lewandowska-Tomaszczyk B. (2012): *Narodowy Korpus Języka Polskiego*. Warszawa.
- Stępień M. (2014): *Wyrażenia parentetyczne w strukturze wypowiedzi – właściwości semantyczne, składniowe, prozodyczne*. Warszawa.
- Weiss D. (2005): *Nowe przyimki o pochodzeniu imiesłowowym?* [In:] *Przysłówki i przyimki. Studia ze składni i semantyki języka polskiego*. M. Grochowski (ed.). Toruń, pp. 177–207.
- Żabowska M. (2009): *Wyrażenia metatekstowe w funkcji parentezy*. [In:] *Język z różnych stron widziany*. M. Skarżyński, A. Czelakowska (eds) Kraków, pp. 157–166.
- Żabowska M. (2020): *Szyk wyrażen jako cecha dystynktywna w polu komentarzy metatekstowych: [Adv] mówiąc, \_ vs. mówiąc [Adv], \_*. „Poradnik Językowy” 9, pp. 21–37.