Language Sciences 102 (2024) 101603

Contents lists available at ScienceDirect

Language Sciences

journal homepage: www.elsevier.com/locate/langsci

A corpus-based study of maximizer–adjective patterns in Croatian

Ivan Lacić

Department of Classical Philology and Italian Studies, Alma Mater Studiorum – University of Bologna, via Zamboni 32, 40126 Bologna (BO), Italy

ARTICLE INFO

Article history:
Received 3 August 2023
Received in revised form 23 October 2023
Accepted 8 November 2023
Available online xxx

Keywords:
Cluster analysis
Collostructional analysis
Correspondence analysis
Croatian
Maximizers

ABSTRACT

Maximizers represent a subclass of degree modifiers that convey the highest degree to which a property can be carried out. This paper studies five Croatian near-synonymous maximizers (all meaning "completely, totally"), viz. *posve*, *potpuno*, *sasvim*, *skroz,* and *totalno*, as a part of <maximizer + adjective> construction. It is assumed here that analysed pairings act as (semi)-prefabricated units with maximizers that impose particular modes of construal. To analyse the subtle semantic differences of examined maximizers, we shall turn to the distributional hypothesis and examine contexts in which maximizers occur. Using a combination of analytical statistics (collostructional analysis) and multifactorial methods (hierarchical agglomerative cluster analysis and correspondence analysis), we aim to examine similarities (proximities) and differences (distances) between analysed constructions in order to understand intricate relationships among maximizers, fostering valuable insights into their semantics. The findings of this study provide insight into the interplay of the Croatian maximizers and adjectives.

1. Introduction

Degree modifiers (DMs), also known as intensifiers (Bolinger, 1972; Quirk et al., 1985), constitute linguistic items employed to modify other elements in terms of degree. They are typically adverbs, as in *very expensive*, *extremely hot*, *quite intelligent*, *somewhat luxurious*. Throughout the 20th century and into the present day, DMs have been a subject of considerable scholarly interest. Pioneering studies by researchers like Stoffel (1901) have offered extensive inventories of intensifying adverbs in contemporary and earlier English forms, shedding light on their origins. In recent years, due to the emergence of computerised corpora, there has been a renewed focus on intensifiers, and their semantics, as well as variability and capacity for change, have captured considerable attention (*inter alia* Paradis, 1997, 2001; Tagliamonte, 2008).

Despite their significance, DMs in the Croatian language have not undergone extensive analysis, and when studied, the investigations have primarily adopted a cross-linguistic perspective. For instance, Pavić Pintarić and Frleta (2014) examined the typology of upwards intensifiers in English, German, and Croatian, utilising a relatively limited parallel corpus of Harry Potter novels (mean of 240 sentences per language). Similarly, Batinić et al. (2015) explored the intensifying function of German modal particles and their analogous modal expressions in Croatian and English, investigating whether these particles could express varying degrees of intensity and types of intensification. Matešić and Memišević (2016) focused on evaluative adjectives and their modifiers in scientific texts from distinct domains (linguistics and medicine) in Croatian and English. Additionally, Nigoević (2020) conducted a comprehensive contrastive study of intensification in Croatian and Italian,

E-mail address: ivan.lacic2@unibo.it.

https://doi.org/10.1016/j.langsci.2023.101603
0388-0001/

showcasing the principal linguistic means used for intensification in both languages through examples gathered from diverse corpora. In spite of these prior works, a comprehensive examination of DMs in Croatian is still lacking, emphasising the need for further research in this area. Specifically, there has been relatively modest attention dedicated to exploring collocational pairings between Croatian DMs and adjectives, and studies employing analytical statistics and multifactorial methods to address this inquiry are, to the best of the author's knowledge, yet to be undertaken. Furthermore, the semantics of Croatian DMs, particularly near-synonymous ones, has not received any substantive examination thus far.

This study aims to fill the aforementioned research gaps by examining five Croatian degree modifiers, namely maximizers—*posve*, *potpuno*, *sasvim*, *skroz*, and *totalno*—as a part of <maximizer + adjective> construction. All five maximizers are Croatian equivalents of "completely, totally, perfectly". This analysis, as shown in Examples 1–5, focuses exclusively on prototypical contexts where the examined maximizers are used as degree modifiers of adjectives[1] (examples from *hrWaC* 2.2):

(1) Toliko o nejasnoćama, a sad o onome što je *sasvim* jasno.
("So much for the ambiguities, and now for what is perfectly clear.")

(2) Naše putovanje krenulo je *skroz* opušteno i tako je ostalo tijekom cijelog našeg boravka.
("Our trip started off completely relaxed and stayed that way throughout our stay.")

(3) Spomenuta Sedgwickova reakcija na darwinizam je *posve* razumljiva […]
("The mentioned Sedgwick's reaction to Darwinism is completely understandable […]")

(4) […] a izgled stranica i predlošci su *potpuno* prilagodljivi vašem sadržaju i željama.
("[…] and the layout of the pages and templates are completely customisable to your content and wishes.")

(5) […] ma kakve su joj to koščobe na ramenima, noge *totalno* ružne.
("[…] what are those bones on her shoulders, her legs are totally ugly.")

Since the scrutinised maximizers are considered near-synonyms, elements characterised by similarities in meaning while exhibiting different distributions across various contexts (Taylor, 2003), this investigation endeavours to measure conceptual content similarities alongside construal differences and subsequently represent them graphically. Such a method, relying on the Behavioural Profile approach (Divjak and Gries, 2006), is expected to unveil subtle semantic distinctions between these near-synonymous constructions (Desagulier, 2014).

In order to analyse subtle semantic differences between maximizers, we shall turn to the distributional hypothesis and build upon the assumption that a "difference of meaning correlates with difference of distribution" (Harris, 1970: 785). This principle asserts that a correlation between distributional similarity and meaning similarity enables us to infer the latter from the former. Drawing inspiration from Desagulier's (2014, 2015) methodological approach, this study aims to explore distinctions among the aforementioned maximizers by examining their collocational profiles as indicative of their divergent semantics. The working hypothesis, in alignment with Desagulier's framework (2014), postulates that an overlap in collocation preferences among *skroz*, *sasvim*, *posve*, *potpuno*, and *totalno* would suggest a shared conceptual content, classifying them as near-synonyms. Conversely, discrepancies would reveal that these five modifiers impose distinct manners of construal.

The goal of this paper is twofold. Firstly, we are interested in examining the similarities in the collocational preferences of maximizers *skroz*, *sasvim*, *posve*, *potpuno,* and *totalno* to determine the extent to which their conceptual content is shared. Secondly, we focus on recognising the differences among the five analysed maximizers by observing the distinct construals these adverbs impose on the conceptual content of the modified adjectives. By scrutinising the similarities and differences in the collocational profiles of maximizers, we aim to gain valuable insights into their semantics.

The paper is structured as follows. Section 2 briefly reviews some of the key concepts regarding maximizers. Section 3 illustrates the corpus and the methodology. Section 4 presents the results. In Section 5, we discuss and interpret the key findings.

2. Maximizers

Maximizers represent quantificational lexical items that convey the highest degree (maximal amount or quantity on a given/implied scale) to which a property can be carried out (Quirk et al., 1985; Israel, 2001; Kennedy, 2003). In English, examples of maximizers include *absolutely*, *completely*, *entirely*, *totally*, *fully*, and *perfectly*, while in Croatian we encounter *skroz*, *sasvim*, *potpuno*, *posve, totalno, krajnje* and *maksimalno*[2]. This section discusses the relationship between maximizers and adjectives, as well as the phenomena of near-synonymous maximizers.

2.1. Maximizers and adjectives

Degree modifiers represent a subclass of degree words (Bolinger, 1972) that provide the specification of degree pertaining to the words they modify. Having considered different degree modifier classifications (Quirk et al., 1985; Allerton, 1987;

[1] Certain maximizers may also function as modifiers of nominal phrases (cf. *totalno* in *totalno ste 20. stoljeće* "you are totally 20th century" (Vidaković Erdeljić, 2023)).

[2] To keep the present analysis manageable and enhance the clarity of the upcoming data visualizations, the decision was made to concentrate on the five maximizers mentioned earlier, with the intention of exploring *krajnje* and *maksimalno* in a future study.

Paradis, 1997), it can be argued that degree modifiers divide into reinforcers and downtoners. Within the reinforcers category, we include maximizers (*potpuno* "completely", *totalno* "totally") and boosters (*vrlo* "very", *jako* "very"), while approximators (*skoro* "almost", *gotovo* "almost"), moderators (*prilično* "rather"), and diminishers (*neznatno* "slightly", *blago* "slightly") fall under the downtoners class.

The behaviour of maximizers is influenced by the scale structure of the adjectives they modify. It is well known that the scales used by gradable adjectives differ in whether they include a maximal and/or a minimum value (Rotstein and Winter, 2004; Kennedy and McNally, 2005; Kennedy, 2007). Accordingly, a typology of adjectives can be established based on their scale structure. For instance, Kennedy and McNally (2005) argue four types of scales: fully open ( ), lower-closed [), upper-closed ( ], and fully closed [ ]:

i. Open scale adjectives (0,1): *visok* "tall", *skup* "expensive";
ii. Lower closed scale adjectives [0,1): *prljav* "dirty", *mokar* "wet";
iii. Upper closed scale adjectives (0,1]: *čist* "clean", *suh* "dry";
iv. Closed scale adjectives [0,1]: *zatvoren* "closed", *pun* "full".

Given their inherent expression of an absolute degree, indicating the attainment of an endpoint, maximizers seamlessly combine with adjectives inherently associated with a boundary.

Considering solely its primary "maximizing reading"[3], maximizers, as exemplified in 7–11, exclusively co-occur with upper or totally closed-scale adjectives. In the absence of a maximum point, these expressions will be infelicitous.

(7) Boca je *potpuno* {?visoka/?prljava/čista/zatvorena}
("The bottle is completely {?tall/?dirty/clean/closed }") (applies to 7–11)
(8) Boca je *posve* {?visoka/?prljava/čista/zatvorena}
(9) Boca je *skroz* {?visoka/?prljava/čista/zatvorena}
(10) Boca je *sasvim* {?visoka/?prljava/čista/zatvorena}
(11) Boca je *totalno* {?visoka/?prljava/čista/zatvorena}

The present study's theoretical foundation encompasses Paradis' research (1997, 2001, 2005) on <degree modifier + adjective> pairings in English. Adopting a cognitive perspective, Paradis (1997: 26) emphasises that "in context, the use of degree modifiers is constrained by the semantic features of the collocating adjective on two dimensions: totality and scalarity". Central to her work is foregrounding of the theory of 'bidirectionality of semantic pressure', a model suggesting that the character of the adjective within any pairing of <degree modifier + adjective> dictates the type of modifier that can modify it. Simultaneously, the nature of the degree modifier influences the selection and interpretation of a compatible adjective component. Here "character" and "nature" refers to a way in which the gradeability is conceptualised. For instance, collocations with degree modifiers exhibiting a totality, "either-or" mode of construal, i.e. maximizers (*completely closed*, *totally open*), dictate that the adjective must have the same mode of construal, thus resulting in its interpretation as a totality (limit) adjective. Owing to the bidirectional relationship between modifiers and adjectives, the model claims that, concomitantly, the choice of a totality adjective reaffirms its modifier's totality mode of construal. Consequently, it becomes evident that the study of degree modifiers, and, thus, also maximizers, necessitates consideration of the context in which these modifiers appear, particularly their morphosyntactic environment, and an analysis of the co-occurrences of maximizers and adjectives within constructional patterns.

2.2. Maximizers as near-synonyms

Synonyms are commonly divided into two categories: absolute synonyms and near-synonyms. Absolute synonyms are exceedingly rare or even non-existent. In fact, Cruse (1986) observes that the synonymy relationship is inherently unstable. Over time, one of the synonyms may become dysfunctional and fall into disuse, or semantic and stylistic nuances may emerge, differentiating the words from each other. On the other hand, near-synonyms, "lexical items whose senses are identical in respect of 'central' semantic traits, but differ [...] in 'minor' or 'peripheral traits'" (Cruse, 1986: 237), are relatively common (consider pairs *big/huge*, *fog/mist*, *brave/courageous*). Near-synonyms are often not completely intersubstitutable (Inkpen and Hirst, 2006) and generally differ from various points of view (semantic, syntactic, pragmatic).

Degree modifiers, hence, also maximizers, are often seen as cognitive synonyms (Paradis, 1997) or near-synonyms. The use of maximizers is typically discretionary, and in most contexts, they can be substituted with other maximizers, as their content is essentially functional: *skroz*, *sasvim*, *posve*, *potpuno*, and *totalno* modify the qualities expressed by the lexical bases they moderate. Although substituting one maximizer for the other does not alter the proposition's truth value (as all five maximizers should theoretically express the same degree of moderation), it is expected to reveal that <maximizer + adjective> pairings are not wholly free (Paradis, 1997; Kennedy, 2003). In fact, a considerable part of the analysed <maximizer + adjective>

[3] For example, *totalno* can have non-degree readings (e.g. *Iskreno prvih dana sam bio totalno mrtav* [...](hrWaC) "Honestly the first few days I was totally dead [...]") and partitive readings (e.g. *Prale su totalno, ali totalno prljav wc* [...](hrWaC) "They were cleaning totally, totally dirty toilet [...]"; a reading could be "every part of the toilet is dirty", not "the toilet is dirtier than it can otherwise possibly be") (Kennedy and McNally, 2005).

constructions are presumed to be (semi)-prefabricated units[4], with maximizers imposing particular modes of construal (Paradis, 2008). As analysed maximizers are considered near-synonyms, it is conceivable that they would possess the same conceptual content, albeit construed in distinctive ways. These distinct collocation patterns may potentially reflect nuanced differences in schematic-domain profiling (Desagulier, 2014).

3. Method

This paper adopts the methodology proposed by Desagulier (2014, 2015) and combines analytical statistics with multifactorial methods. Analytical statistics involves the application of various statistical methods and techniques to analyse and interpret data, encompassing descriptive statistics, inferential statistics, regression analysis, correlation analysis, and more. This paper will employ two methods from a family of methods known as collostructional analysis (Stefanowitsch and Gries, 2003; Gries and Stefanowitsch, 2004a) to examine the distribution and collocational preferences of five analysed maximizers. The data obtained by simple collexeme analysis (SCA) and multiple distinctive collexeme analysis (MDCA) will act as input for two multifactorial analysis methods: hierarchical agglomerative cluster analysis (HACA) and correspondence analysis (CA)[5].

Considering that maximizers "occur frequently in non-academic writings such as informal texts, books and periodicals" (Xiao and Tao, 2007: 246), the corpus of choice was hrWaC 2.2 (Ljubešić and Klubička, 2014), a web corpus compiled from the .hr domain. Corpus was created in 2014 and consists of 1,211 billion words of written Croatian. hrWaC was examined via SketchEngine's (Kilgarriff et al., 2004) search interface.

Guided by one of the fundamental maxims of corpus linguistics, "a word is known by the company it keeps" (Firth, 1957), it is presumed that the context in which a variable (be it lexical or phrasal) appears provides valuable insights into its syntactic and semantic characteristics (Sinclair, 1991; Biber et al., 1999). A straightforward approach to analysing the context of a variable involves extracting its collocates and identifying those that frequently co-occur alongside the variable. Collocates, referring to words or phrases typically found close to each other within a text, are extracted following a decision regarding a specific range (span) that will be analysed. Nonetheless, the literature on intensifiers lacks definitive guidance regarding the ideal exploration span, as Desagulier (2014) indicated. For instance, Kennedy (2003: 473) examines a "window of two words" on both sides to capture collocations that intervening words might separate.

This study will focus on the prototypical contexts in which a degree modifier (maximizer) immediately precedes the adjective it modifies[6]. Additionally, relatively infrequent cases of the nominal copulative predicate in which the sequence <maximizer + adjective> is interrupted with a present or past form of the verb *biti* "be", e.g.

Totalno	*je*	"be-3SG.PRS."	različit.	Totalno	*je bio*	"be-3SG.PST."	različit.
Totally	is		different.	Totally	was		different.
("(S)he/It is totally different.")				("(S)he/It was totally different.")			

will likewise not be considered in order to facilitate and expedite the subsequent analyses. In fact, as pointed out by Desagulier (2014), embracing a constructional approach, i.e. treating the <modifier + adjective> sequence as a construction, and limiting the semantic investigation to the syntactic frame of the construction, helps minimize the risk of obtaining irrelevant data. However, it is noteworthy that for Croatian, owing to its specific syntax, this approach might potentially result in the omission of interesting data. Therefore, we encourage a more thorough investigation of "interrupted" constructions, which, due to practical constraints, were not explored in this study.

To further enhance the analysis and differentiate adjectives that exhibit a significant association with the analysed maximizers from those that frequently occur in the corpus irrespective of context, reliance solely on raw counts and basic relative frequencies will be avoided. Instead, a collostructional analysis approach, specifically SCA and MDCA, will be employed. The way these methods are designed allows us to account for the two-way semantic influences between maximizers and adjectives, effectively filtering out adjectives with a high overall token frequency in the corpus. SCA, as indicated by Desagulier (2014), will serve two primary purposes. First, it will assist in identifying the adjectives that exhibit the strongest attraction to the <maximizer + adjective> construction by quantifying the bidirectional relationship between a modifier and the modified adjective. Additionally, SCA will aid in confirming the functional synonymy of *posve*, *potpuno*, *skroz*, *sasvim*, and *totalno* by revealing an overlap in the selection of adjectives. The extent of this overlap indicates how closely these five maximizers align

[4] In fact, as indicated by Desagulier (2014), it is possible "that entrenched types of the <moderator + adjective> construction provide access to conceptual structures in ways that are not necessarily compositional".

[5] All analyses were carried out using R Statistical Software (v4.2.2; R Core Team 2022).

[6] A brief preliminary study was conducted to examine the syntactic behaviour of *posve*, *potpuno*, *skroz*, *sasvim*, and *totalno* as adjective modifiers. As expected, these five maximizers typically precede the modified adjective and <adjective + maximizer> pairings are highly infrequent and limited to particular contexts. In fact, a construction <adjective + maximizer> tends to occur exclusively with *skroz* and *totalno*. Given that these constructions deviate from the typical word order, possibly as a result of the incorporation of elements from spontaneous speech into writing, and exhibit an informal tone, it is unsurprising that they tend to co-occur with maximizers associated with lower formality. Furthermore, a diachronic study of Croatian degree modifiers should reveal if this observation suggests the possibility of an ongoing process of grammaticalization, which is not as advanced as that observed in other maximizers which are subject to more pronounced syntactic constraints.

with similar content domains, bolstering the argument for their classification as near-synonyms. However, acknowledging that near-synonyms also present nuanced differences, a complementary method called MDCA is employed to contrast the five near-synonymous constructions (<posve + adjective>, <potpuno + adjective>, <sasvim + adjective>, <skroz + adjective>, <totalno + adjective>). As specified by the adjective *distinctive*, MDCA, unlike SCA, filters out overlapping collocates and focuses solely on the adjectives that are idiosyncratic to each modifier. The ensuing analysis enables the classification of distinctive adjectives based on their function and meaning, thereby providing a deeper understanding of the individual functional specificities of the five maximizers.

Results of SCA and MDCA play pivotal roles in two usage-based techniques that aim to capture semantic relations among near-synonyms based on multiple factors: hierarchical agglomerative cluster analysis (HACA) and correspondence analysis (CA).

HACA is employed to test the existence of the unified maximizer paradigm within the broader paradigm of Croatian adjectival degree modifiers. The most attracted collexemes of 19 Croatian degree modifiers (see pp. 14–15 for an inventory of DMs), including five maximizers under examination, will serve as input for the analysis.

CA, another distance-based clustering technique, visually represents cross-tabulations on a two-dimensional plot, enabling the mapping of correlations between lexical items. In this study, CA demonstrates how modifiers imply specific construals based on the adjectives they are associated with. The input for CA is derived from the cross-tabulation of frequencies of the 30 most distinctive collexemes of each of the five maximizers obtained through the MDCA described earlier.

These analytical methods collectively contribute to a comprehensive exploration of the relationships and patterns among the five examined maximizers, thereby enriching our understanding of their semantic characteristics.

4. Results

Section 3 has illustrated the methodology used in this study. This section presents the results of simple collexeme analysis, hierarchical cluster analysis, multiple distinctive collexeme analysis and, finally, correspondence analysis.

4.1. Simple collexeme analysis

Given our consideration of maximizers *skroz*, *sasvim*, *posve*, *potpuno*, and *totalno* as near-synonyms, i.e. alternative ways of expressing that the extreme value on a scale has been reached, we anticipate discovering both significant similarities as well as differences between these modifiers.

Drawing upon the approaches commonly found in collexeme analysis, which frequently focus on the top attracted collexemes, this paper will examine the most attracted collexemes of five <maximizer + adjective> constructions (<posve + adjective>, <potpuno + adjective>, <sasvim + adjective>, <skroz + adjective>, and <totalno + adjective>) in hrWaC corpus.

To quantify the degree of association (attraction and/or repulsion) between a linguistic unit (such as a word or construction) and its collocates (i.e., words or constructions that co-occur with it), i.e. to calculate the collostruction strength of, in our case, a given adjective ADJ for a specific <maximizer + adjective> construction CONS, simple collexeme analysis (SCA) necessitates four frequencies: the raw frequency of ADJ in CONS, the raw frequency of ADJ in all other constructions except CONS, the frequency of CONS with adjectives other than ADJ and, finally, the frequency of all other constructions then CONS with that of all other adjectives then ADJ (Stefanowitsch and Gries, 2003). The frequencies are organised in a 2 × 2 contingency table, and the process is conducted for every adjective that occurs in <maximizer + adjective> construction in the corpus. Upon calculating the association measure, it becomes possible to rank all linguistic elements by their association with the examined construction.

Besides information contained in the four subcomponents of the input contingency table, two supplementary values are required for SCA: the size of the units of analysis (sometimes referred to as corpus size) and the general frequency of the examined constructions.

In our study, units of analysis (corpus size)[7] consist of 137,344,804 adjectival tokens, while the frequency of the examined constructions is as follows: frequency of <posve + adjective> = 53,287; frequency of <potpuno + adjective> = 153,496; frequency of <skroz + adjective> = 14,830; frequency of <sasvim + adjective> = 94,709; frequency of <totalno + adjective> = 25,952. In this study, the log-likelihood ratio[8] (G^2), "the most frequently used [association] measure" (Gries, 2019: 150), was the measure of choice. Table 1 indicates the top 10 most strongly attracted adjectives, i.e. collexemes of *posve*, *potpuno*, *skroz, sasvim,* and *totalno* based on their collostruction strength (log-likelihood ratio (G^2) value).

[7] Following the approach outlined by Proisl (2022), which emphasizes defining units of analysis not based on a class of constructions (Stefanowitsch and Gries, 2003), but rather as word class forms that meet the constraints of the target slot in the examined construction, the units of analysis were defined as all adjective tokens present in hrWaC.

[8] Due to space limitations, we will refrain from providing a detailed elaboration on the rationale for favouring one association measure over another. Advantages and disadvantages of some of the most used association measures are discussed, *inter alia*, in Gries (2019).

Table 1
Top 10 collexemes of posve, potpuno, skroz, sasvim, *and* totalno *in hrWaC.*

posve			***potpuno***		
adjective	**frequency in construction**	G^2	**adjective**	**frequency in construction**	G^2
jasan	2500	11789.74	*nov*	15,532	30770.16
drugačiji	1892	11155.40	*drugačiji*	5071	29330.85
drukčiji	1342	9613.77	*drukčiji*	1764	9907.56
nov	4612	7872.12	*različit*	3514	9213.97
siguran	1322	4261.34	*pogrešan*	1622	8451.11
normalan	1182	3971.87	*besplatan*	1949	7362.07
različit	1310	3597.15	*nepoznat*	1903	7359.73
razumljiv	578	3412.17	*jasan*	2538	6939.81
suprotan	577	2553.16	*suprotan*	1585	6892.05
nepotreban	433	2076.03	*nepotreban*	1339	6633.94
skroz			***sasvim***		
adjective	**frequency in construction**	G^2	**adjective**	**frequency in construction**	G^2
drugačiji	374	1945.86	*normalan*	6346	35155.33
normalan	407	1530.26	*drugačiji*	5102	34529.17
drukčiji	163	894.35	*dovoljan*	5036	30057.49
okej	104	765.84	*solidan*	2757	20072.73
mokar	105	589.93	*jasan*	3852	17104.28
lud	150	577.81	*drukčiji*	1913	12890.00
nebitan	100	553.61	*pristojan*	1649	11192.59
simpa	74	538.80	*nov*	7304	11067.66
cool	94	504.48	*logičan*	1611	9520.23
zadovoljan	154	417.83	*običan*	1685	6647.76
totalno					
adjective	***frequency in construction***	G^2			
nebitan	501	3827.50			
drugačiji	677	3569.47			
glup	537	2485.75			
lud	431	2072.74			
promašen	315	2018.83			
drukčiji	331	1914.44			
nepotreban	302	1658.65			
nesposoban	271	1657.70			
kriv	478	1567.09			
zbunjen	269	1534.50			

The degree of attraction between the construction and the collexeme (and vice versa[9]) is very significant at $p < 0.00001$. As anticipated, SCA reveals both similarities and differences in the collocational preferences of maximizers. The presence of overlap in the selection of collexemes indicates similarity in the maximizer paradigm: adjectives *drugačiji* "different" and *drukčiji* "different", synonyms derived from the numeral adjective *drugi* "second", appear in the top 10 collexemes of each maximizer, being the most distinctive of *sasvim* (coll. strength = 34529.17). Furthermore, while no adjectives co-occur with four maximizers, several adjectives, such as *normalan* "normal", *nov* "new", *nepotreban* "unnecessary", and *jasan* "clear", co-occur with three maximizers. That said, since the number of overlapping collexemes is rather high, discerning differences in the collocational behaviour of maximizers is challenging, and it is believed that a finer-grained analysis, including the classification of the most attracted collexemes in semantic classes at this stage, would not yield more insightful outcomes. Nevertheless, it is noteworthy that the most frequently denoted semantic category by the collexemes of all maximizers is that of *difference*:

i. posve: *drugačiji* "different", *drukčiji* "different", *različit* "different", *suprotan* "contrary";
ii. potpuno: *drugačiji*, *drukčiji*, *različit*, *suprotan*;
iii. skroz: *drugačiji, drukčiji*;
iv. sasvim: *drugačiji*, *drukčiji*;
v. totalno: *drugačiji*, *drukčiji*.

One possible way to filter away overlapping collexemes is through multiple distinctive collexeme analysis (MDCA) (Gries and Stefanowitsch, 2004a). Prior to conducting MDCA, consistent with Desagulier (2014), the internal coherence within the Croatian maximizer paradigm will be analysed by performing hierarchical agglomerative cluster analysis (HACA) with 19 Croatian degree modifiers as input. The primary objective of HACA is to investigate whether the five maximizers cluster together based on their shared preferred collexemes.

[9] It is worth emphasizing that the log-likelihood ratio (G^2) is a symmetrical association measure; hence, it assumes a bidirectional dependence between the construction and the collexeme, viz. the fact that the collexeme attracts the construction to the same degree as the construction attracts the collexeme.

4.2. Hierarchical agglomerative cluster analysis

Hierarchical agglomerative cluster analysis (HACA) encompasses a diverse array of multifactorial techniques designed to reveal underlying structures within data, particularly identifying clusters of similar objects based on their inter-object distances (Everitt et al., 2011). As described by Levshina (2015), HACA depicts analysed entities as either leaves or branches within a clustering tree, commonly referred to as a dendrogram. Unlike a conventional tree that extends from the root to the branches, a dendrogram grows in the opposite direction. Each object (in this case, a constructional profile vector) is initially treated as an individual cluster or "leaf". Subsequently, the most similar objects, i.e. those with the smallest inter-object distances, are progressively merged, resulting in a unified tree structure.

As noted in Gries and Stefanowitsch (2004b) and Stefanowitsch and Gries (2005), and consistent with Gries and Stefanowitsch (2010) and Desagulier (2014), this study employs HACA to investigate the grouping patterns of 19 Croatian degree modifiers of adjectives based on their preferred collexemes. The set[10] of degree modifiers of adjectives—excerpted from grammars (Barić et al., 1997; Silić and Pranjković, 2005), few relevant works on degree modifiers in Croatian (Pavić Pintarić and Frleta, 2014; Nigoević, 2020), and *hrWaC* corpus[11]—and including the five maximizers under analysis, consists of: *blago* "slightly", *donekle* "somewhat", *dosta* "rather, pretty", *izrazito* "extremely", *jako* "very", *malo* "a bit", *naročito* "especially, particularly", *osobito* "especially, particularly", *posve* "completely, totally", *potpuno* "completely, totally", *previše* "too + ADJ", *prilično* "rather, pretty", *sasvim* "completely, totally", *skroz* "completely, totally", *stvarno* "really", *totalno* "completely, totally", *veoma* "very", *vrlo* "very", and *zaista* "very"[12].

The objective of HACA is to determine how five examined maximizers cluster based on their preferred collexemes: if *posve*, *potpuno*, *skroz*, *sasvim*, and *totalno* are found to cluster together, it would suggest that they form a cohesive and consistent paradigm (Desagulier, 2014).

After formulating the query, the concordances were carefully reviewed to exclude examples where an adverb does not act as a degree modifier but serves as a quantifier of a noun following the adjective (e.g. *mnogo* in *mnogo lijepih ljudi* "a lot of pretty people"). Subsequently, for each of the aforementioned 19 degree modifiers, the 1000[13] most frequent adjectival collocates were extracted from the *hrWaC* corpus, and SCA was conducted. Due to the large corpus size, working with complete datasets was impractical, leading to the decision to conduct HACA using the 30 most attracted collexemes of each modifier. After their extraction and the cancellation of duplicate collexemes (e.g. ones appearing with two or more modifiers), a set of 290 adjective types was formed. Consequently, a 19-by-290 co-occurrence table containing the frequency of each adverb–adjective pair type was submitted to HACA.

The hypothesis of independence regarding the input data can be rejected ($\chi^2 = 3{,}926{,}118$;[14] df = NA; p-value = 4.998e–04). To assess the strength of the relationship between row variables (adjectives) and column variables (chosen degree modifiers), Cramér's V can be calculated. It is computed by taking the square root of the χ^2 statistic divided by the product of the sum of all observations and the number of columns minus one. In our case, $V = 0.5967$, which indicates a strong association between variables.

HACA converts the input contingency table into a distance object, namely a dissimilarity matrix (Table 2), to which a chosen amalgamation rule is applied. The distances will indicate the degree of (dis)similarity between constructions based on the proportions of contextual variable values found in the vectors. Smaller distances correspond to constructions with more similar vectors, while greater distances indicate more dissimilar vectors (Levshina, 2015; Desagulier, 2017).

Table 2
A dissimilarity matrix (sampled).

	osobito	**posve**	**potpuno**	**previše**	**prilično**
osobito	0.0000	279.1664	281.0000	263.4204	271.1566
posve	279.1664	0.0000	222.4293	281.0000	275.0213
potpuno	281.0000	222.4293	0.0000	281.0000	275.9040
previše	263.4204	281.0000	281.0000	0.0000	272.6793
prilično	271.1556	275.0213	275.9040	272.6793	0.0000

In this study, after converting the contingency table into a table of distances, it was decided to employ the Canberra distance measure as it is the best-suited one for dealing with many empty occurrences (frequencies of 0) that we encounter in the input data (Divjak and Gries, 2006; Desagulier, 2014, 2017; Gries, 2021). To produce a compact final cluster, clusters were amalgamated using Ward's method (Ward, 1963). The function used is hclust(). Finally, to validate the cluster, i.e. to examine how strongly the data support it, we applied multiscale bootstrap resampling and computed bootstrapping-based cluster significance values using pvclust() package. The resulting dendrogram is displayed in Fig. 1.

[10] Naturally, selected set of degree modifiers is not exhaustive and does not encompass all degree modifiers of adjectives in Croatian. Given the absence of existing works defining a paradigm of degree modifiers in Croatian (cf. Paradis, 1997 for the English paradigm), the modifiers were chosen at the author's discretion, aiming to represent different classes of degree modifiers.

[11] To identify modifiers of adjectives appearing in the typical context of <degree modifier + adjective>, a simple query ([tag = "R.*"][tag = "A.*"]) was utilised. The extracted modifiers were then arranged based on their frequency and subjected to analysis.

[12] *Stvarno* and *zaista* can also be interpreted as a modality modifier (with a reading of 'in truth') (Paradis, 1997).

[13] The SketchEngine platform, which hosts *hrWaC* corpus, poses a download limit of 1000 items from each list.

[14] Since cell counts are small, Pearson's Chi-squared test with Monte Carlo simulated p-value (based on 2000 replicates) was executed to get a p-value without assuming asymptotically normal behaviour.

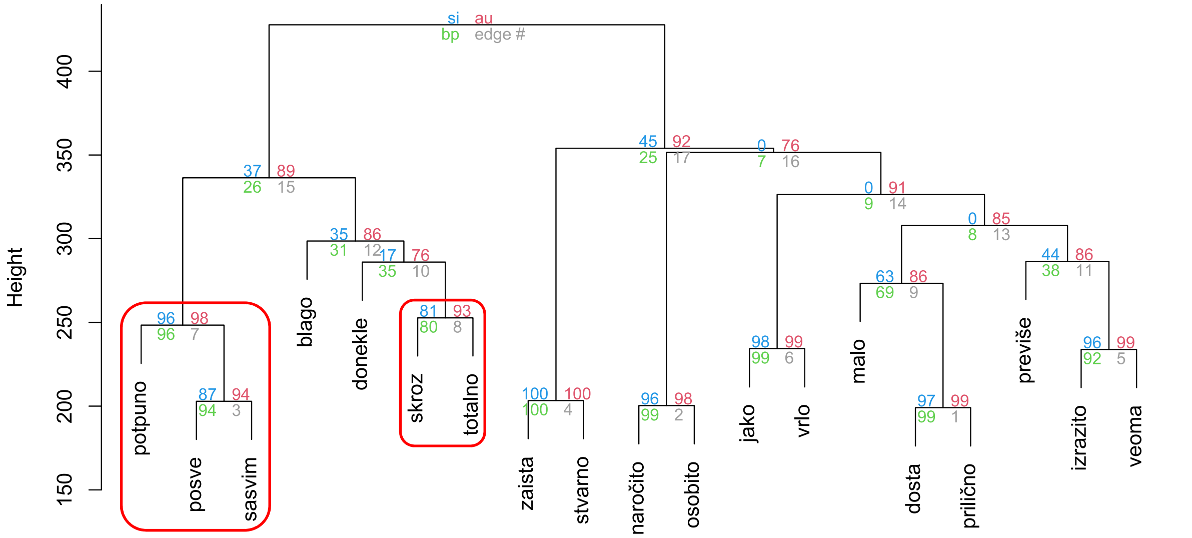

Fig. 1. *Cluster dendrogram of 19 degree modifiers of adjectives in Croatian, clustered according to their adjectival collexemes (distance: Canberra; cluster method: Ward.D).*

We can start inspecting the dendrogram from bottom to top. The closer the two lower elements are merged on the tree, the greater their similarity. Conversely, the higher the merge occurs, the greater the dissimilarity between the merged elements. We observe that each subcluster (node) is accompanied by four values: the value below the node on the right indicates its rank (in our case, from 1 to 17, as 17 clusters have been generated), while the remaining three numbers specify three types of probability values. The number on the top left represents the “Selective Inference” *p*-value (SI)[15], the one on the top right represents “Approximately Unbiased” *p*-value (AU), while the one on the lower left indicates the “Bootstrap Probability” value (BP). As indicated by multiple authors, (*inter alia* Desagulier, 2014; Levshina, 2015), as well as the package documentation (Suzuki and Shimodaira, 2006), the AU value is deemed a more exact measure than BP. That said, both measures share a common logic: the closer the value is to 100, the stronger the cluster. On the other hand, SI *p*-value acknowledges the consideration that clusters in the dendrogram are chosen based on data, which contradicts the fundamental premise of traditional statistical hypothesis testing, i.e. the fact that null hypotheses are typically selected before examining the data. For this motive, SI *p*-value is often preferred to AU and BP values in assessing the stability and robustness of clusters (Shimodaira and Terada, 2019).

In our case, when observing SI and AU values, it becomes evident that generated subclusters are rather strong, i.e. well supported by the data. It is also noteworthy that low SI values imply low BP values, while AU values remain rather high (cf. e.g. clusters 10, 13, 14 and 16). In such instances, it may be inferred that AU values could be “false positives”, i.e. they might not effectively reflect the clusters’ actual strength. This observation could challenge the notion put forth regarding AU’s superior precision compared to BP. With this in mind, it becomes evident that a comprehensive evaluation of the validity and reliability of clusters in HACA necessitates considering SI, AU, and BP values collectively to make well-informed decisions.

The resulting dendrogram displays a moderate level of homogeneity among Croatian adjective degree modifiers, with some notable discrepancies between the function of the modifiers and their clustering. Several observations can be made:

i. Cluster 15 brings together maximizers and diminishers. It breaks down into a cluster of sole maximizers (7) and a cluster of maximizers and diminishers (12). Cluster 7 groups together *potpuno*, *posve*, and *sasvim*, with *posve* and *sasvim* forming a separate subcluster (3), while cluster 12 groups together diminishers *blago* and *donekle* with maximizers *skroz* and *totalno* (forming a separate subcluster (8));
ii. Cluster 17 predominantly comprises boosters and breaks down into cluster 4 (*zaista* and *stvarno*), cluster 2 (*naročito* and *osobito*), cluster 6 (*jako* and *vrlo*), cluster 11 (*previše*, *izrazito*, and *veoma*), and, finally, cluster 9 (*malo*, *dosta*, and *prilično*). Clusters 2, 4, 6, and 11 bring together all boosters and confirm that the Croatian booster paradigm is relatively homogeneous. Cluster 4 (*zaista* and *stvarno*) with SI, AU and BP scores of 100 is particularly strong. Especially interesting is cluster 9 since it is composed of one diminisher (*malo*), and it further breaks down into cluster 1, containing two

[15] Following Shimodaira (2019), four versions of SI values were examined: the default pvclust()result; recomputation via scaleboot compatible to pvclust(); linear model ($k = 2$); quadratic model ($k = 3$). While the $k = 3$ is expected to exhibit lower bias compared to $k = 2$, it tends to have a higher level of variance. As endeavours to implement scaleboot() with a wider range of scales for the purpose of decreasing *p*-value variance, while keeping the distance measure and clustering method unchanged, proved to be problematic, we opted to retain the default pvclust()results as they were deemed sufficiently reliable.

moderators (*dosta* and *prilično*). The reason for this clustering is not immediately evident. Considering that the other two diminishers (*blago* and *donekle*) are indicated as members of cluster 15, it is intriguing why diminisher *malo* collocates in cluster 17. However, these questions merit a whole other discussion that exceeds the limits of this study.

Particular attention should be paid to the clustering profiles of the five maximizers, denoted in red rectangles. As mentioned earlier, the maximizers in question do not form a single unified cluster, indicating that the Croatian maximizer paradigm is not entirely homogenous. Cluster 7 brings together *potpuno*, *posve*, and *sasvim*, with *posve* and *sasvim* forming a special subcluster (cluster 3), while cluster 8 brings together *skroz* and *totalno*[16]. It is essential to indicate that the SI values of the two clusters are rather high and range from 81 for cluster 8 to 96 for cluster 7.

One possible explanation for the clustering split in the maximizer paradigm lies in the formality of the modifiers: *potpuno*, *posve,* and *sasvim* can be considered more formal degree modifiers, whereas *skroz* and (especially) *totalno* mainly appear in a lower, more informal register. As the subsequent analysis will reveal (Section 4.3.), distinct collexemes of *skroz* and *totalno* differ from those of *potpuno*, *posve,* and *sasvim* regarding their formality level. This is in line with the use of *totally* in English (Bordet, 2017: 11) which "tend to co-occur with adjectives [...] belonging to colloquial language (*cool*, *awesome*, *hot*, *lame*, *rad*, *psyched*, etc.) and denoting more intense feelings or judgements".

Following the methodology proposed by Desagulier (2014), the multiple distinctive collexeme analysis (MDCA) will be employed to further investigate the differences among the maximizers. This method will allow for a more detailed examination and a better understanding of the individual characteristics of each maximizer.

4.3. Multiple distinctive collexeme analysis

(Multiple) Distinctive collexeme analysis (Gries and Stefanowitsch, 2004a; Stefanowitsch, 2013) contrasts two (DCA) or more (MDCA) constructions in their respective synchronic collocational preferences. While the compared constructions may be completely unrelated, MDCA is particularly suited for the study of related, near-synonymous constructions. Unlike SCA, MDCA allows one to identify collexemes that are idiosyncratic to specific constructions while going beyond the raw frequency and abstracting away from shared common elements. By focusing on usage-based, pattern-specific properties through systematic statistical investigation (Stefanowitsch and Flach, 2020), MDCA determines whether there are asymmetries in the relative frequencies of the co-occurring lexical elements and identifies collexemes that occur significantly more frequently with one construction compared to the other, ranking them based on their degree of distinctiveness (Hilpert, 2006). By emphasising these elements, MDCA sheds light on the subtle semantic and functional distinctions between constructions that may initially appear (near-)synonymous and that would be much more challenging to identify using the more traditional approaches.

The input for MDCA differs from one for SCA. Instead of quantifying the attraction between a lexical item and a construction, MDCA contrasts three or more near-synonymous constructions in their respective collocational preferences. As indicated by Flach (2021), input is a data frame and can be formatted either as a raw frequency list (one observation per line, with collexeme 1 in column 1 and collexeme 2 in column 2) or as an aggregated frequency list which must contain a third column with the frequency of the construction. The MDCA script used in this study is based on Flach's (2021) collex.covar() function, which can handle more than two constructions.

Tables 3–7 present the ten most distinctive collexems for each of the five analysed maximizers.

Table 3
The 10 most distinctive adjectives of posve

< posve + adjective >	**observed frequency**	**expected frequency**	**coll. str. log-likelihood**	***ΔP*1**	***ΔP*2**
jasan	2500	1410.7	875.82	0.02688	0.12389
drukčiji	1342	807.1	370.43	0.01320	0.10499
siguran	1322	867.5	255.50	0.01215	0.08310
neobičan	227	92	180.68	0.00333	0.22879
razumljiv	578	335.8	178.62	0.00598	0.11314
neuobičajen	156	57.3	150.60	0.00244	0.26849
različit	1310	963.6	138.35	0.00855	0.05713
točan	311	188.3	82.45	0.00303	0.10189
neočekivan	161	82.1	74.16	0.00195	0.15001
originalan	138	66.5	73.84	0.00176	0.16767

[16] The high correlation among the analysed maximizers is further supported by the Spearman's ρ correlation coefficient. The correlation coefficient ranges from -1 to $+1$, indicating a perfect correlation when it equals -1 or $+1$. In our study, we observed an almost perfect correlation between *potpuno* and *posve* ($\rho = 0.9380$), and a similarly strong correlation between *sasvim* and *posve* (0.8824). Additionally, *skroz* showed the highest correlation with *totalno* (0.8246), while *totalno* exhibited the highest correlation with *posve* (0.8985).

Table 4
The 10 most distinctive adjectives of potpuno

< potpuno + adjective >	observed frequency	expected frequency	coll. str. log-likelihood	$\Delta P1$	$\Delta P2$
isti	3908	2161.7	2661.32	0.02297	0.36281
opremljen	1344	608.6	1928.98	0.00967	0.53646
besplatan	1949	1012.6	1662.29	0.01239	0.41181
nov	15,532	12470.2	1471.13	0.04028	0.11945
uništen	1373	682.4	1373.08	0.00908	0.44953
spreman	1658	941.0	1013.70	0.00943	0.33914
gol	1089	619.6	657.55	0.00617	0.33631
ispravan	984	552.5	625.59	0.00568	0.34664
zdrav	1564	982.5	625.00	0.00765	0.26349
obnovljen	555	271.4	587.68	0.00373	0.46286

Table 5
The 10 most distinctive adjectives of skroz

< skroz + adjective >	observed frequency	expected frequency	coll. str. log-likelihood	$\Delta P1$	$\Delta P2$
dobar	546	117.6	906.39	0.03506	0.15209
mali	329	65.0	594.28	0.02160	0.16892
čudan	110	9.7	387.36	0.00821	0.42705
okej	104	10.9	323.71	0.00762	0.35438
simpa	74	4.9	319.68	0.00566	0.58598
simpatičan	76	6.0	289.74	0.00573	0.48301
crn	158	33.5	263.17	0.01018	0.15387
desni	49	2.4	259.50	0.00381	0.78930
loš	85	9.9	243.09	0.00614	0.31305
kratak	83	10.2	227.72	0.00596	0.29491

Table 6
The *10 most* distinctive adjectives of sasvim

< sasvim + adjective >	observed frequency	expected frequency	coll. str. log-likelihood	$\Delta P1$	$\Delta P2$
dovoljan	5036	1545.0	10754.53	0.05474	0.67232
drugi	5175	1966.8	6702.32	0.05030	0.48770
normalan	6346	2712.8	6275.48	0.05697	0.40383
solidan	2757	842.9	5894.92	0.03001	0.67044
pristojan	1649	508.3	3434.52	0.01789	0.66006
običan	1685	616.8	2337.88	0.01675	0.50999
dobar	1982	831.4	2003.27	0.01804	0.40846
logičan	1611	700.7	1486.67	0.01427	0.38289
konkretan	674	233.1	1059.81	0.00691	0.55465
mali	1066	459.4	1003.36	0.00951	0.38807

Table 7
The *10 most* distinctive adjectives of totalno

< totalno + adjective >	observed frequency	expected frequency	coll. str. log-likelihood	$\Delta P1$	$\Delta P2$
drugačiji	677	45.4	3650.96	0.03208	0.93358
glup	537	55.0	1909.48	0.02342	0.60184
promašen	315	27.5	1285.84	0.01462	0.71604
različit	390	53.0	1078.77	0.01714	0.43650
lud	431	65.6	1074.47	0.01859	0.38261
cool	238	27.5	759.90	0.01071	0.52573
loš	137	9.7	692.93	0.00647	0.89673
zbunjen	269	45.4	605.59	0.01137	0.33799
čudan	154	16.0	539.59	0.00702	0.59292
debilan	129	11.2	527.40	0.00599	0.71849

Tables 3–7 present additional information, precisely the ΔP value (Ellis, 2007; Ellis and Ferreira-Junior, 2009), which addresses certain limitations of G^2. Unlike G^2, ΔP (i) is asymmetric and does not conflate p(word2|word1) and p(word1|word2) into just one value, allowing it to distinguish cases in which collexeme 1 strongly attracts collexeme 2, but the collexeme 2 does not strongly attract collexeme 1 or vice versa; (ii) reflects association (+effect), but it does not reflect frequency (+effect −freq.) meaning that the change in the size of the corpus does not affect the association value. ΔP divides into two distinct values: $\Delta P1$ assesses the predictive capacity of the collexeme (slot 2) for the construction (slot 1), whereas $\Delta P2$ quantifies the predictiveness of the construction (slot 1) for the collexeme (slot 2) (Gries and Ellis, 2015; Gries, 2019). As

anticipated, when examining <maximizer + adjective> constructions, it is evident that the constructions are more predictive of the adjectives than vice versa ($\Delta P2$ values are significantly higher than $\Delta P1$). Among the five examined maximizers, *totalno* stands out as the most predictive when considering its top 10 most distinctive collexemes, with a mean $\Delta P2$ of 0.61424. The highest level of predictiveness is witnessed in the case of *totalno drugačiji*, where the maximizer *totalno* almost impeccably ($\Delta P2 = 0.93358$) anticipates its collexeme.

With overlapping collexemes (adjectives highly correlated with two or more analysed modifiers) filtered out by MDCA, the collocational behaviour tendencies of the five maximizers become more apparent.

Upon careful examination of the distinctive collexemes from Tables 3–7, it becomes evident that *totalno* displays the most semantically intriguing collocational profile, primarily attracting negatively connotated collexemes (*glup*, *lud*, *čudan*, *debilan*) and those indicating difference (*drugačiji*, *različit*). Notably, the only adjective that does not fall into these two categories, as seen in Table 7, is *cool*[17]. To delve deeper into this observation, a manual examination of the 200 most distinctive adjectives for *totalno* was conducted to determine their polarity (positive—neutral—negative connotation). Fig. 2 displays the results.

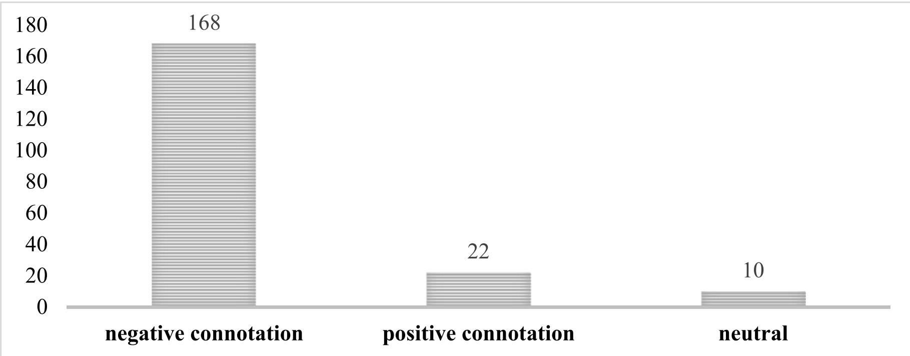

Fig. 2. *Histogram representing the connotation of the top 200 distinctive collexemes of* totalno.

The analysis revealed that out of the 200 adjectives, 168 (84 %) were negatively connotated, while the remaining 32 were divided into 22 positively connotated adjectives (e.g. *simpatičan* "nice", *zabavan* "funny", *trendi* "trendy") and 10 neutral adjectives (e.g. *različit* "different", *dječji* "children's"). While a more thorough examination is necessary, it is beyond the scope of the present study, and therefore, we will not be conducting it at this time.

While SCA and MDCA have provided valuable insights, it is essential to recognize their limitations stemming from the limited number of selected collexemes. On the other hand, as Desagulier (2014) pointed out, increasing the number of collexemes might complicate drawing meaningful generalizations from the data. Accordingly, instead of relying solely on a comparison of SCA and MDCA output frequency tables, it is advisable to adopt a technique that enables us to visualise the data regarding the attraction between (a) maximizers, (b) adjectives, and (c) maximizers *and* adjectives by converting the initial matrix into a low-dimensional space. This approach should enable us to understand these three elements' interplay better. In line with Desagulier's (2014, 2015) approach, this study will utilise the output of MDCA as input for correspondence analysis (CA).

4.4. Correspondence analysis

Simple Correspondence Analysis, commonly referred to just as Correspondence Analysis (henceforth CA), is a multifactorial exploratory statistical technique utilised for exploring relationships and patterns within categorical data (Benzécri, 1973; Greenacre, 2017). Its primary function is to transform the original matrix, viz. a contingency table, into a lower-dimensional space, typically a two-dimensional plot. By visually representing the data, CA allows for the identification of associations, clusters, patterns, and trends in complex categorical data (Desagulier, 2014, 2015, 2017). In CA, each row and each column are represented as a point in Euclidean space, and the proximity between points indicates the strength of association between the respective categories. The difference between profiles is measured using the χ^2-distance, a measure similar to the Euclidean distance but weighted by the inverse of the corresponding value in the average row profile, ensuring that rows deviating strongly from the average profile are positioned farther from other rows. The same principle applies to columns, where labels are located close if they exhibit similar proportions of counts in each row, indicating similar profiles. However, caution must be exercised when interpreting the mutual proximity of rows and columns, as "there is no direct interpretation of row-to-column or column-to-row distances" (Levshina, 2015: 371).

[17] A predilection of *totalno* for collocations predominantly involving (non-adapted) Anglicisms, particularly adjective *cool* or its adapted form *kul*, was confirmed also by Vidaković Erdeljić (2023) who has conducted a manual inspection of 500 occurrences of *totalno* in the Tweet-hr corpus. In author's examination, 53 examples of collocations with words borrowed from English were identified.

Table 8
Input for CA (sampled).

MAXIMIZER *DISTINCTIVE COLLEXEME*	posve	potpuno	skroz	sasvim	totalno
dovoljan	147	68	32	5036	0
normalan	1182	1278	407	6346	63
solidan	70	0	55	2757	0
automatski	10	210	0	0	0
glup	29	105	72	60	537
seksi	0	0	4	0	48
ljudski	94	266	3	90	4
prazan	115	972	119	58	90
čudan	235	32	110	55	154
simpatičan	8	0	76	61	0

Table 8 displays a sample of the input used for CA. In line with an approach described by Desagulier (2014), CA was conducted using the 30 most distinctive collexemes of each of the five maximizers and the raw frequency of each construction as input. The hypothesis of independence concerning the input table can be rejected (χ^2 = 144,558;[18] df = NA; *p*-value = 4.998e–04). This rejection confirms the interdependence between the choice of a maximizer and the choice of the adjective, aligning with Paradis' (1997) theory of the bidirectionality of semantic pressure. In addition, Cramér's *V* amounts to 0.4450. The value of $V = 0.4450$ indicates a significant association between the rows and the columns, supporting the notion of a meaningful relationship between adjectives and maximizers.

CA uses the input frequencies to juxtapose (a) line profiles, i.e. distinctive collexemes (adjectives); (b) column profiles, i.e. maximizers; (c) line profiles and column profiles, i.e. adjectives and maximizers. Even though the input table is rather small-scaled, attempting to analyse it with the naked eye could be challenging, and “any tendency we infer from raw frequencies may be flawed” (Desagulier, 2017). Therefore, it is highly recommended to use CA.

The CA() function from the FactoMineR package was employed to run CA. Figs. 3 and 4 (version with collexemes in English) display the biplot output of CA.

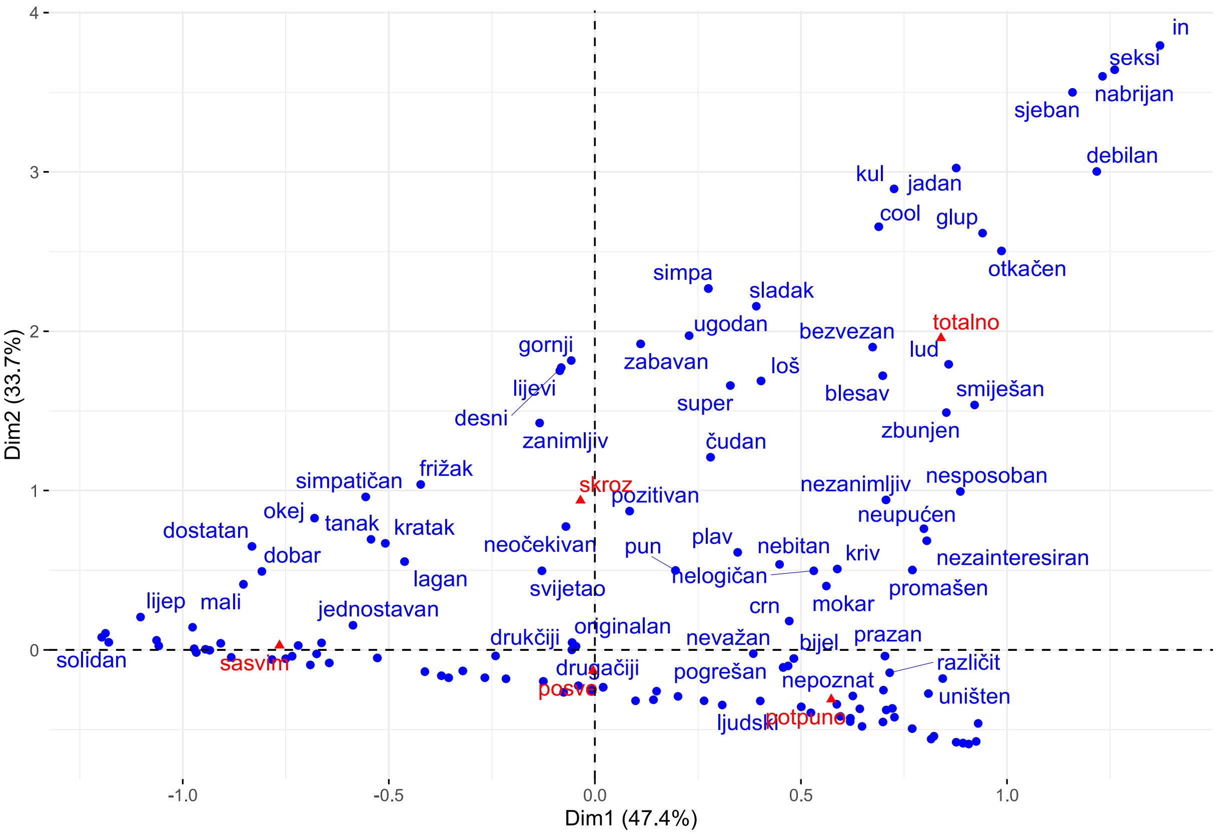

Fig. 3. *CA biplot of the <maximizer + adjective> construction in hrWaC*

[18] Pearson's Chi-squared test with simulated *p*-value (based on 2000 replicates).

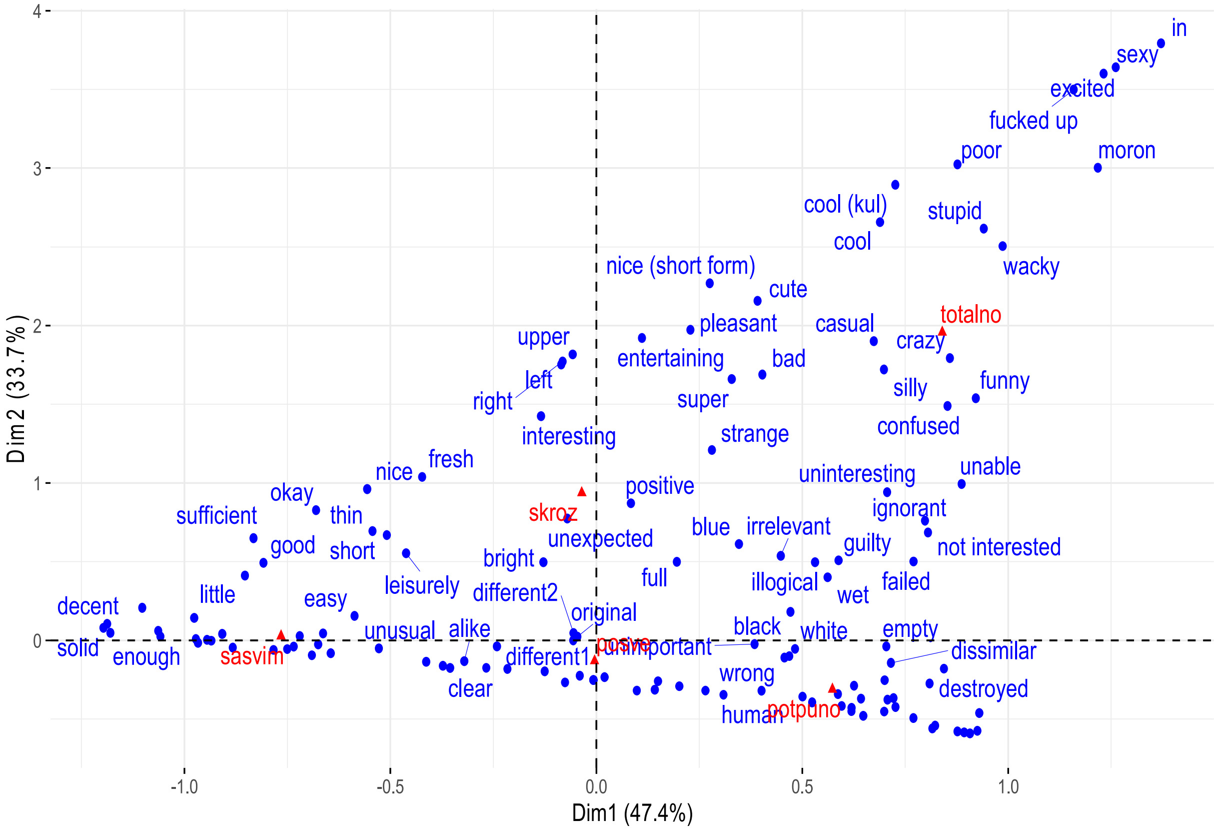

Fig. 4. *CA biplot of the <maximizer + adjective> construction in hrWaC [translated].*

Let us examine how the plot is built. In CA, the plot is constructed using two principal axes of inertia, which intersect to define the average profile of all points in the data cloud. The technique decomposes the overall inertia (Φ^2)—obtained by dividing the χ^2 statistic by the total sample size—by identifying representative dimensions that condense as much information as possible with each axis corresponding to a dimension. Typically, a plot displays only two dimensions, selected based on their eigenvalues, which measure the amount of information (variation) present along each axis (Levshina, 2015; Greenacre, 2017). In this analysis, the first axis (dimension 1) represents 47.45 % of the Φ^2, while the second axis (dimension 2) represents 33.68 % of the Φ^2. Although there are third and fourth dimensions with eigenvalues of 13.26 % and 5.62 %, they are not included in the plot. Whilst including additional dimensions can provide a more comprehensive understanding of the relationships between the analysed variables, the first two dimensions already account for 81.13 % of the variation contained in the input table, allowing for a sufficiently accurate interpretation of the results. To retain the clarity of the plot and facilitate meaningful interpretations, 64 points remained unlabelled.

We can start analysing the plot and how it juxtaposes five maximizers by contrasting the two main dimensions. Along the horizontal axis, dimension 1 contrasts *potpuno* to *sasvim* while *posve* is located on the vertical axis. *Posve*, thus, along with *skroz*, which also collocates itself on the vertical axis, do not significantly contribute to either dimension, indicating that these two maximizers are relatively indifferent to the characteristics of the adjective they modify. On the other hand, on the vertical axis, dimension 2 opposes *totalno* and *skroz* which are located above the horizontal axis, i.e. at the higher part of the plot, to *potpuno*, *sasvim*, and *posve* which can be found below the horizontal axis, i.e. at the lower part of the cloud. This is consistent with results obtained by the HACA (Fig. 1), where five modifiers form two separate clusters, one containing *potpuno*, *posve*, and *sasvim*, and another composed of *totalno* and *skroz*. However, a more in-depth interpretation of the plot concerning the division of labour among maximizers is rather tricky to spot since the cloud is very granular.

A possible way to lower the plot's granularity and try to explain the modifiers' specificities and division of labour among them is to annotate the adjectives for semantic classes. Deriving the annotation scheme for specific semantic annotation tasks can be expedited by utilising external sources, e.g. the database WordNet (Miller, 1995). However, in the case of the Croatian Wordnet (CroWN) (Raffaelli et al., 2008), the classification of adjectives into semantic classes and domains based on the relations expressed in the database is not provided. Consequently, the annotation process had to be performed manually. Following the classification proposed by Hundsnurscher and Splett (1982), adjectives were split up into 15 semantic classes, and, in line with GermaNet (Hamp and Feldweg, 1997), a special class for pertainyms was added.[19] Furthermore, each semantic class was broken down into several subclasses. In the end, annotated adjectives were categorized into 39 semantic classes. Table 9 indicates semantic classes and subclasses of adjectives used in the analysis.

Table 9
Semantic classes of adjectives.

BEHAVIOR-RELATED	*Behaviour*
	Character
	Inclination
	Sympathy
BODY-RELATED	*Affliction*
	Appearance
GENERAL	*Evaluation*
	Norm
MATERIAL-RELATED	*Dampness*
	Ripeness
	Stability
	State
MOOD-RELATED	*Feeling*
PERCEPTIONAL	*Colour*
	Lightness
	Taste
PERTAINYMS	
QUANTITY-RELATED	*Costs*
RELATIONAL	*Accuracy*
	Certainty
	Completeness
	Correspondence
	Difficulty
	Functioning
	Linking
	Privative
	Reference
	Requirements
	Security
	Validity
SPATIAL	*Dimension*
	Direction
	Existence
	Localization
	Spatial distribution
SPIRIT-RELATED	*Experience*
	Intelligence
	Knowledge
TEMPORALITY-RELATED	*Age*

[19] This decision is not exempt from several well-known methodological and theoretical concerns. First, the manual annotation was carried out by the author, making it inherently subjective as it relies on the semantic intuitions of a single individual. Second, categorizing each adjective into only one semantic class implicitly negates polysemy, the capacity of each adjective to carry multiple meanings. From the usage-based research paradigm adopted in this work, it is evident that meaning should not be treated as a fixed and stable category but as a dynamic one since it is in flux and consistently distributed in the co-occurring contexts of the linguistic unit (Bybee, 2007; Diessel, 2019). While a classification like this one can hardly reflect the complexity of the semantic space (Perek, 2016), it should surely be able to shed light on some specificities of each modifier.

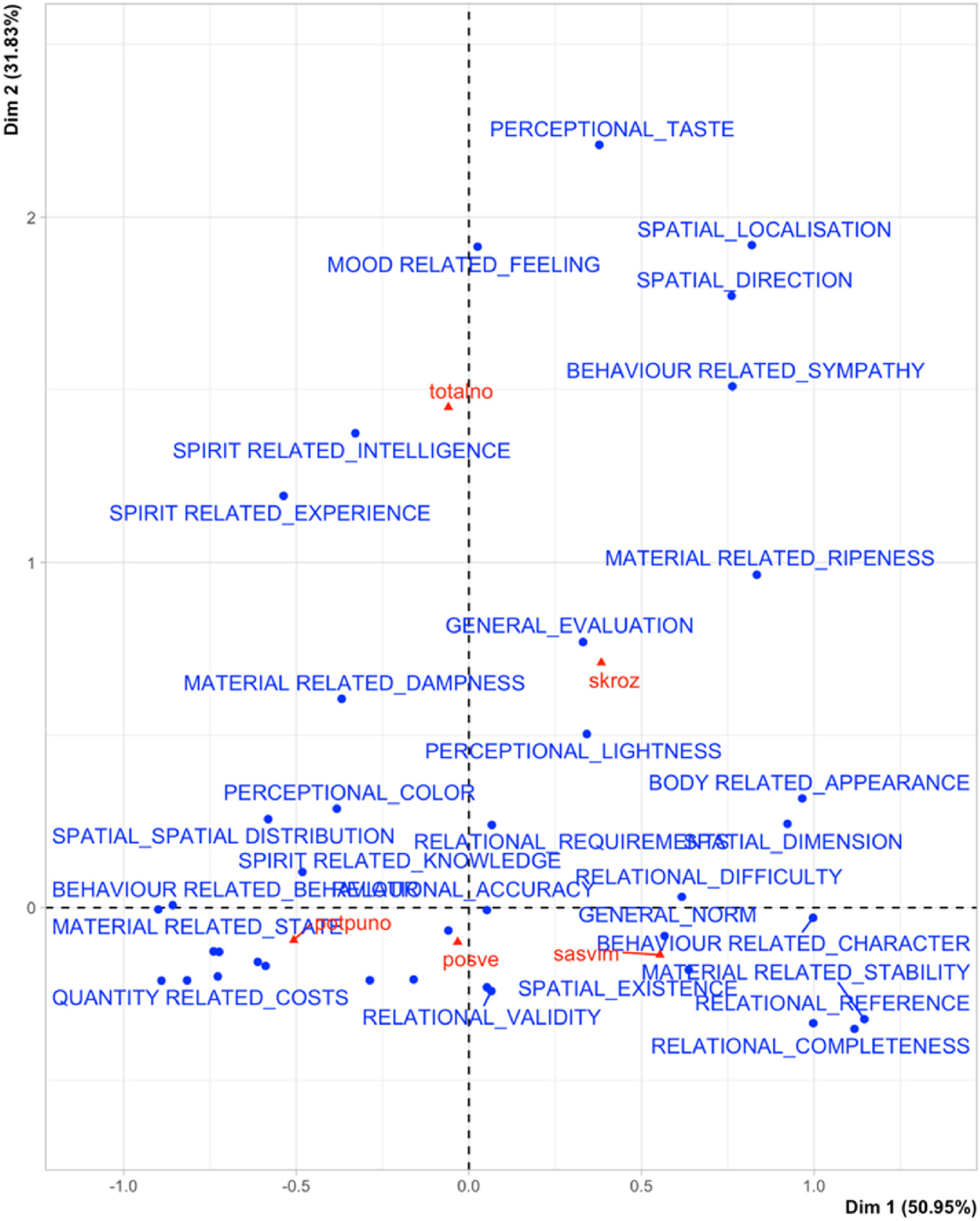

Fig. 5. *CA biplot of the <maximizer + adjective> construction in hrWaC including semantic annotation.*

Fig. 5 represents a CA plot of <maximizer + adjective> construction including semantic annotation. In this biplot, the first axis (dimension 1) represents 50.95 % of the Φ^2, while the second axis (dimension 2) represents 31.83 % of the Φ^2, accounting for 82.78 % of the variation present in the input table. The third and fourth dimensions, with eigenvalues of 12.46 % and 4.76 %, are not included in the plot. Upon observing Fig. 5 biplot, it is apparent that the relative position of maximizers has somewhat changed with respect to Figs. 3 and 4. Nevertheless, the central division identified in the first biplot, where dimension 2 opposes *totalno* and *skroz* to *potpuno*, *sasvim,* and *posve*, remains evident.

The analysis has revealed the specificities of each maximizer regarding the semantic class of adjectives it attracts. Since the semantic annotation is rather detailed, drawing generalizations from the data is challenging. However, we can still notice that:

i. *totalno* attracts adjectives referring to intelligence, more precisely to stupidity (*glup* "stupid", *debilan* "moronic", *blesav* "silly"), and psychological states (*sjeban* "fucked up", *nabrijan* "pumped", *jadan* "pathetic");
ii. *skroz* attracts adjectives expressing positive or negative evaluation (*dobar* "good", *loš* "bad", *bezvezan* "unexciting, dreary") and spatiality (*gornji* "upper", *desni*, "right", *lijevi* "left");
iii. *potpuno* felicitously combines with adjectives indicating temporality (*nov* "new", *suvremen* "contemporary"), certainty (*određen* "determined", *jednoznačan* "unambiguous"), and linking (*neovisan* "independent", *slobodan* "'free", *odvojen* "separated");
iv. *sasvim* accompanies adjectives denoting conformity or deviation from the norm (*običan* "regular", *prosječan* "average", *drugačiji* "different", *čudan* "strange, odd") and simplicity (*jednostavan* "simple");
v. *posve* is usually related with adjectives signaling accuracy (*preopćenit* "overly general", *pogrešan* "incorrect", točan "correct"), resemblance (*sličan* "similar", *nalik* "alike", *različit* "different", *isti* "same"), and certainty (*izvjestan* "certain, indubitable", *izgledan* "certain, indubitable").

Considering the granularity of the division in semantic classes, it was decided not to include an additional layer of annotation regarding the connotation of the adjectives (positive vs. negative). Although the decision makes it challenging to observe a division of labour in intensifying particular complementary meanings (e.g. positive vs. negative attitude), it also has some advantages. For example, it is possible to notice how:

i. there is a division of labour inside a semantic class of spatiality—*skroz* modifies direction (*desni* "right", *lijevi* "left") and localization (*gornji* "upper"), whereas *sasvim* modifies dimension (*mali* "small", *kratak* "short") and existence (*realan* "real");
ii. there is a division of labour inside a semantic class of body-related adjectives—*sasvim* modifies appearance (*lijep* "beautiful", *seksi* "sexy"), whereas *potpuno* modifies affliction (*zdrav* "healthy").

Finally, particular meanings do not show distinctiveness with any specific maximizer, e.g. adjectives denoting requirements (*zadovoljavajuć* "satisfactory", *prihvatljiv* "acceptable", *dostatan* "sufficient") are very productive with *sasvim*, *potpuno* and *posve* and slightly less productive with *skroz* and *totalno*.

5. Conclusions

This cognitive, usage-based approach research has proposed and integrated several statistical methods to analyse similarities and differences among five near-synonymous <maximizer + adjective> constructions, revealing that the combinations of maximizers and adjectives are not entirely free.

Several points can be made. First, from a methodological point of view, the significance of employing statistics in corpus-based analyses, often lacking in studies of Croatian collocations, was asserted. Collostructional analysis has been favoured as the preferred approach over raw counts and percentage-based methods, as it effectively filters out co-occurring pairs that may exhibit irrationally high or low frequencies, irrespective of corpus size, allowing for a more realistic interpretation of the results. Second, incorporating univariate and multivariate statistics in line with Desagulier (2014, 2015) enabled a better identification of usage patterns and conceptual structures among a set of near-synonyms. Lastly, the obtained results partially support Paradis' (1997) perspective on the cognitive synonymy of English modifiers, showcasing both similarities and differences. While Croatian maximizers share a fundamental functional basis in modifying the degree of an adjective's property to a maximum value, they, as shown, may not always modify the same classes of adjectives and can function within distinct conceptual domains. The degree of entrenchment varies among constructions. Tables 3–7 illustrate the top 10 distinctive collexemes of each maximizer, while correspondence analysis goes a step further. Besides depicting entrenched collocations (e.g. *posve siguran*, *potpuno nov*, *sasvim dovoljan*), viz. ones that, through repeated exposure, become mentally encoded and established as a cognitive routine (Divjak, 2019), CA also represents collocations that are possible but improbable (e.g. *sasvim debilan*, *potpuno okej*, *posve kul*), reflecting in that way the fact that speakers tend to use certain adjectives with certain degree modifiers but can also extend modifiers idiosyncratically to other classes of adjectives. The existence of denser clusters of adjectives around *potpuno* and *sasvim* could be interpreted as a sign that the use of these maximizers is more conservative, i.e. that the set of adjectives that collocate with them is more closed in respect to that of the other three maximizers in the examination.

Due to space limitations and the lack of existing works which would facilitate the study, several aspects of <maximizer + adjective> construction were left to be analysed in the future, e.g. the syntactic behaviour of degree modifiers (alternations) and their grading force (Paradis, 1997).

Despite all the mentioned limitations and the need for further experimental validation of numerous aspects of degree modifier use in Croatian, the adopted approach is believed to bring novel insights into the study of maximizers. Additionally, it contributes to the understanding of constructions and collocations of the Croatian language, an area for which relevant quantitative studies have yet to be undertaken. In that sense, the methodology presented in this paper has the potential to be extended to explore other intensifiers in Croatian and not only as it can be applied to investigate linguistic paradigms in general, especially for studies adhering to (Cognitive) Construction Grammar theoretical framework. Finally, the presented findings not only enhance our understanding of the cognitive processes that guide users' choice of construction but can be instrumentalised in Applied linguistics, particularly for Croatian FL purposes, as identification of distinctive collexemes of each maximizer can serve as valuable information for language teaching planning.

Declaration of competing interest

None.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Data availability

Data will be made available on request.

References

Allerton, D.J., 1987. English intensifiers and their idiosyncrasies. In: Steele, R., Threadgold, T. (Eds.), Language Topics: Essays in Honour of Michael Halliday, vol. 2. John Benjamins, pp. 15–31.
Barić, E., Lončarić, M., Malić, D., Pavešić, S., Peti, M., Zečević, V., Znika, M., 1997. *Hrvatska gramatika* [Croatian Grammar]. Školska knjiga.
Batinić, M., Kresić, M., Pavić Pintarić, A., 2015. The intensifying function of modal particles and modal elements in a cross-linguistic perspective. Rasprave 41 (1), 1–27.
Benzécri, J.-P., 1973. *L'analyse des données. 2. L'analyse des correspondances* [Data Analysis. 2. Correspondence Analysis]. Bordas.
Biber, D., Johansson, S., Leech, G., Conrad, S., Finegan, E., 1999. Longman Grammar of Spoken and Written English. Pearson.
Bolinger, D., 1972. Degree Words. De Gruyter Mouton.
Bordet, L., 2017. From vogue words to lexicalized intensifying words: the renewal and recycling of intensifiers in English. A case-study of *very, really, so* and *totally*. Lexis J. English Lexicol. 10 (1), 1–16.
Bybee, J., 2007. Frequency of Use and the Organization of Language. Oxford University Press.
Cruse, A.D., 1986. Lexical Semantics. Cambridge University Press.
Desagulier, G., 2014. Visualizing distances in a set of near-synonyms: *rather, quite, fairly*, and *pretty*. In: Glynn, D., Robinson, J.A. (Eds.), Corpus Methods for Semantics: Quantitative Studies in Polysemy and Synonymy. John Benjamins, pp. 145–178.
Desagulier, G., 2015. Forms and Meanings of Intensification: A Multifactorial Comparison of *Quite* and *rather*. Anglophonia -. Fr. J. Engl. Linguist. 20. https://doi.org/10.4000/anglophonia.558.
Desagulier, G., 2017. Corpus Linguistics and Statistics with R. Introduction to Quantitative Methods in Linguistics. Springer.
Diessel, H., 2019. The Grammar Network: How Linguistic Structure Is Shaped by Language Use. Cambridge University Press.
Divjak, D., 2019. Frequency in Language: Memory, Attention and Learning. Cambridge University Press.
Divjak, D., Gries, S. Th, 2006. Ways of trying in Russian: clustering behavioral profiles. Corpus Linguist. Linguistic Theory 2 (1), 23–60.
Ellis, N.C., 2007. Language acquisition as rational contingency learning. Appl. Ling. 27 (1), 1–24.
Ellis, N.C., Ferreira-Junior, F., 2009. Constructions and their acquisition: island and the distinctiveness of their occupancy. In: Ruiz, F.J., Ibáñez, M. (Eds.), Annual Review of Cognitive Linguistics, vol. 7. John Benjamins, pp. 187–220.
Everitt, B.S., Landau, S., Leese, M., Stahl, D., 2011. Cluster Analysis, fifth ed. Wiley-Blackwell.
Firth, J.R., 1957. Papers in Linguistics, 1931–1951. Oxford University Press.
Flach, S., 2021. Collostructions: An R Implementation for the Family of Collostructional Methods (Version 0.2.0) [R Script].
Greenacre, M., 2017. Correspondence Analysis in Practice, third ed. CRC Press.
Gries, S. Th, 2019. Ten Lectures on Corpus-Linguistic Approaches: Applications for Usage-Based and Psycholinguistic Research. Brill.
Gries, S. Th, 2021. Statistics for Linguistics with R, third ed. De Gruyter Mouton.
Gries, S. Th, Ellis, N.C., 2015. Statistical measures for usage-based linguistics. Lang. Learn. 65 (1), 228–255.
Gries, S. Th, Stefanowitsch, A., 2004a. Extending collostructional analysis: a corpus-based perspective on alternations. Int. J. Corpus Linguist. 9 (1), 97–129.
Gries, S. Th, Stefanowitsch, A., 2004b. Co-varying collexemes in the into-causative. In: Achard, M., Kemmer, S. (Eds.), Language, Culture, and Mind. CSLI Publications, pp. 225–236.
Gries, S. Th, Stefanowitsch, A., 2010. Cluster analysis and the identification of collexeme classes. In: Rice, S., Newman, J. (Eds.), Empirical and Experimental Methods in Cognitive/Functional Research. CSLI Publications, pp. 73–90.
Hamp, B., Feldweg, H., 1997. GermaNet – a lexical-semantic net for German. In: Vossen, P., Calzolari, N., Wilks, Y. (Eds.), Proceedings of the ACL Workshop Automatic Information Extraction and Building of Lexical Semantic Resources for NLP Applications. Association for Computational Linguistics, pp. 9–15.
Harris, Z.S., 1970. Papers in Structural and Transformational Linguistics. Reidel.
Hilpert, M., 2006. Distinctive collexeme analysis and diachrony. Corpus Linguist. Linguistic Theory 2 (2), 243–256.

Hundsnurscher, F., Splett, J., 1982. *Semantik der Adjektive im Deutschen: Analyse der semantischen Relationen* [Semantics of Adjectives in German: Analysis of Semantic Relations]. Westdeutscher Verlag.
Inkpen, D., Hirst, G., 2006. Building and using a lexical knowledge base of near-synonym differences. Comput. Ling. 32 (2), 223–262.
Israel, M., 2001. Minimizers, maximizers and the rhetoric of scalar reasoning. J. Semant. 18 (4), 297–331.
Kennedy, C., 2007. Vagueness and grammar: the semantics of relative and absolute gradable adjectives. Ling. Philos. 30 (1), 1–45.
Kennedy, C., McNally, L., 2005. Scale structure, degree modification, and the semantics of gradable predicates. Language 81 (2), 345–381.
Kennedy, G., 2003. Amplifier collocations in the British national corpus: implications for English language teaching. Tesol Q. 37 (3), 467–487.
Kilgarriff, A., Rychly, P., Smrz, P., Tugwell, D., 2004. The sketch engine. In: Williams, G., Vessier, S. (Eds.), Euralex 2004 Proceedings. Université de Bretagne-Sud.
Levshina, N., 2015. How to Do Linguistics with R. Data Exploration and Statistical Analysis. John Benjamins.
Ljubešić, N., Klubička, F., 2014. {bs,hr,sr}WaC –Web corpora of Bosnian, Croatian and Serbian. In: Bildhauer, F., Schäfer, R. (Eds.), Proceedings of the 9th Web as Corpus Workshop (WaC-9). Association for Computational Linguistics, pp. 29–35.
Matešić, M., Memišević, A., 2016. Pragmatics of adjectives in academic discourse: from qualification to intensification. Jezikoslovlje 17 (1), 179–206.
Miller, G.A., 1995. WordNet: a lexical database for English. Commun. ACM 38 (11), 39–41.
Nigoević, M., 2020. *Intenzifikacija u jeziku: s primjerima iz hrvatskog i talijanskog jezika* [Intensification in Language: With Examples from the Croatian and Italian Language]. Filozofski fakultet Sveučilišta u Splitu.
Paradis, C., 1997. Degree Modifiers of Adjectives in Spoken British English. Lund University Press.
Paradis, C., 2001. Adjectives and boundedness. Cognit. Ling. 12 (1), 47–65.
Paradis, C., 2005. Ontologies and construals in lexical semantics. Axiomathes 15, 541–573.
Paradis, C., 2008. Configurations, construals and change: expressions of DEGREE. Engl. Lang. Ling. 12 (2), 317–343.
Pavić Pintarić, A., Frleta, Z., 2014. Upwards intensifiers in the English, German and Croatian language. J. Foreign Lang. 6 (1), 31–48.
Perek, F., 2016. Using distributional semantics to study syntactic productivity in diachrony: a case study. Linguistics 54 (1), 149–188.
Proisl, T., 2022. *Use words, not constructions!* A new perspective on the unit of analysis in collostructional analysis. Int. J. Corpus Linguist. 27 (3), 349–379.
Quirk, R., Greenbaum, S., Leech, G., Svartvik, J., 1985. A Comprehensive Grammar of the English Language. Longman.
R Core Team, 2022. *R: A Language and Environment for Statistical Computing* [Computer Software]. R Foundation for Statistical Computing.
Raffaelli, I., Bekavac, B., Agić, Ž., Tadić, M., 2008. Building Croatian WordNet. In: Tanács, A., Csendes, D., Vincze, V., Fellbaum, Ch, Vossen, P. (Eds.), Proceedings of the Fourth Global WordNet Conference (GWC 2008). University of Szeged, pp. 349–359.
Rotstein, C., Winter, Y., 2004. Total adjectives vs. partial adjectives: scale structure and higher-order modifiers. Nat. Lang. Semant. 12 (3), 259–288.
Shimodaira, H., 2019. Computing selective inference p-values of clusters using pvclust and scaleboot. Available on: http://stat.sys.i.kyoto-u.ac.jp/prog/scaleboot/pvclust.pdf.
Shimodaira, H., Terada, Y., 2019. Selective inference for testing trees and edges in phylogenetics. Frontiers in Ecology and Evolution 7, 1–15.
Silić, J., Pranjković, I., 2005. *Gramatika hrvatskoga jezika za gimnazije i visoka učilišta* [Grammar of the Croatian Language for High Schools and Universities]. Školska knjiga.
Sinclair, J., 1991. Corpus, Concordance, Collocation. Oxford University Press.
Stefanowitsch, A., 2013. Collostructional analysis. In: Hoffman, T., Trousdale, G. (Eds.), The Oxford Handbook of Construction Grammar. Oxford University Press, pp. 209–306.
Stefanowitsch, A., Flach, S., 2020. 'Too big to fail but big enough to pay for their mistakes': a collostructional analysis of the patterns [too ADJ to V] and [ADJ enough to V]. In: Corpas, G., Colson, J.P. (Eds.), Computational and Corpus-Based Phraseology. John Benjamins, pp. 247–272.
Stefanowitsch, A., Gries, S. Th, 2003. Collostructions: investigating the interaction of words and constructions. Int. J. Corpus Linguist. 8 (2), 209–243.
Stefanowitsch, A., Gries, S. Th, 2005. Covarying collexemes. Corpus Linguist. Linguistic Theory 1 (1), 1–43.
Stoffel, C., 1901. Intensives and down-toners: a study in English adverbs. Carl Winter's Universitätsbuchhandlung.
Suzuki, R., Shimodaira, H., 2006. Pvclust: an R package for assessing the uncertainty in hierarchical clustering. Bioinformatics 22 (12), 1540–1542.
Tagliamonte, S., 2008. So different and pretty cool! Recycling intensifiers in Toronto, Canada. Engl. Lang. Ling. 12 (2), 361–394.
Taylor, J.R., 2003. Near synonyms as co-extensive categories: 'high' and 'tall' revisited. Lang. Sci. 25 (3), 263–284.
Vidaković Erdeljić, D., 2023. Msm da sam totalno cool for the summer: uporaba intenzifikatora u korpusu hrvatskih tweetova [I think that I'm totally cool for the summer: the use of intensifiers in the corpus of Croatian tweets]. In: Mikić Čolić, A., Glušac, M. (Eds.), *Standardni i nestandardni idiomi* [Standard and Non-standard Idioms]. Hrvatsko društvo za primijenjenu lingvistiku, pp. 251–262.
Ward, J.H., 1963. Hierarchical grouping to optimize an objective function. J. Am. Stat. Assoc. 58, 236–244.
Xiao, R., Tao, H., 2007. A corpus-based sociolinguistic study of amplifiers in British English. Socioling. Stud. 1 (2), 241–273.