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

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 multi-factorial 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, Battini 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/© 2023 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).
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 (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 opuštano 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 predlości su potpuno prilagodljivi vasem 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 košobe na ramenima, noge totalno rúzne.
        (“[…] 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 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. 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. stolječe “you are totally 20th century” (Vidaković Erdejić, 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.

---
Boca je potpuno “tall”, skup “expensive”;
ii. Lower closed scale adjectives [0,1]: prljav “dirty”, mokar “wet”;
iii. Upper closed scale adjectives [0,1]: cist “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 felicitous.

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 mrav [...] [hrWaC] “Honestly the first few days I was totally dead [...]”) and partitive readings (e.g. Prele 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, with maximizers imposing particular modes of construal. 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.

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 colloclatical 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).

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 hrWaC2.2 (Ljubesić 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. 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. |
| Totally is different. |
| Totalno je bio “be-3SG.PST.” različit. |
| Totally was 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 (skroz + adjective, potpuno + adjective, sasvim + adjective, posve + 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 (skroz + 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 (skroz + 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 $2 \times 2$ contingency table, and the process is conducted for every adjective that occurs in <skroz + 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\(^8\) ($\chi^2$), “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 ($\chi^2$) 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 adjectival 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).
The degree of attraction between the construction and the collexeme (and vice versa⁸) 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 *drugaciji* “different” and *drukčiji* “different”, synonyms derived from the numeral adjective *drug* “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

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

⁸ 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 (Baric et al., 1997; Silić and Pranjković, 2005), few relevant works on degree modifiers in Croatian (Pavlic Pintaric and Frleta, 2014; Nigojević, 2020), and hrWaC corpus$^{10}$—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”, previse “too + ADJ”, prilicno “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 adjectival 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 $\chi^2$ 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>
<thead>
<tr>
<th></th>
<th>osobito</th>
<th>posve</th>
<th>potpuno</th>
<th>previse</th>
<th>prilicno</th>
</tr>
</thead>
<tbody>
<tr>
<td>osobito</td>
<td>0.0000</td>
<td>279.1664</td>
<td>281.0000</td>
<td>263.4204</td>
<td>271.1566</td>
</tr>
<tr>
<td>posve</td>
<td>279.1664</td>
<td>0.0000</td>
<td>222.4293</td>
<td>281.0000</td>
<td>275.0213</td>
</tr>
<tr>
<td>potpuno</td>
<td>281.0000</td>
<td>222.4293</td>
<td>0.0000</td>
<td>281.0000</td>
<td>275.0400</td>
</tr>
<tr>
<td>previse</td>
<td>263.4204</td>
<td>281.0000</td>
<td>281.0000</td>
<td>0.0000</td>
<td>272.6793</td>
</tr>
<tr>
<td>prilicno</td>
<td>271.1566</td>
<td>275.0213</td>
<td>275.0400</td>
<td>272.6793</td>
<td>0.0000</td>
</tr>
</tbody>
</table>

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 multidisciplinary 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.
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), 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 (narocito 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

<table>
<thead>
<tr>
<th>&lt;posve + adjective&gt;</th>
<th>observed frequency</th>
<th>expected frequency</th>
<th>coll. str. log-likelihood</th>
<th>(J^p_1)</th>
<th>(J^p_2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>jasan</td>
<td>2500</td>
<td>1410.7</td>
<td>875.82</td>
<td>0.02688</td>
<td>0.12389</td>
</tr>
<tr>
<td>drukciji</td>
<td>1342</td>
<td>807.1</td>
<td>370.43</td>
<td>0.01320</td>
<td>0.10499</td>
</tr>
<tr>
<td>siguran</td>
<td>1322</td>
<td>867.5</td>
<td>255.50</td>
<td>0.01215</td>
<td>0.08310</td>
</tr>
<tr>
<td>neobican</td>
<td>227</td>
<td>92</td>
<td>180.68</td>
<td>0.00233</td>
<td>0.22879</td>
</tr>
<tr>
<td>razumljiv</td>
<td>578</td>
<td>335.8</td>
<td>178.62</td>
<td>0.00598</td>
<td>0.11314</td>
</tr>
<tr>
<td>neobicajen</td>
<td>156</td>
<td>57.3</td>
<td>150.60</td>
<td>0.00244</td>
<td>0.26849</td>
</tr>
<tr>
<td>razlicit</td>
<td>1310</td>
<td>963.6</td>
<td>138.35</td>
<td>0.00855</td>
<td>0.05713</td>
</tr>
<tr>
<td>tocanc</td>
<td>311</td>
<td>188.3</td>
<td>82.45</td>
<td>0.00303</td>
<td>0.10189</td>
</tr>
<tr>
<td>neskekivkan</td>
<td>161</td>
<td>82.1</td>
<td>74.16</td>
<td>0.00195</td>
<td>0.15001</td>
</tr>
<tr>
<td>originalan</td>
<td>138</td>
<td>66.5</td>
<td>73.84</td>
<td>0.00176</td>
<td>0.16767</td>
</tr>
</tbody>
</table>

\(^{16}\) The high correlation among the analysed maximizers is further supported by the Spearman’s \(\rho\) 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).
Tables 3–7 present additional information, precisely the DP value (Ellis, 2007; Ellis and Ferreira-Junior, 2009), which addresses certain limitations of $G^2$. Unlike $G^2$, DP (i) is asymmetric and does not conflate $p(\text{word2}|\text{word1})$ and $p(\text{word1}|\text{word2})$ 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. DP divides into two distinct values: $DP_1$ assesses the predictive capacity of the collexeme (slot 2) for the construction (slot 1), whereas $DP_2$ 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 (\( \Delta P_2 \) values are significantly higher than \( \Delta P_1 \)). Among the five examined maximizers, totalno stands out as the most predictive when considering its top 10 most distinctive collexemes, with a mean \( \Delta P_2 \) of 0.61424. The highest level of predictiveness is witnessed in the case of totalno drugačiji, where the maximizer totalno almost impeccably (\( \Delta P_2 = 0.93358 \)) 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, cudan, debilan) and those indicating difference (drugaciji, razlicit). Notably, the only adjective that does not fall into these two categories, as seen in Table 7, is cool\(^\text{17}\). 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.

![Histogram representing the connotation of the top 200 distinctive collexemes of totalno.](image)

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. razlicit “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 \( \chi^2 \)-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 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 \( \chi^2 = 144.558; 18 \text{ df} = \text{NA}; p\text{-value} = 4.998 \times 10^{-4} \). 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 \( \text{V} \) amounts to 0.4450. The value of \( \text{V} = 0.4450 \) 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.

---

**Table 8**

*Input for CA (sampled).*

<table>
<thead>
<tr>
<th>MAXIMIZER</th>
<th>posve</th>
<th>potpuno</th>
<th>skroz</th>
<th>sasvim</th>
<th>totalno</th>
</tr>
</thead>
<tbody>
<tr>
<td>dovoljan</td>
<td>147</td>
<td>68</td>
<td>32</td>
<td>5036</td>
<td>0</td>
</tr>
<tr>
<td>normalan</td>
<td>1182</td>
<td>1278</td>
<td>407</td>
<td>6346</td>
<td>63</td>
</tr>
<tr>
<td>solidan</td>
<td>70</td>
<td>0</td>
<td>55</td>
<td>2757</td>
<td>0</td>
</tr>
<tr>
<td>automatski</td>
<td>10</td>
<td>210</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>glup</td>
<td>29</td>
<td>105</td>
<td>72</td>
<td>60</td>
<td>537</td>
</tr>
<tr>
<td>seksi</td>
<td>0</td>
<td>0</td>
<td>4</td>
<td>0</td>
<td>48</td>
</tr>
<tr>
<td>ljudski</td>
<td>94</td>
<td>266</td>
<td>3</td>
<td>90</td>
<td>4</td>
</tr>
<tr>
<td>prazan</td>
<td>115</td>
<td>972</td>
<td>119</td>
<td>58</td>
<td>90</td>
</tr>
<tr>
<td>cudan</td>
<td>235</td>
<td>32</td>
<td>110</td>
<td>55</td>
<td>154</td>
</tr>
<tr>
<td>simpatican</td>
<td>8</td>
<td>0</td>
<td>76</td>
<td>61</td>
<td>0</td>
</tr>
</tbody>
</table>

---

18 Pearson’s Chi-squared test with simulated \( p \)-value (based on 2000 replicates).
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 ($\Phi^2$) — obtained by dividing the $\chi^2$ 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 $\Phi^2$, while the second axis (dimension 2) represents 33.68% of the $\Phi^2$. 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.

![Fig. 4. CA biplot of the <maximizer + adjective> construction in hrWaC [translated].](image)
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>
<thead>
<tr>
<th>Table 9</th>
<th>Semantic classes of adjectives.</th>
</tr>
</thead>
<tbody>
<tr>
<td>BEHAVIOR-RELATED</td>
<td>Behaviour</td>
</tr>
<tr>
<td></td>
<td>Character</td>
</tr>
<tr>
<td></td>
<td>Inclination</td>
</tr>
<tr>
<td></td>
<td>Sympathy</td>
</tr>
<tr>
<td>BODY-RELATED</td>
<td>Affliction</td>
</tr>
<tr>
<td></td>
<td>Appearance</td>
</tr>
<tr>
<td>GENERAL</td>
<td>Evaluation</td>
</tr>
<tr>
<td>MATERIAL-RELATED</td>
<td>Dampness</td>
</tr>
<tr>
<td></td>
<td>Ripeness</td>
</tr>
<tr>
<td></td>
<td>Stability</td>
</tr>
<tr>
<td></td>
<td>State</td>
</tr>
<tr>
<td>MOOD-RELATED</td>
<td>Feeling</td>
</tr>
<tr>
<td>PERCEPTIONAL</td>
<td>Colour</td>
</tr>
<tr>
<td></td>
<td>Lightness</td>
</tr>
<tr>
<td></td>
<td>Taste</td>
</tr>
<tr>
<td>PERTAINYMS</td>
<td>Costs</td>
</tr>
<tr>
<td>QUANTITY-RELATED</td>
<td>Accuracy</td>
</tr>
<tr>
<td>RELATIONAL</td>
<td>Certainty</td>
</tr>
<tr>
<td></td>
<td>Completeness</td>
</tr>
<tr>
<td></td>
<td>Correspondence</td>
</tr>
<tr>
<td></td>
<td>Difficulty</td>
</tr>
<tr>
<td></td>
<td>Functioning</td>
</tr>
<tr>
<td></td>
<td>Linking</td>
</tr>
<tr>
<td></td>
<td>Priveative</td>
</tr>
<tr>
<td></td>
<td>Reference</td>
</tr>
<tr>
<td></td>
<td>Requirements</td>
</tr>
<tr>
<td></td>
<td>Security</td>
</tr>
<tr>
<td></td>
<td>Validity</td>
</tr>
<tr>
<td>SPATIAL</td>
<td>Dimension</td>
</tr>
<tr>
<td></td>
<td>Direction</td>
</tr>
<tr>
<td></td>
<td>Existence</td>
</tr>
<tr>
<td></td>
<td>Localization</td>
</tr>
<tr>
<td></td>
<td>Spatial distribution</td>
</tr>
<tr>
<td>SPIRIT-RELATED</td>
<td>Experience</td>
</tr>
<tr>
<td></td>
<td>Intelligence</td>
</tr>
<tr>
<td></td>
<td>Knowledge</td>
</tr>
<tr>
<td></td>
<td>Age</td>
</tr>
</tbody>
</table>

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 $\phi^2$, while the second axis (dimension 2) represents 31.83 % of the $\phi^2$, 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”, los “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 (nevrisan “independent”, slobodan “free”, oduvojen “separated”);
iv. sasvim accompanies adjectives denoting conformity or deviation from the norm (obican “regular”, prosječan “average”, drugačiji “different”, cudan “strange, odd”) and simplicity (jednostavan “simple”);
v. posve is usually related with adjectives signaling accuracy (prepočet “overly general”, pogresan “incorrect”, točan “correct”), resemblance (slican “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”, seks “sexy”), whereas potpuno modifies affliction (zdрав “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 kuf), 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


Bordet, L., 2017. From vogue words to lexicalized intensifying words: the renewal and recycling of intensifiers in French, German and Croatian. Rasprave 42 (2), 199–216.


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.


