EJI (English Journal of Indragiri): Studies in Education, Literature, and Linguistics
Vol. 8. No. 2, July 2024
ISSN (Print). 2549-2144, ISSN (Online). 2589-5140
Homepage: https://ejournal-fkip.unisi.ac.id/index.php/eji/

Exploring Grammatical and Semantic Profiles of ANGRY and MAD: A Corpus-Based Study

Ida Ayu Saskara Tranggana Suari [1], Gede Primahadi Wijaya Rajeg [2], I Nengah Sudipa [3]
[1,2,3]Universitas Udayana, [2]Centre for Interdisciplinary Research on the Humanities and Social Sciences (CIRHSS) - Universitas Udayana, [2]University of Oxford
Email: [1]*tranggana.suari@gmail.com*

Article History:
Received : 04/05/2024
Accepted : 10/07/2024
Published : 16/07/2024

Keywords:
corpus linguistics, grammatical profile, semantic profile, synonyms, usage-based approach

Abstract

This paper investigates how two near-synonymous emotion words (i.e., ANGRY and MAD) would differ in their grammatical (i.e., morphological) and semantic profiles based on data from the *Corpus of Contemporary American English* (COCA). Morphologically, we analysed the distribution of verbal inflectional morphologies of these words overall and across different text types/genres in COCA. Semantically, we explored the preferred semantic category of the Experiencer and Stimulus collocates of ANGRY and MAD. In that way, we adopted both quantitative and qualitative methods. We showed that the base verbal form *anger* is very prominent across genres compared to the base form *madden*. Similarly, the *-ed* and third-person singular *-s* forms are more predominant for ANGER than for MADDEN. In contrast, the *-ing* form *maddening* is far more common than *angering*. Semantically, *angry* predominantly collocates with Kinship-based Experiencer than *mad*. Both adjectives attract distinct types of Experiencer nouns frosm the Social and political relation fields. In terms of the Stimulus, *mad* attracts collocates from the semantic field of Emotion and values more predominantly than *angry*. In sum, corpus analyses help reveal grammatical and semantic differences between near-synonyms such as ANGRY and MAD

Kata Kunci:
linguistik korpus, profil gramatikal, profil semantik, sinonim, pendekatan berbasis penggunaan

Abstrak

Kajian korpus menyangkut penggunaan kata-kata dalam penggunaan bahasa di sehari-hari. Mempertimbangkan hal ini, makalah ini menyelidiki dua kata sifat emosi yang hampir sinonim (yaitu, *ANGRY* dan *MAD*) akan berbeda dalam perilaku gramatikal (yaitu morfologi) dan semantiknya berdasarkan data dari *Corpus of Contemporary*

EJI (English Journal of Indragiri): Studies in Education, Literature, and Linguistics
Vol. 8. No. 2, July 2024
ISSN (Print). 2549-2144, ISSN (Online). 2589-5140
Homepage: https://ejournal-fkip.unisi.ac.id/index.php/eji/

American English (COCA). Secara morfologis, kami menganalisis distribusi morfologi infleksi verbal kedua kata tersebut secara keseluruhan dan di berbagai jenis/genre teks di COCA. Secara semantik, kami menjelajahi tipe semantik kolokat peran *Experiencer* (Pengalam) dan *Stimulus* dari *ANGRY* dan *MAD*. Untuk tujuan tersebut, kami menerapkan metode kuantitatif dan kualitatif. Kami menunjukkan bahwa bentuk verbal dasar *anger* sangat menonjol di berbagai genre dibandingkan dengan bentuk dasar *madden*. Demikian pula, bentuk *-ed* dan orang ketiga tunggal *-s* lebih dominan untuk *anger* dibandingkan *madden*. Sebaliknya, bentuk *-ing* untuk *maddening* jauh lebih umum daripada *angering*. Secara semantik, *angry* (dibandingkan *mad*) lebih banyak dikaitkan dengan Pengalam dari ranah Kekerabatan (*kinship*). *Angry* dan *mad* juga berasosiasi dengan jenis Pengalam khas dari bidang hubungan sosial dan politik. Terkait dengan tipe semantik peran *Stimulus*, *mad* condong berkolokasi dengan Stimulus jenis Emosi dan nilai-nilai yang lebih dominan daripada *angry*. Sebagai penutup, kajian korpus kuantitatif membantu mengungkap perbedaan nuansa gramatikal dan semantik dua kata yang bersinonim.

INTRODUCTION

It is a known fact that English, as a globally used language, has a vast and rich vocabulary which results in many words having similar meanings. For instance, take the words “sad”, “sorrowful”, and “gloomy”. All these words express the feeling of sadness, yet each word carries a slightly different nuance. Essentially, “sad” is an emotion that conveys the transient condition of being unhappy. It might vary from mild disappointment to severe sorrow. “Sorrowful”, on the other hand, conveys a deeper sense of melancholy and loss. It is a strong emotion brought on by a specific tragedy or loss. Lastly, “gloomy” can be defined as a mood that is so oppressive and dense with darkness or bleakness that it is frequently used to characterise such moods. “Gloomy” can also apply to internal emotional states as well as exterior conditions (i.e. the weather or surroundings). Synonymy is a complex concept since two lexemes never have the same range of (morpho-)syntactic and semantic occurrences

(Kreidler, 1998:97). Even in cases when they do share occurrences and predict the same class, synonymy is not straightforward in referring terms, and their references are probably going to be different (Kreidler, 1998:97). For two terms in a language to have the same meaning and occur in the exact same settings would be inefficient.

The tendency today is people often use synonyms without paying attention to their correct usage in a sentence, especially emotional terms to express our feelings. In a cross-linguistic study between the Polish and English languages by Dziwirek and Lewandowska-Tomaszczyk (2010), Slavic speakers emphasise using verbs to communicate emotion, while English speakers select adjectives to convey feelings, reactions, assessments, physical and mental states. In relation to feelings, while English speakers overwhelmingly favour adjectives and participles, Polish frequently utilise verbs, especially reflexive verbs (Dziwirek and Lewandowska-Tomaszczyk, 2010:20). The study also uncovered several startling previously overlooked trends, such as the fact that FEAR and SADNESS are the most developed emotion domains in both languages. For the domain of SADNESS for instance, English has a significantly higher number of lexical expressions than Polish, and that the most notable difference between the two languages is in the semantic area of DISGUST/DISLIKE (Dziwirek and Lewandowska-Tomaszczyk, 2010:25).

Understanding how subtle grammatical variations contribute to complex semantic distinctions is essential to investigate the linguistic patterns of synonyms (Janda, 2016). In this paper, we continue this line of research into analysing the English adjectival concepts ANGRY and MAD. Words from the domain of ANGER are focused on because they have been studied in many cross-linguistic research from the domain of metaphor (e.g., Kövecses, 2000) and cross-linguistic research on their grammatical behaviour (e.g., Dziwirek & Lewandowska-Tomaszczyk, 2010). In this study, we attempt to look at the internal, within-language variation between these synonyms. These are chosen as the subjects for investigation with the intention of analysing how their morphological profile of inflectional verbal forms as well as semantic profiles

of Experiencer and Stimulus collocates contribute to the distinctions of their usages or meanings (Janda & Lyashevskaya, 2011; Liu, 2016).

We investigated two research questions. Firstly, what are the distribution of the inflectional forms (i.e., the grammatical/morphological profile) of ANGRY and MAD when used as verbs (i.e., as "anger" and "madden") overall and across different text types/genres? This is answered in Section 3.2. Secondly, what do the Experiencer (3.3) and Stimulus (3.4) collocates reveal regarding the semantic profiles of ANGRY and MAD? These adjectives both convey a sense of 'rage' but the specific usage can be different, as we will show in this paper. This study adopts the corpus linguistic method that involves analysing a corpus which is a collection of data concerning actual language use (Gries, 2009).

To establish a valid foundation for this research, there were five literatures reviewed in this article. Xiao and McEnery (2006) discussed a cross-linguistic study of collocation, semantic prosody and near synonyms between English and Chinese. The finding showed that despite the obvious differences between English and Chinese, the two languages' collocational patterns and semantic prosodies of near synonyms are comparable (take CAUSE and zao4cheng2 [造成] as examples). Another article by Effendi, Amalia and Lalita (2020) analysed the synonyms "announce", "declare", and "state" based on data from the British National Corpus (BNC) to highlight how corpus linguistics can be used to teach vocabulary and meaning in use. An undergraduate thesis by Gumelar (2018) discussed the Linguistic Features in Heavy Metal Songs from 2000-2018 and made use of AntConc as the tool for Corpus Analysis. Furthermore, Phitayakorn (2017) researched near-synonyms, namely "advise", "recommend", and "suggest", drawing data from the 6th edition of the Longman Dictionary of Contemporary English and the BNC. Lastly, Harta, Rajeg and Rahayuni (2023) analysed the semantic preferences of the verbal synonyms "hurry" and "rush" via the nouns that collocate with them in the Corpus of Contemporary American English (COCA). The *Concepticon* (List et al., 2024), a cross-linguistic semantic catalogue, was utilised in Harta et al.'s study to classify the collocates based on

EJI (English Journal of Indragiri): Studies in Education, Literature, and Linguistics
Vol. 8. No. 2, July 2024
ISSN (Print). 2549-2144, ISSN (Online). 2589-5140
Homepage: https://ejournal-fkip.unisi.ac.id/index.php/eji/

their semantic field. Our paper incorporates Harta et al.'s (2023) method of using the *Concepticon* for categorising the semantic field of Experiencer and Stimulus Collocates of ANGRY and MAD.

METHOD

To answer our first research question, we used the Part-of-Speech [POS]/Word Class tagging in COCA to retrieve the morphological, inflectional profiles of the verbs "anger" and "madden". COCA provided a [POS] feature to select any type of Part of Speech to assign to the two target verbs (see Rajeg, 2020 for the audiovisual tutorial of COCA). For instance, to see the relative frequency of "anger" as verbs, we used the List feature of COCA and searched for the different inflectional forms using the search pattern ANGER_v (see Figure 1). The search result is shown in Figure 2.

Figure 1 List feature in COCA for searching the frequency of a pattern.

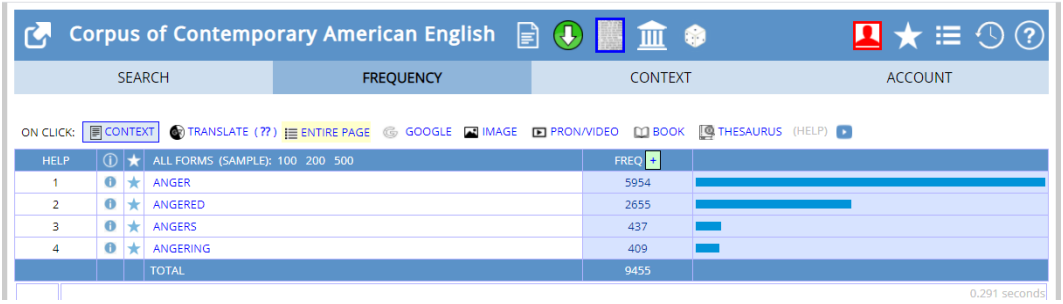


Figure 2 Inflectional Verbal Forms of ANGER.

The data we analysed is from the output of the Chart feature of COCA for the first research question. The specific inflectional forms we are interested in are the base form, past participle form, third person singular *-s* form, and gerund *-ing*

EJI (English Journal of Indragiri): Studies in Education, Literature, and Linguistics
Vol. 8. No. 2, July 2024
ISSN (Print). 2549-2144, ISSN (Online). 2589-5140
Homepage: https://ejournal-fkip.unisi.ac.id/index.php/eji/

form of the two verbs. We search these inflectional forms independently in the Chart feature using the POS feature in COCA.

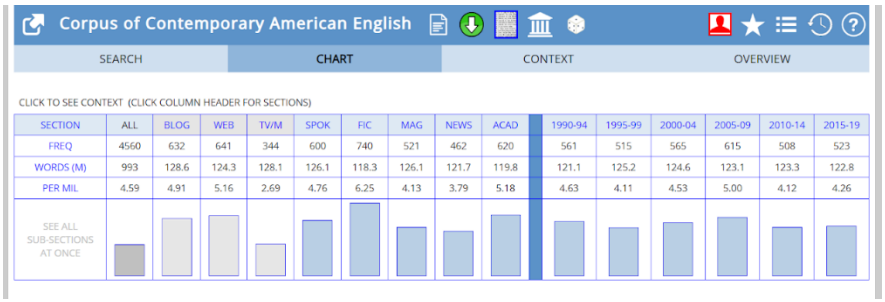


Figure 3 Output of the Chart feature.

The second research question concerns the Experiencer and Stimulus collocates of “angry” and “mad”. To determine Experiencer collocates, we used the Compare feature of COCA (see Figure 4).

Figure 4 Interface of the Compare feature.

In the Compare feature, two words are being compared in terms of their collocates: “angry” is in this case chosen as Word1 while “mad” as Word2. Then, we focus on the noun collocates appearing one word after “anger” and “mad”. The output will show the nouns preferences of these two words, and this will be discussed in the findings and discussion section.

The next step was searching the Stimulus collocates of “angry” and “mad”. For this, we utilised the Word feature of COCA (see Figure 5). Firstly, the target word such as “angry” was typed in. Then, by clicking the button saying, ‘See

EJI (English Journal of Indragiri): Studies in Education, Literature, and Linguistics
Vol. 8. No. 2, July 2024
ISSN (Print). 2549-2144, ISSN (Online). 2589-5140
Homepage: https://ejournal-fkip.unisi.ac.id/index.php/eji/

detailed info for word', COCA generated a more elaborate information of said word, including topics, collocates, and related words.

Figure 5 Word feature interface.

To identify the pattern candidate representing the Stimulus collocates, the section Clusters was used (see Figure 6). For "anger", we focus on the phrasal pattern "angry *at* NOUN", "angry *with* NOUN", "angry *about* NOUN", and "angry *over* NOUN".

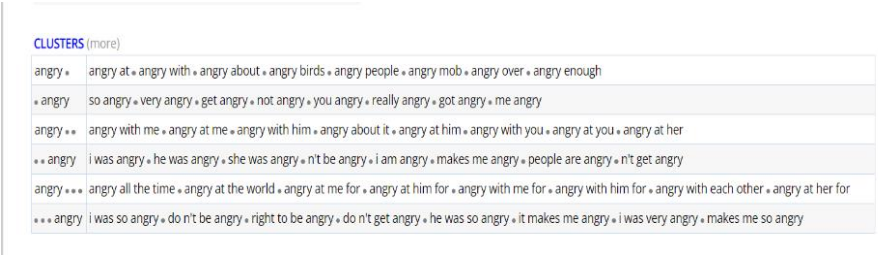


Figure 6 The CLUSTERS section in the output of the Word feature.

The NOUN slots in these patterns are analysed as representing the Stimulus of ANGRY.

Then, given that the pattern has been identified, we searched for the manifestation of the pattern with specific nouns using the List feature of COCA. In the List feature, each specific pattern above was searched for by typing, for instance "angry at NOUN" (Figure 7), onto the search field then clicking the "Find Matching Strings" button. COCA would then generate up to 100 data of the noun collocations with the given pattern (see the snippet of the output in Figure 8). This procedure was applied to "mad" as well.

EJI (English Journal of Indragiri): Studies in Education, Literature, and Linguistics
Vol. 8. No. 2, July 2024
ISSN (Print). 2549-2144, ISSN (Online). 2589-5140
Homepage: https://ejournal-fkip.unisi.ac.id/index.php/eji/

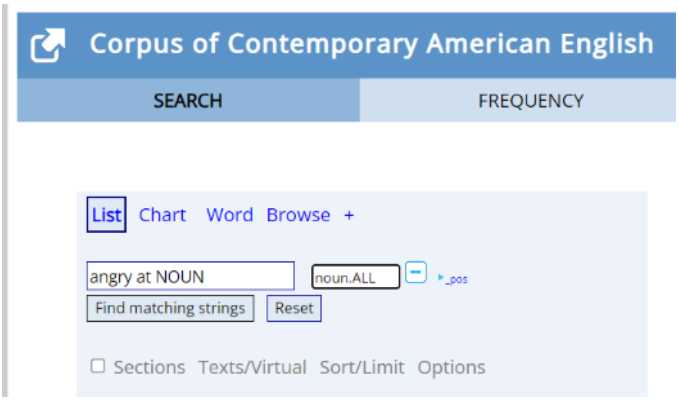


Figure 7 The List feature for searching the pattern [angry *at* NOUN].

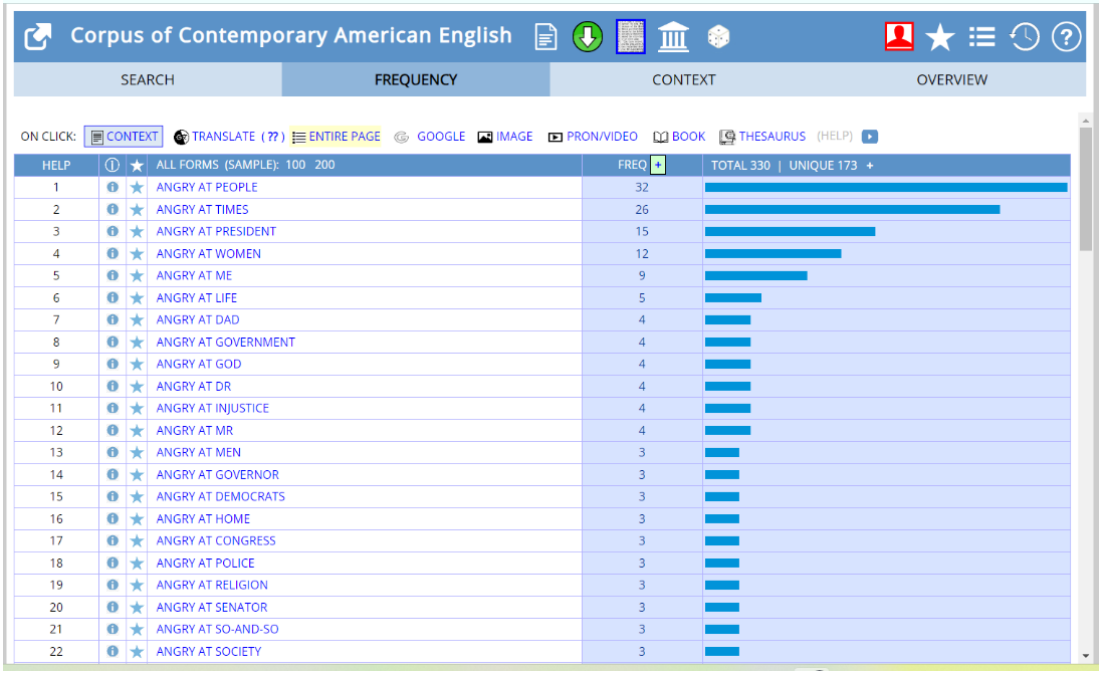


Figure 8 A snippet of the result of the 100 data sample for [angry *at* NOUN]

The collected data was analysed quantitatively and qualitatively. The quantitative method was used to identify and compare the frequencies of the morphological profiles and preferences of the collocates. These were then presented into a chart and/or tables. The qualitative method was used to determine the semantic field/type of the Experiencer and Stimulus collocates. The data was descriptively presented through a collection of words or phrases that clarify how the adjective was distinguished in its morphological profile, and collocational semantics. The underlying data and the programmatic R codes, especially to

EJI (English Journal of Indragiri): Studies in Education, Literature, and Linguistics
Vol. 8. No. 2, July 2024
ISSN (Print). 2549-2144, ISSN (Online). 2589-5140
Homepage: https://ejournal-fkip.unisi.ac.id/index.php/eji/

produce the visualisations, can be accessed at https://doi.org/10.5281/zenodo.11108458 (Suari et al., 2024) or https://github.com/complexico/anger-mad-coca.

FINDINGS AND DISCUSSION

This section presents the results and discussions concerning the two research questions. Section 3.1 provides the overall relative frequency of "angry" and "mad" across different text types/genres in COCA. Then, the first research question on the grammatical/morphological profile is addressed in Section 3.2. Finally, Sections 3.3 and 3.4 respectively discuss the results of collocational analyses for the Experiencer and Stimulus types of "angry" and "mad".

Relative Frequency of "Angry" and "Mad"

The fact that "angry" and "mad" appear to be similar in meaning, it is necessary to determine the extent to which they also have similar usages (i.e., occurrences) across different text types in COCA. Figure 9 shows this information.

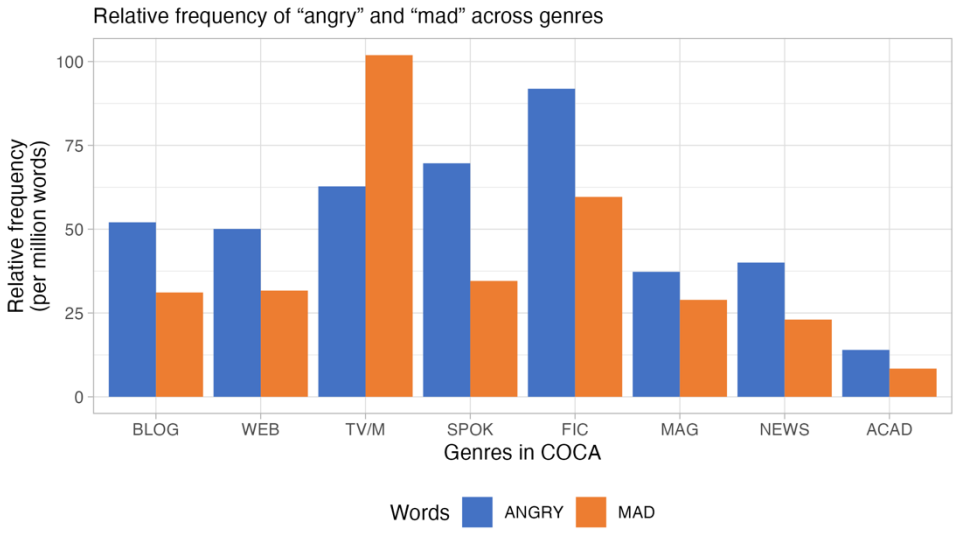


Figure 9 Relative frequency of "angry" and "mad".

EJI (English Journal of Indragiri): Studies in Education, Literature, and Linguistics
Vol. 8. No. 2, July 2024
ISSN (Print). 2549-2144, ISSN (Online). 2589-5140
Homepage: https://ejournal-fkip.unisi.ac.id/index.php/eji/

From Figure 9, it is evident that the adjective "angry" has a higher frequency rate than "mad" overall, aside from the TV/Movies Genre, where "mad" appears to be higher. Furthermore, in writing genres such as Blog, Web, Magazine, News, and Academic text, "mad" occurred less likely. Respectively, "angry" has the highest relative frequency appearing in the Fiction Genre. On the contrary, "mad" individually has the highest relative frequency on TV/Movies Genre. These data implied that when compared, ANGRY and MAD do seem to have quite a distinct differential distribution across genres/text types.

Grammatical Profiles of Angry and Mad as Verbs (Namely "Anger" and "Madden")

The first research question in this study seeks to explore the range of inflectional verbal forms of "angry" and "mad" as verbs (namely as "anger" and "madden") across genres. Figure 10 and Figure 11 show the distribution of the verbal forms for the verbal lemma ANGER and MADDEN respectively in the base verb form, *-ed* form, third person singular form, and in the *-ing* form.

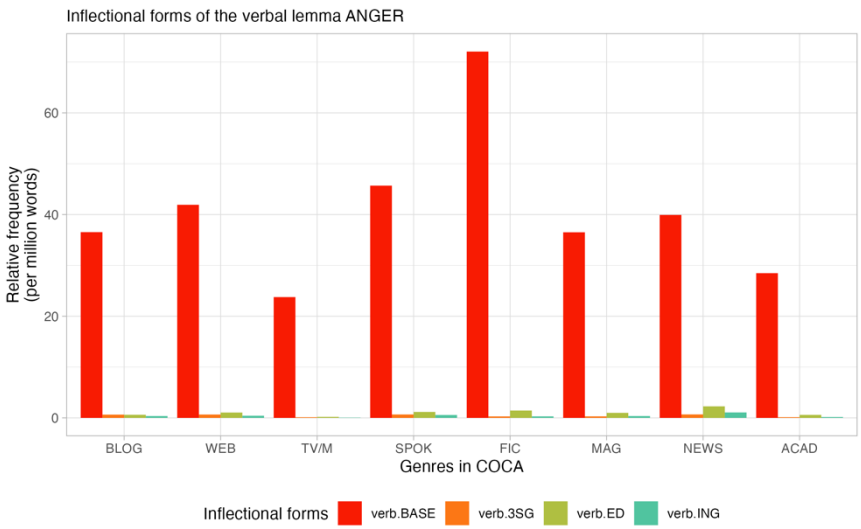


Figure 10 Inflectional verbal forms of the lemma ANGER across genres.

EJI (English Journal of Indragiri): Studies in Education, Literature, and Linguistics
Vol. 8. No. 2, July 2024
ISSN (Print). 2549-2144, ISSN (Online). 2589-5140
Homepage: https://ejournal-fkip.unisi.ac.id/index.php/eji/

Figure 10 shows that "anger" in the base inflectional verbal form (cf. the red bars) exhibits a considerable variation across genres and is obviously significantly higher than the remaining three verbal forms. It is also apparent from this chart that the other verbal forms namely "angered", "angers", and "angering" occurred much less frequently overall across genres. Furthermore, the highest frequency of the base form "anger" is apparent in the genre Fiction with the relative frequency of 72,06 occurrences per million words. The form "anger" occurred the least in the TV/Movies genre (23,76 occurrences per million words). These findings suggest that as the base form ANGER is much more common than the other three verbal forms.

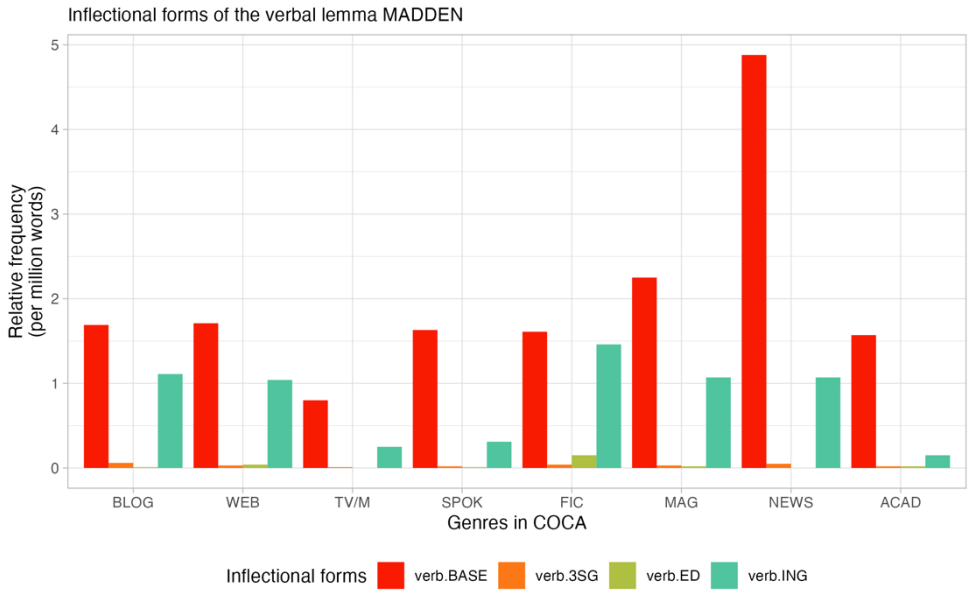


Figure 11 Inflectional verbal forms of the lemma MADDEN across genres.

Like ANGER, the verbal lemma MADDEN (Figure 11) is also realised more frequently in the base form across genres. However, when the base forms of these two verbs are compared (see Figure 12 below), "anger" still seems to rank significantly higher than "madden" across the whole genres.

EJI (English Journal of Indragiri): Studies in Education, Literature, and Linguistics
Vol. 8. No. 2, July 2024
ISSN (Print). 2549-2144, ISSN (Online). 2589-5140
Homepage: https://ejournal-fkip.unisi.ac.id/index.php/eji/

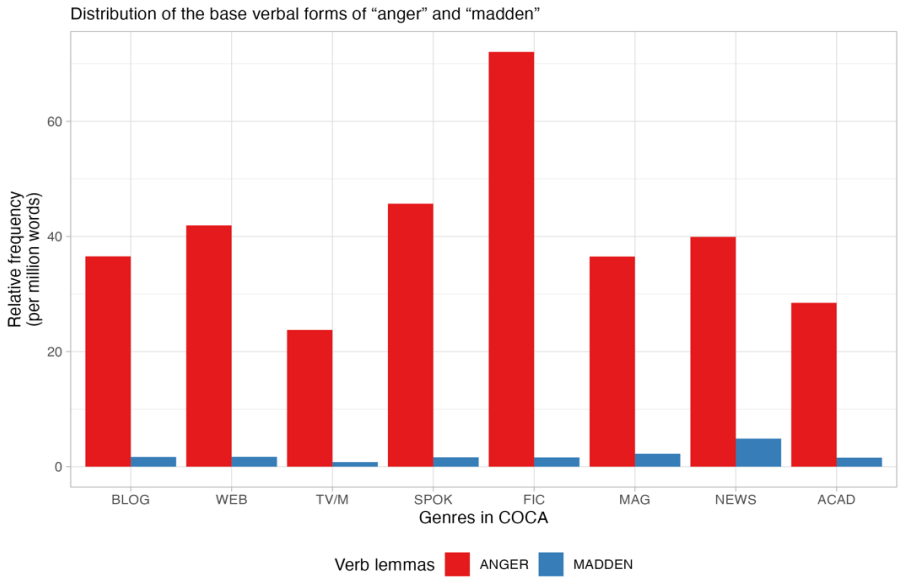


Figure 12 Distribution of the base verbal forms for the lemma ANGER and MADDEN.

Unlike ANGER, Figure 11 above further demonstrates that MADDEN seems to have two dominating verbal forms, namely the base form and the *-ing* form. The Academic genre is where the *-ing* form of MADDEN occurs the least. Figure 13 below visualises the direct comparison of the *-ing* form between the verbal lemmas ANGER and MADDEN across genres.

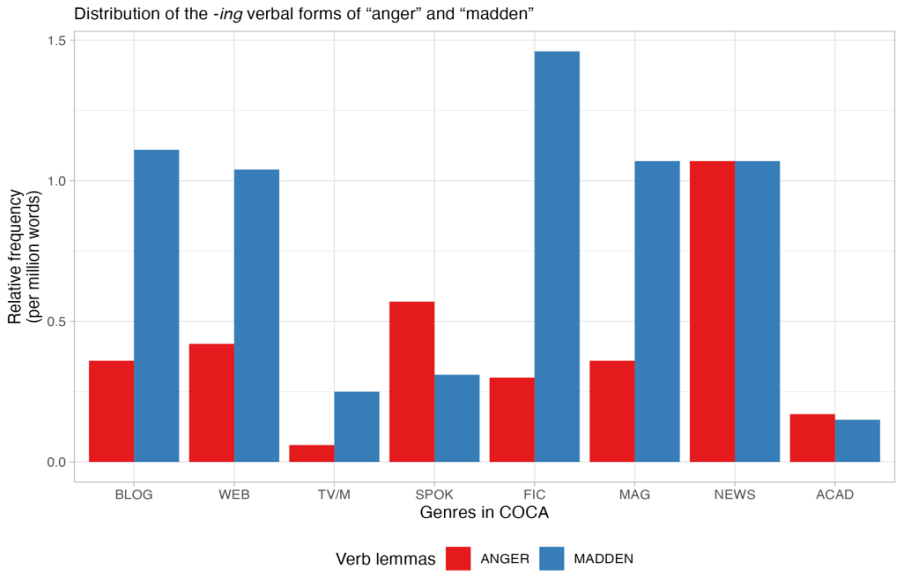


Figure 13 Distribution of the *-ing* verbal forms for the lemma ANGER and MADDEN.

EJI (English Journal of Indragiri): Studies in Education, Literature, and Linguistics
Vol. 8. No. 2, July 2024
ISSN (Print). 2549-2144, ISSN (Online). 2589-5140
Homepage: https://ejournal-fkip.unisi.ac.id/index.php/eji/

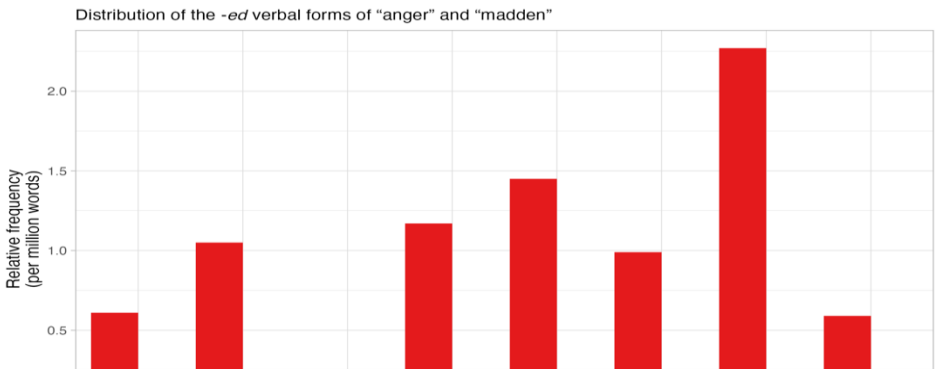
The distribution of the *-ing* form between MADDEN and ANGER is clearly different. In most genres (except in Academic, Spoken, and News), the form "maddening" (cf. the blue bars) occurs much more frequently than "angering" (the red bars). The following sentences were extracted from COCA to illustrate the *-ing* forms between these two verbs.

(1) "Hating your fat is an acceptable form of selfhatred and it's saddening and **maddening** at the same time. # But you just go swim, have fun,"

(2) "Sorry for the language Moe. That in essence is O'Bama's problem he is **angering** the people he doesn't wants to show up at the polls on Nov. 6th"

Example (2) shows that "angering" is used in a transitive context with direct object (i.e., *the people*). Meanwhile, the *-ing* form "maddening" in (1) is used intransitively as participial adjective (like *embarrassing, terrifying*). Future research is needed to determine the extent to which the *-ing* form of ANGER and MADDEN is differentiated according to their occurrences in transitive vs. intransitive context. Overall, this brief comparison for the *-ing* form of ANGER and MADDEN begins to reveal how semantically similar words such as the verbal lemma ANGER and MADDEN can have different distributional pattern, among others, along this morphological (grammatical) profile. Let us now consider the distribution of the *-ed* form of ANGER and MADDEN in

Figure 14 below.



EJI (English Journal of Indragiri): Studies in Education, Literature, and Linguistics
Vol. 8. No. 2, July 2024
ISSN (Print). 2549-2144, ISSN (Online). 2589-5140
Homepage: https://ejournal-fkip.unisi.ac.id/index.php/eji/

Figure 14 Distribution of the *-ed* verbal forms for the lemma ANGER and MADDEN.

The form "angered" is predominant than "maddened" (

Figure 14 above). The verbal lemma ANGER is not only much more frequent than MADDEN in the base and *-ed* forms, but also in the *-s* third-person singular form (Figure 15 below). Further interesting finding from

Figure 14 is that for MADDEN, the *-ed* form is most frequently used in the Fiction genre compared to the other genres.

EJI (English Journal of Indragiri): Studies in Education, Literature, and Linguistics
Vol. 8. No. 2, July 2024
ISSN (Print). 2549-2144, ISSN (Online). 2589-5140
Homepage: https://ejournal-fkip.unisi.ac.id/index.php/eji/

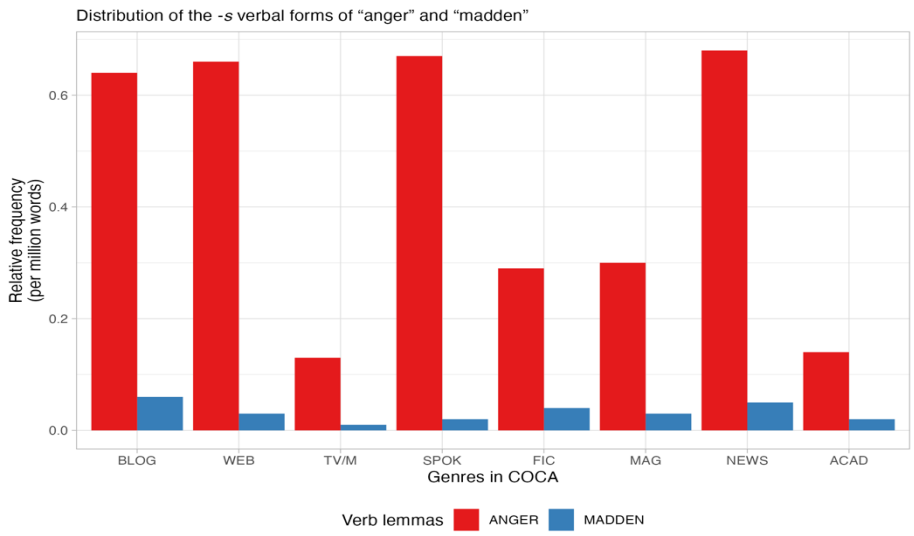


Figure 15 Distribution of the *-s* (third-person singular) verbal forms for the lemma ANGER and MADDEN.

The results discussed in this section demonstrate that near synonymous verb (al lemma)s such as ANGER and MADDEN may differ in their grammatical behaviours. More specifically, the lemma ANGER is more predominantly than MADDEN to occur in three morphological forms: the base, *-ed*, and *-s* forms; on the contrary, the *-ing* form is more common for MADDEN. Further implication of these results is that grammatical rule of inflectional morphology cannot be equally applicable. That is, by rule, we would expect that "angering" and "maddening" should be equally likely but the data show that there is asymmetry (cf. Figure 13). Quantitative corpus data allows us to measure the proportion of the application of a given inflectional rule to a certain set of words, the idea of Grammatical Profile (Janda, 2016).

Semantic Profiles of Experiencer Collocates

Pertaining to the second problem in this study, we identified the collocates that may refer to the Experiencer arguments of "angry" and "mad". The specific constructional, collocational pattern to identify this Experiencer is [*angry*/*mad* + NOUN] collocation; the NOUN slot will be further analysed to determine whether its filler refers to Experiencer of "angry" or "mad". This study utilised the

EJI (English Journal of Indragiri): Studies in Education, Literature, and Linguistics
Vol. 8. No. 2, July 2024
ISSN (Print). 2549-2144, ISSN (Online). 2589-5140
Homepage: https://ejournal-fkip.unisi.ac.id/index.php/eji/

Compare feature in COCA to generate right-side noun collocations from both words ("angry" was considered as WORD1 (W1) and "mad" as WORD2 (W2) in the Compare feature search interface). Table 1 shows a snippet of the output of the Compare feature for "angry" and "mad".

Table 1 Snippet of the (lightly edited) output of the Compare feature.

No	COLLOCATE	W1	W2	W1/W2	SCORE	WORD
1	BIRDS	755	1	755	580.3	angry
2	MOB	302	0	604	464.2	angry
3	CROWD	130	0	260	199.8	angry
...	...	...	...	...	...	...
63	GHOST	20	2	10	7.7	angry
64	KID	27	3	9	6.9	angry
...	...	...	...	...	...	...
80	BEAR	21	10	2.1	1.6	angry
81	WOMAN	103	54	1.9	1.5	angry
84	MEN	188	1106	0.2	0.1	angry

No	COLLOCATE	W2	W1	W2/W1	SCORE	WORD
1	SCIENTIST	411	0	822	1,069.50	mad
2	DASH	258	0	516	671.3	mad
3	HATTER	175	0	350	455.4	mad
...	...	...	...	...	...	...

The upper panel of Table 1, especially the green-highlighted cells, shows noun collocates that are strongly associated with "angry" while the lower panel is strongly associated with "mad". The color-coding of the output in COCA shows the direction and strength of the association. The greenish cell shows positive (typical) association of the collocates with the compared word in question; the white cells show neutral association while the reddish cell shows negative (atypical) association (e.g., the noun *men* is in reddish, pink cell as it is negatively associated with "angry"; this can be seen from the frequency with which *men* co-occur with "angry" [W1 column; 118 times] versus "mad" [W2 column; 1,106 times]). The important quantitative information is shown, among others, within the columns W1, W2, and SCORE. For instance, for the upper panel, the noun

birds co-occurs with "angry" (W1) for 755 times (in *angry birds* pattern) while *birds* only collocates once with "mad" (W2) (in the *mad birds* pattern). In contrast, in the lower panel, the noun collocate *scientist* never co-occurs with "angry" (W1; hence, the frequency of zero in W1 column) (in the *angry scientist* pattern) meanwhile the collocation *mad scientist* occurs 411 times.

The SCORE column indicates a value of association between the words in the COLLOCATE column and one of the compared words. This score reflected how strongly the words collocate with "angry' and "mad". The higher the value of SCORE, the stronger and more frequent the words collocate with each adjective. For instance, for "angry" in the upper panel, *birds*, *mob*, and *crowd* are the top-three most strongly associated right-side noun collocates (hence, Experiencer) for "angry". In contrast, while *ghost* and *kid* are also green-highlighted, thus positively associated with "angry", the strength of their association is lower as shown by their rankings (i.e., at the rank 63 and 64 respectively). In total there are 78 noun collocates that are green highlighted for (i.e., having positive association with) "angry". Meanwhile, there are only 32 noun collocates with green highlight for "mad".

The next step is qualitative, semantic analysis for the nouns inside the COLLOCATE column. These nouns were categorised into semantic field/types by referring to the *Concepticon* catalogue of semantic concepts from cross-linguistic data (List et al., 2024). Subsequently, the collocates that are considered as representing the Experiencer argument (and are green highlighted only) are filtered from the whole table and then visualised based on their co-occurrence frequencies (in logarithmic scale to the base of 10) with "angry" and "mad"; these frequencies are taken from the values within the W1 and W2 columns. The nouns were also lemmatised (i.e., *scientists* and *scientist* are brought under the same lemma, namely *scientist*). The visualisation, which could be thought of a semantic map of Experiencer for "angry" and "mad", is shown in Figure 16.

EJI (English Journal of Indragiri): Studies in Education, Literature, and Linguistics
Vol. 8. No. 2, July 2024
ISSN (Print). 2549-2144, ISSN (Online). 2589-5140
Homepage: https://ejournal-fkip.unisi.ac.id/index.php/eji/

Figure 16 Semantic map for the co-occurrence of noun collocate Experiencer with "angry" and "mad".

It is clear from Figure 16 that "angry" and "mad" differ in terms of their Experiencer collocate types. This is shown by the box shading manually overlaid on the collocates. For instance, the nouns *demonstrator*, *citizen*, *mob*, *crowd*, and others towards the lower right corner are strongly associated with "angry" but they have negative association with "mad". In contrast, "mad" attracts different type of Experiencer, which is not typical for "angry", namely *scientist*, *hatter*, *bomber*, *bastard* (towards the upper left corner with pink shading). Moreover, royal individual such as *queen* and *king* tend to be "mad" rather than "angry". The noun *man* is distinctive for "mad" while *people* and *person* are for "angry". Even though all the aforementioned distinctive collocates for "angry" and "mad" largely belong to the same semantic field (either as individual or group from the Social and political relation field), the specific entity is different (e.g., *angry citizen* is much more common than *mad citizen*, but *mad scientist* is much more common than *angry scientist*).

EJI (English Journal of Indragiri): Studies in Education, Literature, and Linguistics
Vol. 8. No. 2, July 2024
ISSN (Print). 2549-2144, ISSN (Online). 2589-5140
Homepage: https://ejournal-fkip.unisi.ac.id/index.php/eji/

A broader distinction based on the semantic field is shown by the predominance of Kinship-based nouns for "angry" than "mad". For "mad", only *bastard* belongs to the Kinship field that is attracted to it in the pattern *mad bastard*. In contrast, "angry" is strongly associated with many more Kinship nouns such as *mother*, *dad*, *kid*, *child*, *wife*, *parent* and so on (see Figure 16). Therefore, the pattern *angry mother/dad/kid/child* is more common than for "mad". Similarly, there are more Animal-based nouns attracted to "angry" (i.e., *bee*, *bird*, and *bull*) than "mad" (*dog* and *cow*).

Semantic Profiles of Stimulus Collocates

To identify the Stimulus collocates of "angry" and "mad", this study utilised the List feature in COCA. According to Kroeger (2004), a stimulus is the object of perception, cognition, or emotion. The constructional pattern in identifying the Stimulus is that the two adjectives followed by prepositions then the nouns, namely [angry/mad *at*/*about*/*over*/*with* NOUN]. Each prepositional variation of the construction for each "angry" and "mad" was searched for individually/independently in COCA List feature. COCA generated a maximum of one hundred sample. Figure 17 shows the snippet of the output.

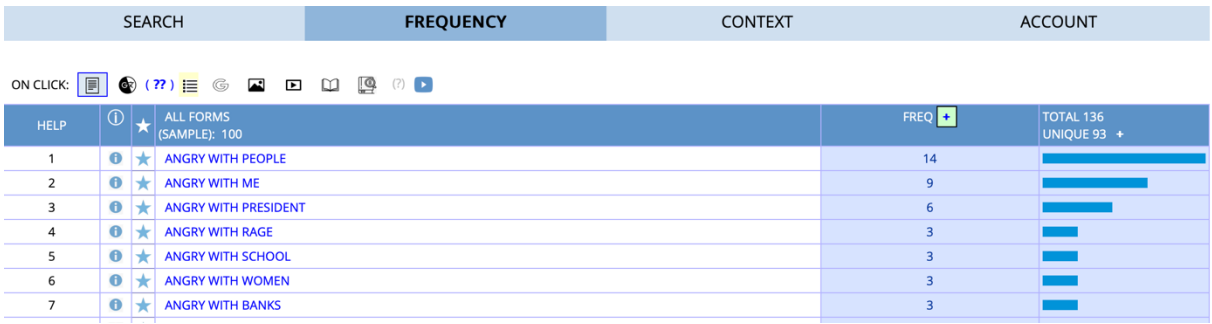


Figure 17 Snippet of the output of List feature after searching for the pattern [angry *with* NOUN].

Such an output was stored in Google Spreadsheet. For further analyses, we limited the pattern that has a frequency greater than three (values in the FREQ column in Figure 17). The qualitative analysis involved classifying the semantic field of the filler of the NOUN collocate slot in the construction. Furthermore, the

EJI (English Journal of Indragiri): Studies in Education, Literature, and Linguistics
Vol. 8. No. 2, July 2024
ISSN (Print). 2549-2144, ISSN (Online). 2589-5140
Homepage: https://ejournal-fkip.unisi.ac.id/index.php/eji/

noun collocates were lemmatised (such that *banks* and *bank* are normalised into *bank*). Quantitative analysis includes contrasting the log-frequency of the noun collocates in the combined constructional patterns between "angry" and "mad" and visualised it into a semantic map akin to Figure 16. Figure 18 firstly provides an overview of the proportion of the four constructional patterns across the two words.

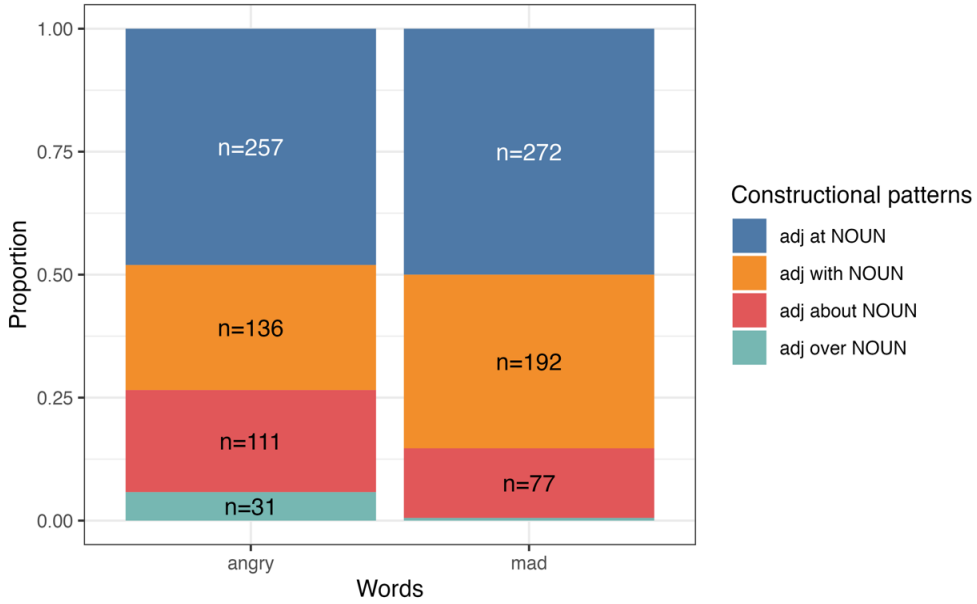


Figure 18 Proportion of the constructional patterns for the Stimulus collocates of "angry" and "mad".

From Figure 18, we can see that the patterns [angry *at* NOUN] and [mad *at* NOUN] are predominant for both words. However, "mad" has a higher relative frequency for the [adj *with* NOUN] pattern than "angry". The least frequent construction for the two adjectives is the [adj *over* NOUN] pattern. Next, Figure 19 below visualises the combined log-frequency of the NOUN collocates across the four constructional patterns with "angry" and "mad".

EJI (English Journal of Indragiri): Studies in Education, Literature, and Linguistics
Vol. 8. No. 2, July 2024
ISSN (Print). 2549-2144, ISSN (Online). 2589-5140
Homepage: https://ejournal-fkip.unisi.ac.id/index.php/eji/

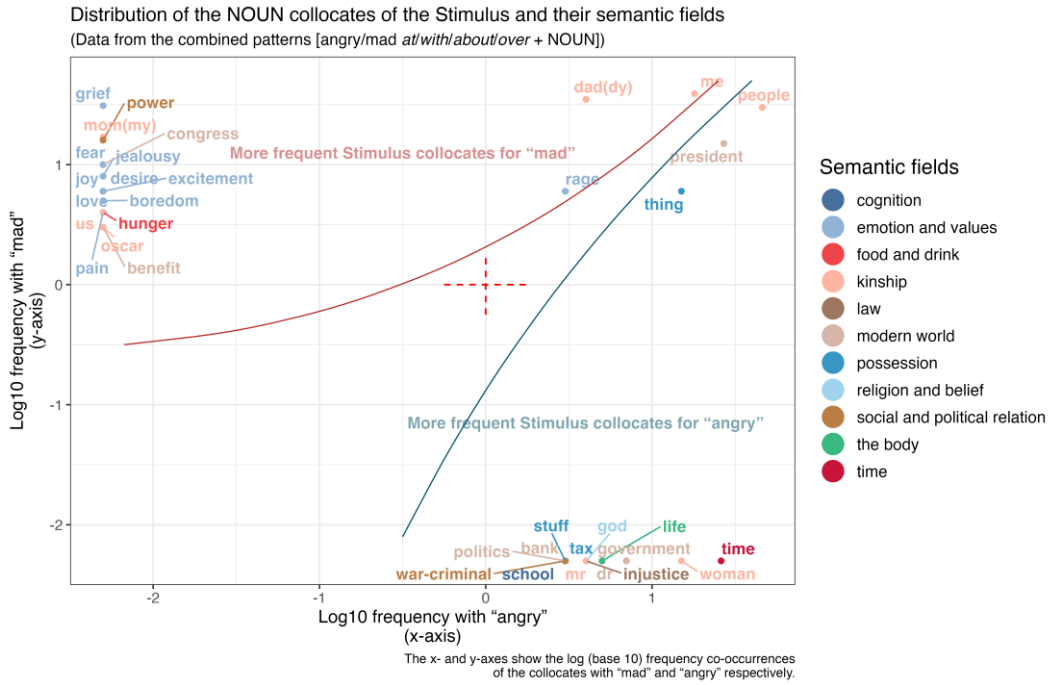


Figure 19 Semantic map for the co-occurrence of noun collocates Stimulus with "angry" and "mad".

The first clear pattern from Figure 19 is the predominance of collocates from the semantic field of Emotion and values (i.e., *grief*, *fear*, *jealousy*, *joy*, *desire*, *excitement*, *love*, *pain*, *rage*, and *boredom*) as the Stimulus for "mad". It suggests that "mad" is used for a more interconnected relationship with our cognition or emotions rather than "angry". What is more interesting is that the connection of "mad" with these emotion/values Stimulus appears in a single constructional pattern, namely [mad *with* NOUN] as shown in Table 2 below.

Table 2 The top expressions (frequency >=3) realising the [mad *with* NOUN] pattern referring to Emotion and Values.

NO	MAD *WITH* NOUN	FREQUENCY
1	MAD WITH GRIEF	31
2	MAD WITH POWER	16
3	MAD WITH FEAR	10
4	MAD WITH JOY	8
5	MAD WITH JEALOUSY	8
6	MAD WITH RAGE	6
7	MAD WITH LOVE	6
8	MAD WITH EXCITEMENT	6
9	MAD WITH DESIRE	6
10	MAD WITH BOREDOM	5

EJI (English Journal of Indragiri): Studies in Education, Literature, and Linguistics
Vol. 8. No. 2, July 2024
ISSN (Print). 2549-2144, ISSN (Online). 2589-5140
Homepage: https://ejournal-fkip.unisi.ac.id/index.php/eji/

11	MAD WITH HUNGER	4
12	MAD WITH PAIN	4
...	...	...
	TOTAL	192

In contrast to “mad”, the more frequent Stimulus collocates for “angry” are quite varied in terms of their semantic field (cf. Figure 19). A predominant Stimulus type comes from the Modern World field, especially related to the institutional domain, such as *politics*, *bank*, *government*, *dr.*, and *president*. Another typical Stimulus noun more frequent for “angry” belongs to the Possession field, namely the generic *stuff* and *thing*, and the more specific noun *tax*. In a way, *tax* itself is closely related to governmental institutional aspect that can trigger anger than madness. Finally, both “angry” and “mad” can have Stimulus from the Kinship-based nouns. However, those typical for “mad” tend to be more personal, such as *mom(my)*, *dad(dy)*, and even *me*. As in Figure 16, the semantic co-occurrence map in Figure 19 also demonstrates different semantic profiles of Stimulus typically co-occurring with “angry” and “mad”.

To gain a more detailed picture of the distribution of the Stimulus nouns with the constructional pattern, we visualised the frequency of the nouns with the constructions in Figure 20 below.

EJI (English Journal of Indragiri): Studies in Education, Literature, and Linguistics
Vol. 8. No. 2, July 2024
ISSN (Print). 2549-2144, ISSN (Online). 2589-5140
Homepage: https://ejournal-fkip.unisi.ac.id/index.php/eji/

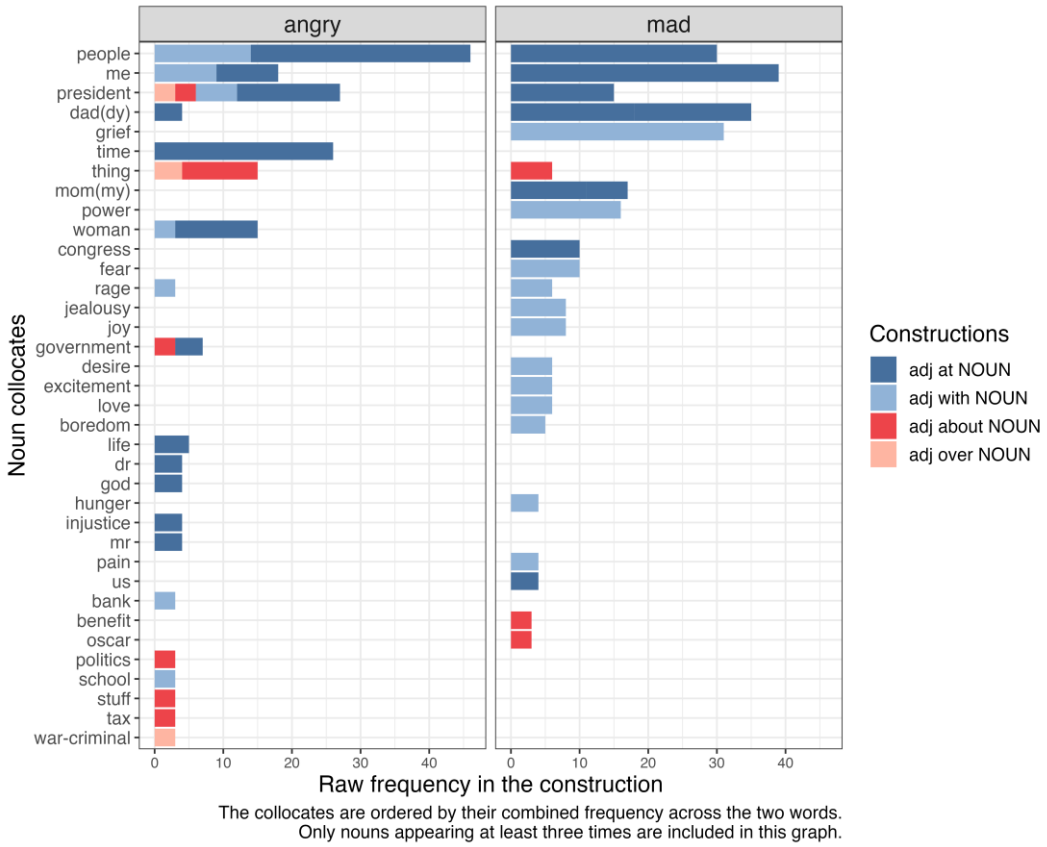


Figure 20 Distribution of the Stimulus collocate by construction.

To illustrate the interpretation of Figure 20, we can see for instance that the top-three nouns namely *people*, *me*, and *president* do occur with the two words. However, their co-occurrence with "mad" is only within the [adj *at* NOUN] construction. Meanwhile, the three nouns can appear with "angry" in more than one constructional pattern type, such as the noun *president*, which can appear in all the four constructions, though most frequently with [adj *at* NOUN]. Nevertheless, we can see other noun collocates that appear only with either "angry" or "mad" and within one construction type (e.g., *congress* and *mom(my)* only appear with "mad" in the [adj *at* NOUN] construction).

CONCLUSIONS

The purpose of this study was to investigate how morphological (the grammatical) and semantic profiles could help distinguish two words that are semantically similar. In terms of the grammatical profile, our first research question, we focus on the frequency of the inflectional verbal forms for the words (see §3.2 for detailed results). The findings indicate that the *-ing* form is the more predominant profile for the verbal lemma MADDEN than for ANGER (cf. Figure 13). In contrast, ANGER dominates the other inflectional forms (i.e., the base, *-ed*, and the *-s* third-person singular forms) compared to MADDEN. As for the second research question, we aimed at analysing the (semantic field of the) noun collocates representing the Experiencer and Stimulus arguments in a predicate headed by "angry" and "mad" (§3.3-§3.4). For the Experiencer, both "angry" and "mad" can attract nouns referring to the Social and Political Relation semantic field, however the specific nouns are different (cf. Figure 16). Furthermore, "anger" attracts more Experiencer nouns belonging to Kinship than "mad". Regarding the Stimulus, we found that "mad" co-occurs mostly with Emotion and Values noun collocates (Figure 19). On the contrary, the predominant Stimulus type for "anger" comes from the Modern World field, especially related to the institutional domain, such as *politics*, *bank*, *government*, *dr.*, and *president*.

These findings suggested that near-synonymous words such as *angry* and *mad* demonstrate distinct grammatical/morphological and semantic profiles. These differences can be revealed by inspecting large amount of textual data and applying qualitative and quantitative methods, as in corpus linguistic method, to detect usage patterns for how these words are used. Future research can build on this study by inspecting different emotion near-synonyms to determine the extent to which their grammatical and semantic profiles differ.

ACKNOWLEGMENT

We would like to thank the anonymous reviewers for their constructive feedback to this paper. The remaining flaws are our own. This work is the outcome of an undergraduate thesis project by the first author within the

CompLexico (Computer-assisted Lexicology and Lexicography) research group (https://www.cirhss.org/complexico/) in the *Centre for Interdisciplinary Research on the Humanities and Social Sciences* (CIRHSS) at Udayana University.

REFERENCES

Dziwirek, K., & Lewandowska-Tomaszczyk, B. (Eds.). (2010). *Complex emotions and grammatical mismatches: A contrastive corpus-based study*. De Gruyter Mouton.

Effendi, I. S., Amalia, R., & Lalita, S. A. (2020). Are Synonyms Always Synonymous? A Corpus-assisted Approach to Announce, Declare, and State. *ASIAN TEFL Journal of Language Teaching and Applied Linguistics*, *5*(1). https://doi.org/10.21462/asiantefl.v5i1.110

Gries, S. Th. (2009). What is Corpus Linguistics? *Language and Linguistics Compass*, *3*(5), 1225–1241. https://doi.org/10.1111/j.1749-818X.2009.00149.x

Gumelar, A. (2018). *A CORPUS BASED ANALYSIS OF LINGUISTIC FEATURES IN HEAVY METAL SONGS FROM 2000-2018*. Brawijaya University.

Harta, I. M. L., Rajeg, G. P. W., & Rahayuni, N. K. S. (2023). Distinguishing semantic preference of *hurry* and *rush* via collocational patterns: A corpus-based study. *International Journal of Education, Language, Literature, Arts, Culture, and Social Humanities*, *1*(3), 1–12.

Janda, L. A. (2016). Linguistic profiles: A quantitative approach to theoretical questions. *Język i Metoda*, 127–145.

Janda, L. A., & Lyashevskaya, O. (2011). Grammatical profiles and the interaction of the lexicon with aspect, tense, and mood in Russian. *Cogl*, *22*(4), 719–763. https://doi.org/10.1515/cogl.2011.027

Kövecses, Z. (2000). The concept of ANGER: Universal or culture specific? *Psychopathology*, *33*, 159–170.

Kreidler, C. (1998). *Introducing English Semantics*. Routledge.

List, J., Tjuka, A., van Zantwijk, M., Blum, F., Barrientos Ugarte, C., Rzymski, C., Greenhill, S., & Forkel, R. (2024). *CLLD Concepticon 3.2.0* (3.2.0) [dataset]. https://doi.org/10.5281/zenodo.596412

Liu, M. (2016). Emotion in lexicon and grammar: Lexical-constructional interface of Mandarin emotional predicates. *Lingua Sinica*, *2*(1), 4. https://doi.org/10.1186/s40655-016-0013-0

Phitayakorn, W. (2017). *A CORPUS-BASED STUDY OF ENGLISH SYNONYMS: ADVISE, RECOMMEND, AND SUGGEST*. Thamsat University.

Rajeg, G. P. W. (2020). Tutorials for Corpus of Contemporary American English (COCA). *Figshare*. https://doi.org/10.6084/m9.figshare.12362756

Suari, I. A. S. T., Rajeg, G. P. W., & Sudipa, I. N. (2024). *Supplementary materials for "Exploring Grammatical and Semantic Profiles of ANGRY and MAD: A Corpus-based Study"* (v1.0.0) [dataset]. Zenodo. https://doi.org/10.5281/zenodo.11108458

Xiao, R., & McEnery, T. (2006). Collocation, Semantic Prosody, and Near Synonymy: A Cross-Linguistic Perspective. *Applied Linguistics*, *27*(1), 103–129. https://doi.org/10.1093/applin/ami045