

Decoding “Word of the Year”: Analyzing Words of Five Categories Spanning 2004-2022

Shiman Xu^{1,*}

¹*School of International Studies, Zhejiang University, Hangzhou, China*

*Corresponding author: xushiman1019@foxmail.com

Abstract: *The paper begins with an introduction to the concept of "Word of the Year" and its significance in tracking linguistic and cultural shifts. It emphasizes categorizing chosen words into specific thematic areas to understand prevailing trends. It focuses on five categories: Internet/Social Media, Political and Economic, Social and Cultural, Environmental, and Technological. The analysis section examines each category individually, providing insights into the vocabulary trends, frequency of word selections, and notable shifts. Statistical analysis and graphical representations are used to present cumulative counts, proportions, and regression analysis for each category, facilitating comprehensive comparisons across the chosen time frame. The findings shed light on linguistic and cultural trends that have shaped the English language. The paper concludes with a discussion of the broader implications and future prospects of "Word of the Year" studies, emphasizing their role in understanding the dynamic nature of language and its intersection with society.*

Keywords: “Word of the Year”, Linguistic Analysis, Dictionary, Societal Trends, English Language

1. Introduction

Words serve as invaluable tools for conveying assertions, ideas, aspirations, and uncertainties. However, they can also become battlegrounds for ideological conflicts. The tradition of selecting the "Word of the Year" has permeated linguistics, lexicography, and language organizations worldwide. This practice involves identifying a single word that encapsulates the essence, prevailing trends, and consequential events of a given year. It offers insights into the dynamic nature of language and its profound interconnection with societal development, both globally and locally. This paper aims to examine "Word of the Year" selections in English, alongside analogous practices in other languages, by scrutinizing the historical context and selection processes employed by different dictionaries and language organizations. Furthermore, it seeks a comprehensive understanding of the ever-evolving social, cultural, and political landscape that shapes our languages and societies and will shed light on shared global trends and regional idiosyncrasies spanning the period from 2004 to 2022.

2. Literature Review

An extensive literature search has been conducted using keywords including "Word of the Year," "linguistic analysis," "dictionary," and "societal trends". Notable academic contributions include Richard M. Rollins' "Words as Social Control: Noah Webster and the Creation of the American Dictionary" (1976, *American Quarterly*) and Grace M. Burton's "Word Problems—a Mirror of Society" (1974, *The Arithmetic Teacher*). Additionally, domestic studies such as "The Use and Performance of 'Word of the Year' in Hong Kong-style Chinese: A Case Study of the Top Ten New Words and Catchphrases in 2022" (Hong Shuang, 2023), "Yearly Buzzwords, Vividly Capturing Social Life" (Peng Xunwen, 2023), and "The Analysis of Collins Dictionary: Words of 2021 under Conceptual Blending Theory" (Bu Yiming, 2022) have provided valuable insights. Observations indicate a relative scarcity of recent scholarly attention from foreign researchers in this field, while domestic research has primarily focused on specific dictionaries or the compilation and summarization of annual buzzwords within the Chinese discourse. Others have adopted the conceptual integration theory to study yearly buzzwords.

3. The Definition of "Word of the Year"

The word(s) of the year refers to any of various assessments as to the most important word(s) or

expression(s) in the public sphere during a specific year. Sometimes, the "Word of the Year" is judged to reflect the ethos, mood, or preoccupations of that particular year and to have lasting potential as a word of cultural significance.

4. "Word of the Year" Across Different Dictionaries and Languages

4.1. "Word of the Year" in English

4.1.1. Paper Dictionaries and Online Dictionaries with a Global User Base

International standard dictionaries serve as globally recognized lexicons, providing established and widely accepted linguistic references for English learners worldwide. These prominent language resources are the cornerstone for English language acquisition and comprehension on a global scale, including *Cambridge Dictionary*, *Collins English Dictionary*, *Webster's Dictionary*, and *Oxford English Dictionary*.

Webster's Dictionary, published by Merriam-Webster, Inc., the oldest dictionary publisher in the United States [1]. In 1996, Merriam-Webster launched its first website, which provided free access to an online dictionary and thesaurus [2]. The lists of Merriam-Webster's Words of the Year (for each year) feature the ten words of the year from the English language. These word lists started in 2003 and have been published at the end of each year. At first, Merriam-Webster determined its contents by analysing page hits and popular searches on its website. Since 2006, the list has been determined by an online poll and by suggestions from visitors to the website [3]. From 2008 onwards, however, user submissions have not been a deciding factor. Merriam-Webster's Word of the year is determined by data: the word must be frequently looked up at Merriam-Webster.com, and it must show a significant increase in lookups over the previous year.

Oxford English Dictionary is published by Oxford University Press. British prime minister Stanley Baldwin described the OED as a "national treasure"[4]. The candidates for the Word of the Year are drawn initially from the Oxford Dictionaries New Monitor Corpus, a research programme which collects around 150 million words of current English in use each month, using automated search criteria to scan new web content. The final "Word of the Year" selection team is made up of lexicographers and consultants to the dictionary team, and editorial, marketing, and publicity staff [5].

Collins English Dictionary is compiled by a team of experienced linguists and lexicographers. The Collins Word of the Year is also not restricted to UK language usage, and words are often chosen that apply internationally as well. The Collins Words of the Year are selected by the Collins Dictionary team across Glasgow and London, consisting of lexicographers, editorial, marketing, and publicity staff, though previously the selection process has been open to the public.

Cambridge Dictionary is widely regarded as a prestigious and highly reputable English dictionary, published by Cambridge University Press. Its authority stems from its strong academic background and linguistic research heritage. The Cambridge "Word of the Year" is led by the data of what users look up.

4.1.2. Paper Dictionaries and Online Dictionaries with a Regional or Specific User Base

Specialized regionalized dictionaries and locally targeted lexicons serve as authoritative language references, catering to specific linguistic communities and providing noteworthy insights into regional language usage, including the American Dialect Society (ADS), Australian National Dictionary Centre, and *Macquarie Dictionary*.

The American Dialect Society (ADS), founded in 1889, is a learned society "dedicated to the study of the English language in North America, and of other languages, or dialects of other languages, influencing it or influenced by it" [6]. It is an association which in its first constitution defined its objective as "the investigation of the spoken English of the United States and Canada" (*Constitution*, 1890). Over the years, its objective has remained the same, only expanded to encompass "the English language in North America, together with other languages or dialects of other languages influencing it or influenced by it" [7]. Since 1991, the American Dialect Society has designated one or more words or terms to be the word of the year. In addition to the "Word of the Year", the society also selects words in other categories that vary from year to year.

The Australian National Dictionary Centre (ANDC) is a major centre for lexicographical research in Australia [8], conducting research into Australian English, which comprises the varieties of the English language native to Australia. By the 1820s, the native-born colonists' speech was recognisably distinct

from speakers in Britain and Ireland [9]. The Australian National Dictionary Centre has announced a Word of the Year each since 2006. The word is chosen by the editorial staff and is selected on the basis of having come to some prominence in the Australian social and cultural landscape during the year [10]. The Word of the Year is often reported in the media as being Australia's word of the year, but the word is not always an Australian word.

The Macquarie Dictionary is a dictionary of Australian English, which is generally considered by universities and the legal profession to be the authoritative source on Australian English. Each year the editors select a short-list of new words added to the dictionary and invite the public to vote on their favourite. The public vote is held in January and results in the People's Choice winner. There is also a word selected by a committee.

4.2. "Word of the Year" in Other Languages

In fact, in addition to English-speaking countries, many other nations across the globe also engage in the annual selection of words that represent their respective countries and languages. These chosen words serve as linguistic snapshots, reflecting the collective concerns, focal points, and societal hot topics of the particular year among their nation's citizens and language users. For instance, in the year 2014, when choosing the "Word of the Year", Chinese replied "law", Japanese replied "duty", German replied "Lichtgrenze" and Singaporean replied "chaos".

5. Analysis of "Word of the Year" in English from 2004 to 2022

Initially, a comprehensive dataset including selections from four renowned dictionaries spanning 2004-2022 was compiled to delineate the temporal evolution of the "Word of the Year" phenomenon in English. These dictionaries comprise the *Cambridge Dictionary*, *Collins English Dictionary*, *Webster's Dictionary*, and *Oxford English Dictionary*. (see Figure 1)

WOTY		2004	2005	2006	2007	2008	2009	2010	2011	2012	
<i>Cambridge Dictionary</i>											
<i>Collins English Dictionary</i>											
<i>Webster's Dictionary</i>		blog	integrity	truthiness	w00t	bailout	admonish	austerity	pragmatic	socialism and capitalism	
<i>Oxford English Dictionary</i>	UK	chav	sudoku	bovvered	carbon footprint	credit crunch	simples	big society	squeezed middle	ommishambles	
	US	podcast	carbon-neutral	locavore	hypemiling	unfriend	refudiate			GIF (noun)	
WOTY		2013	2014	2015	2016	2017	2018	2019	2020	2021	2022
<i>Cambridge Dictionary</i>				austerity	paranoid	populism	nomophobia	upcycling	quarantine	perseverance	homer
<i>Collins English Dictionary</i>		Geek	Photobomb	Binge-watch	Brexit	Fake news	Single-use	Climate Strike	Lockdown	NFT	Permacrisis
<i>Webster's Dictionary</i>		science	culture	*-ism	surreal	feminism	justice	they	pandemic	vaccine	gaslighting
<i>Oxford English Dictionary</i>	UK	selfie	vape	👉	post-truth	youthquake	toxic	climate emergency	No single word chosen	vax	goblin mode
	US										

Figure 1: "Word of the Year" in English from 2004 to 2022

Secondly, all words selected as the "Word of the Year" were meticulously traced back to authoritative websites and resources. Each vocabulary entry was thoroughly researched to uncover its lexical significance at the time of designation. Comprehensive information regarding the rationale behind their selection as the "Word of the Year" was gathered, along with the corresponding social context and background that influenced their prominence.

Thirdly, after comprehending the semantics and contexts of all entries, five classifications have been conducted, including Internet/Social Media Category, Political and Economic Category, Social and Cultural Category, Environmental Category, and Technological Category. Internet/Social Media Category encompasses terms related to online communication, digital technologies, social networking sites, viral trends, and the impact of social media on society. Political and Economic Category includes vocabulary entries concerning politics, government, political ideologies, elections, policymaking, international relations, economic trends, and financial markets. Social and Cultural Category reflects societal and cultural phenomena, trends, and movements, including terms related to societal issues, cultural shifts, identity, diversity, and popular culture. Environmental Category pertains to environmental issues, sustainability, climate change, environmental conservation, renewable energy, ecological awareness, and the impact of human activities on the planet. Technological Category comprises vocabulary entries associated with technological advancements, innovations, digital transformations, emerging technologies, scientific discoveries, artificial intelligence, robotics, and the digital revolution. Spanning 2004 to 2022, the cumulative vocabulary counts for each category in the Word of the Year selections are as follows: 7 for Internet/Social Media, 13 for Political and Economic, 32 for Social and Cultural, 8 for Environmental, and 3 for Technological. Notably, Social and Cultural boasts the highest

aggregate entries, with Political and Economic ranking second. These two major categories significantly exceed the lexical volume of the remaining three.

Following the delineation of the five classifications, a comprehensive analysis was undertaken by computing the relative frequency of "Word of the Year" selections within each category for each year. This involved dividing the count of words belonging to a specific category by the total count of "Word of the Year" selections for that particular year. (See Figure 2)

Word of the Year	Internet/Social Media Category	Political and Economic Category	Social and Cultural Category	Environmental Category	Technological Category	Total number
2004	1/2	0	1/2	0	0	1
2005	0	0	2/3	0	1/3	1
2006	0	0	2/3	1/3	0	1
2007	0	0	1/3	2/3	0	1
2008	0	2/3	0	1/3	0	1
2009	1/3	1/3	1/3	0	0	1
2010	0	1	0	0	0	1
2011	0	0	1	0	0	1
2012	1/3	2/3	0	0	0	1
2013	1/3	0	1/3	0	1/3	1
2014	1/3	0	2/3	0	0	1
2015	1/4	1/4	1/2	0	0	1
2016	0	1/2	1/2	0	0	1
2017	0	1/2	1/2	0	0	1
2018	1/4	0	1/2	1/4	0	1
2019	0	0	1/4	3/4	0	1
2020	0	0	1	0	0	1
2021	0	0	3/4	0	1/4	1
2022	0	0	1	0	0	1

Figure 2: Relative frequency of 'Word of the Year' selections by category

Based on the findings, in order to facilitate an in-depth analysis and comparative examination, it is logical to bifurcate these five categories accordingly. Figure 3 reflects the variation tendency of "Word of the Year" in Social and Cultural Category and Political and Economic Category, while Figure 4 offers information of variation tendency of "Word of the Year" in other three categories.

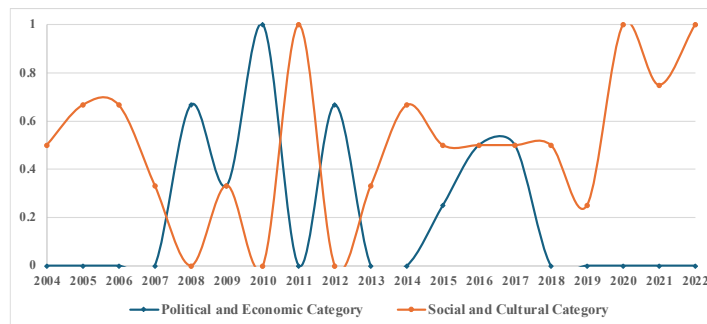


Figure 3: 'Word of the Year' variation tendency in social/cultural & political/economic categories

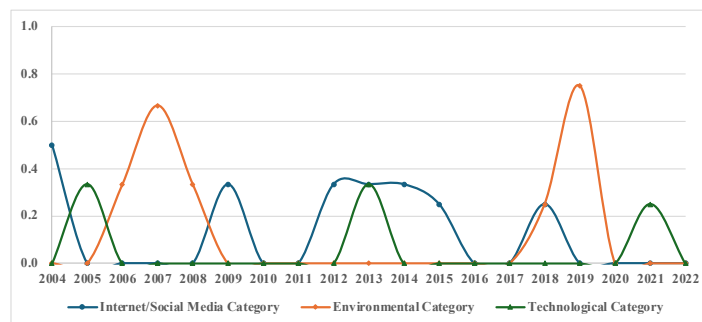


Figure 4: 'Word of the Year' variation tendency in other three categories

Since line graph analysis coupled with linear regression presents advantages including the identification of trends, patterns, and relationships within the data. In the context of this study, it allows for a comprehensive exploration of linguistic trends and patterns over time and anticipate language changes. Through the utilization of line graph analysis, the findings are presented as follows. (See Table 1)

For the Political and Economic Category (Y), the R^2 value of 0.019 suggests a weak relationship between the Word of the Year (X) and this category. The negative regression coefficient (-0.008) indicates a slight decrease in the Political and Economic Category as the Word of the Year increases. On the other hand, for the Social and Cultural Category (Y), the R^2 value of 0.126 indicates a moderate relationship

with the Word of the Year (X). The positive regression coefficient (0.02) suggests a slight increase in the Social and Cultural Category as the Word of the Year increases. Regarding the Internet/Social Media Category (Y), the R^2 value of 0.055 suggests a weak relationship with the Word of the Year (X). The negative regression coefficient (-0.007) implies a slight decrease in the Internet/Social Media Category as the Word of the Year increases. For the Environmental Category (Y), the R^2 value of 0.009 indicates a weak relationship with the Word of the Year (X). The negative regression coefficient (-0.004) suggests a slight decrease in the Environmental Category as the Word of the Year increases. Lastly, for the Technological Category (Y), the R^2 value of 0.003 indicates a very weak relationship with the Word of the Year (X). The negative regression coefficient (-0.001) implies a negligible decrease in the Technological Category as the Word of the Year increases.

Table 1: Line graph analysis

i. X: {Year}; Y: {Political and Economic Category}	ii. X: {Year}; Y: {Social and Cultural Category}
1. R^2 : 0.019	1. R^2 : 0.126
2. Regression coefficients: -0.008 (X) & 15.51 (constant)	2. Regression coefficients: 0.02 (X) & -39.525 (constant)
3. Significance p-values for F-test: 0.576 (X) & 0.5712 (constant)	3. Significance p-values for F-test: 0.136 (X) & 0.141 (constant)
4. Model formula: $y = 15.51 - 0.008 * \text{Year}$	4. Model formula: $y = -39.525 + 0.02 * \text{Year}$
iii. X: {Year}; Y: {Internet/Social Media Category}	iv. X: {Year}; Y: {Environmental Category}
1. R^2 : 0.055	1. R^2 : 0.009
2. Regression coefficients: -0.007 (X) & 14.543 (constant)	2. Regression coefficients: -0.004 (X) & 8.069 (constant)
3. Significance p-values for F-test: 0.334 (X) & 0.330 (constant)	3. Significance p-values for F-test: 0.702 (X) & 0.698 (constant)
4. Model formula: $y = 14.543 - 0.007 * \text{Year}$	4. Model formula: $y = 8.069 - 0.004 * \text{Year}$
v. X: {Year}; Y: {Technological Category}	
1. R^2 : 0.003	2. Regression coefficients: -0.001 (X) & 2.403 (constant)
3. Significance p-values for F-test: 0.817 (X) & 0.813 (constant)	4. Model formula: $y = 2.403 - 0.001 * \text{Year}$

Upon analysing these results, it becomes evident that the Social and Cultural Category has consistently dominated the Word of the Year selections, indicating its significant influence and relevance over the given time span. The category's extensive vocabulary entries suggest a strong connection to societal and cultural phenomena that have captured public attention and shaped discourse. On the other hand, the Political and Economic Category emerges as the second most prominent category. This suggests that political and economic issues have also played a significant role in shaping language and discourse during the studied period. To be more specific, as for "Word of the Year" categorized in "Social and Cultural Category", the line graph reveals two distinct periods with relatively higher values, accompanied by slight fluctuations in the middle. However, the overall average level remains relatively high, indicating a potential upward trend in the future. As for "Political and Economic Category", the line graph illustrates three distinct peaks in the middle, a relatively stable phase, and lower values at both ends. The future trend appears uncertain, with no clear direction. In terms of the "Environmental Category", there are two prominent peaks in the line graph. The first peak occurs around 2007, while the second peak occurs around 2019. In the remaining time periods, the vocabulary count is almost negligible or close to zero. Similarly, as for "Technological Category", there are three minor peaks in total, occurring in 2005, 2013, and 2021, respectively. In the remaining time periods, the vocabulary count within this category is zero. For the "Technological Category", the average value of the vocabulary count within the mentioned category is higher than the previous two categories. However, there is an overall downward trend in the total vocabulary count.

To enhance the credibility of the data, the ARIMA model was utilized for conducting time series analysis. The ensuing results are presented below. (See Table 2)

For the first category, the t-values for the coefficients were 2.097 (0), 1.843 (1), and -2.87 (2), indicating the respective statistical significance of the coefficients. The significance p-values for the F-test were 0.999 (0), 0.998 (1), and 0.049** (2), suggesting a significant relationship between the independent variables and the dependent variable. Regarding the second category, the t-value for the coefficient was 1 (0), suggesting limited statistical significance. The significance p-value for the F-test was 0.994 (0), indicating a lack of overall significance in the model. Moving on to the third category, the

t-values for the coefficients were -0.307 (0), -0.84 (1), and -1.744 (2), suggesting limited statistical significance for the respective variables. The significance p-values for the F-test were 0.925 (0), 0.807 (1), and 0.409 (2), indicating a lack of overall significance. In the fourth category, the t-values for the coefficients were 1.242 (0), 0.17 (1), and -4.041 (2), indicating varying levels of statistical significance. The significance p-values for the F-test were 0.996 (0), 0.971 (1), and 0.001*** (2), suggesting a significant relationship between the independent variables and the dependent variable. Finally, for the fifth category, the t-value for the coefficient was 1 (0), suggesting marginal statistical significance. The significance p-value for the F-test was 0.994 (0), indicating a lack of overall significance.

Table 2: ARIMA model

i. Time series data: {Social and Cultural Category}; Time variable: Year	ii. Time series data: {Internet/Social Media Category}; Time variable: Year
1. R ² : 0.0 2. t-values: 2.097 (0) & 1.843 (1) & -2.87 (2) 3. Significance p-values for F-test: 0.999 (0) & 0.998 (1) & 0.049** (2)	1. R ² : 0.0 2. t-value: 1 (0) 3. Significance p-value for F-test: 0.994 (0)
iii. Time series data: {Political and Economic Category}; Time variable: Year	iv. Time series data: {Environmental Category}; Time variable: Year
1. R ² : 0.0 2. t-values: -0.307 (0) & -0.84 (1) & -1.744 (2) 3. Significance p-values for F-test: 0.925 (0) & 0.807 (1) & 0.409 (2)	1. R ² : 0.0 2. t-values: 1.242 (0) & 0.17 (1) & -4.041 (2) 3. Significance p-values for F-test: 0.996 (0) & 0.971 (1) & 0.001*** (2)
v. Time series data: {Environmental Category}; Time variable: Year	
1. R ² : 0.0 2. t-value: 1 (0) 3. Significance p-value for F-test: 0.994 (0)	Note: *** (3 asterisks), ** (2 asterisks), and * (1 asterisk) denote significance levels of 1%, 5%, and 10%, respectively.

6. Conclusions

In conclusion, based on the first analysis, the R² values for all categories are low, ranging from 0.003 to 0.126, indicating that the year variable explains only a small portion of the variations in the categories. The regression coefficients and significance p-values for the independent variable (year) and constant term suggest that the relationship between the categories and the year variable is not statistically significant in most cases. As for the ARIMA model, the R² values for all categories are 0.0, indicating that the variations in the categories cannot be explained by the year variable alone. Additionally, the t-values and significance p-values suggest that there is no significant relationship between the categories and the year variable.

Based on these findings, the conclusion of the paper could state that the analyzed data does not provide evidence of a significant relationship or trend between the examined categories (Social and Cultural, Internet/Social Media, Political and Economic, Environmental, and Technological) and the year variable. The absence of significant associations between these categories and time suggests that other factors may play a more influential role in determining the observed variations. Further research and exploration of additional variables may be necessary to gain a deeper understanding of the dynamics and factors affecting these categories.

Despite the absence of discernible trends or significant correlations between the number of "Words of the Year" in various categories and the temporal dimension, as well as the limited predictive power of such data, a holistic examination reveals that societal attention is most strongly directed towards social and cultural matters. This enduring focus is expected to persist as a primary area of interest in the foreseeable future. Conversely, the quantity of annual keywords associated with the political and economic domain is subject to considerable fluctuation and uncertainty, as it is influenced by contemporaneous events and influential personalities. Consequently, the dynamics of this category display high volatility and variability, as evidenced by recent observations. Moreover, the remaining three

categories share a distinctive characteristic of exhibiting sporadic peaks in specific years, indicating heightened attention from the public. The values assigned to these categories demonstrate close associations with societal hot topics and popular phenomena prevalent during those respective years. They serve as societal snapshots, enabling relevant authoritative institutions, leaders, and governments to gauge and address prevailing public concerns and aspirations effectively. Consequently, such insights facilitate the provision of more targeted and relevant support and assistance.

To sum up, while the analyzed data reveals no statistically significant relationships between the number of "Words of the Year" and the temporal dimension across the examined categories, it is evident that societal attention primarily revolves around social and cultural dimensions. This focus is expected to persist in the coming years. Attention towards the political and economic domain is characterized by volatility, influenced by current events and prominent figures. The sporadic peaks observed in the remaining three categories provide snapshots of societal interests, shedding light on pressing concerns and informing decision-makers regarding the preferences and needs of the populace.

References

- [1] Correa, Carla (November 3, 2021). "Attention, New Englanders: Fluffernutter Is Now a Word". *The New York Times*. ISSN 0362-4331. Retrieved November 5, 2021.
- [2] "Timeline: Merriam-Webster Milestones", Merriam-Webster, archived from the original on January 13, (2015), retrieved October 14, 2018
- [3] "Merriam-Webster launches 'Word of the Year' online poll". CNET. 2007-11-27. Retrieved 2007-12-29.
- [4] Skapinker, Michael (21 December 2012). "Well-chosen words". *Financial Times*. Archived from the original on 10 December 2022. Retrieved 3 June 2018.
- [5] *Oxford Dictionaries Word of the Year: Frequently Asked Questions* (viewed November 20, 2013).
- [6] Flexner, Stuart B. (December 15, 1985). "One language, highly divisible". *The New York Times*. Retrieved February 19, 2008.
- [7] Auroux, Sylvain (2006). *History of the Language Sciences: an international handbook on the evolution of the study of language from the beginnings to the present*. p. 2366. ISBN 3-11-016736-0.
- [8] "The Australian National Dictionary Centre | National Museum of Australia". www.nma.gov.au. Retrieved 14 October 2017.
- [9] Burridge, Kate (2020). "Chapter 11: History of Australian English". In Willoughby, Louisa (ed.). *Australian English Reimagined: Structure, Features and Developments*. Routledge. p. 181. ISBN 978-0-367-02939-5.
- [10] "Australian National Dictionary Centre's Word of the Year (2016) | Ozwords". ozwords.org. Retrieved October 14, 2017.