

***Lose one's life* and *lose one's job*
with singular *they*: two constructions,
two regional varieties, many practical aspects
of working with mega-corpora**

Lose one's life [stracić życie] i *lose one's job* [stracić pracę]
z zaimkiem *they* odnoszącym się do liczby pojedynczej:
dwie konstrukcje, dwa warianty regionalne języka angielskiego,
wiele praktycznych aspektów pracy z dużymi korpusami tekstów


Karolina RUDNICKA¹

University of Gdańsk

Abstract

The paper compares the usage of singular *they* with two morphologically similar constructions in British and American English. The constructions in question are *lose one's life* and *lose one's job*. The results obtained suggest that singular *they*, at least used with the two constructions in focus of this work, seems to be more widely used in the American variety of English than in the British variety. An additional aim of this work is to present and discuss some practical aspects of working with mega-corpora. The work shows how and where quantitative language studies need to be accompanied by manual and qualitative investigations. The corpora used in this work are the British National Corpus (BNC) and the Corpus of Contemporary American English (COCA).

Keywords: Singular *they*, corpus linguistics, mega-corpora, construction grammar, British English, American English, pronoun

¹  <https://orcid.org/0000-0001-8097-6086>

University of Gdańsk, Faculty of Languages

karolina.rudnicka@ug.edu.pl

Streszczenie

Artykuł porównuje użycie zaimka *they* w odniesieniu do podmiotów w liczbie pojedynczej z dwiema, z punktu widzenia morfologicznego podobnymi do siebie, konstrukcjami w wariancie brytyjskim i amerykańskim języka angielskiego. Badane konstrukcje to *lose one's life* [stracić życie] i *lose one's job* [stracić pracę]. Uzyskane wyniki sugerują, że *they* odnoszące się do liczby pojedynczej jest bardziej rozpowszechnione w amerykańskim wariancie języka angielskiego niż w brytyjskim. Dodatkowym celem artykułu jest zaprezentowanie i przedyskutowanie praktycznych aspektów pracy z dużymi korpusami językowymi. Praca wskazuje w których miejscach badania ilościowe muszą iść w parze z badaniami jakościowymi. Korpusy zastosowane do wykonania badania to Brytyjski Korpus Narodowy (BNC) i Amerykański Korpus Współczesnego Języka Angielskiego (COCA).

Słowa kluczowe: *They* odnoszące się do liczby pojedynczej, językoznawstwo korpusowe, korpusy językowe, gramatyka konstrukcyjna, brytyjski angielski, amerykański angielski, zaimek

1. Introduction: construction grammar, singular *they* and mega-corpora

This work subscribes to construction grammar, and the phrases which are being looked at are termed as constructions. According to Hilpert (2014: 22), *constructions* are “linguistic generalizations that speakers internalize”. Linguistic approaches which see morphemes, words, idioms and abstract phrasal patterns as *constructions*, do not accept a clear-cut division between the syntax and the lexicon, but assume that all constructions belong to a lexicon-syntax continuum – the so-called *construction* (Hoffmann & Trousdale 2013: 1). The two constructions studied here are *lose one's life* and *lose one's job*. They, without doubt, are morphologically similar – both have a verb + possessive pronoun + object order, the verb is in both cases the same one, *to lose*. Also, the situations to which the constructions refer, bear similarities, namely the possessor of an object (the doer) is deprived of the object, the object being in one case – the *life* of the subject, and, in another case, the *job* of the subject.

An interesting aspect is the fact that, according to Cambridge Dictionary Online, *lose one's life* is an idiomatic construction. There is a separate entry devoted to the phrase², however, no such entry is devoted to the construction *lose one's job*. Still, if one consults other online dictionaries, there are some that also list the latter construction as an idiom, for instance Merriam-

² Cambridge Online Dictionary, s.v. *lose your life*, retrieved on May 17, 2020 from <https://dictionary.cambridge.org/dictionary/english/lose-your-life>.

Webster³. The question as for whether one expression is more idiomatic than the other, remains open and, very likely, depends on the context of the concrete utterance. Sentences (1) and (2) exemplify the use of the two constructions.

- (1) His mother Princess Diana lost her life chased by paparazzi. (COCA; 2019)
- (2) I'm sorry that all these folks may lose their jobs, very sorry. (COCA; 2012)

Another construction in focus of this work is the singular *they*, in particular singular *their*, resulting from the use of gender-neutral *they*. Singular *they* is a much-discussed (e.g. Gernsbacher 1997; Baron 2018 & 2020; Solomon 2019; Bradley 2020, McCarthy 2020) personal pronoun, which has a long history of usage in the English-speaking world. According to Baron (2018), the singular usage of *they* can be traced back to the fourteenth century and it “has become the pronoun of choice to replace *he* and *she* in cases where the gender of the antecedent – the word the pronoun refers to – is unknown, irrelevant, or nonbinary, or where gender needs to be concealed”. Along similar lines, Solomon (2019: 69) writes:

Singular *they* has been attested in literature and in common speech for centuries, and continues to be used in spite of nineteenth-century pushback by prescriptive grammarians, because singular *they* fills the lexical gap in English where an epicene (grammatically unisex) third-person singular pronoun should be.

McCarthy (McCarthy 2020: 78) also acknowledges the fact that singular *they* has already become part of the pronoun system and claims that its advent made the system of *she* (female) and *he* (male or generic⁴) “defunct”.

Baron goes one step further and shows that the pronoun-question seems to be one of the most important questions nowadays, its importance reaching far beyond correctness (2020: 1): “ "What's your pronoun?" is an invitation to declare, to honor, or to reject, not just a pronoun, but a gender identity. |And it's a question about a part of speech.”. Finally, it is worth mentioning that singular *they* has recently become Merriam-Webster’s Word of the Year 2019⁵.

³ Merriam-Webster, s.v. *lose one's job*, retrieved on May 20, 2020 from <https://www.merriam-webster.com/dictionary/lose%20one%27s%20job>.

⁴ Before 1970s *he* used to be used as a generic pronoun to refer to both men and women, however a feminist call in the 70s led to its decline (Baron 2020).

⁵ Merriam-Webster’s Word of the Year 2019: *they*. Accessed at July, 27, 2020. Available at <https://www.merriam-webster.com/words-at-play/word-of-the-year/they>.

The present work aims at contributing to the research on singular *they*, and the possessive pronoun *their* used with regard to singular subjects, exemplified in (3), by looking at whether and how it is used with the two constructions of choice.

(3) I suppose it's just that everyone has their own little routines for doing things. (BNC; 1991)

The main goal of the paper is to present an account of the differences between the two varieties of English studied, the British and American English, with regard to the usage and structure of the two studied phrases and, in this context, with a special focus on the usage of singular *they*. The methodology presented in the next part, provides a valuable addition to the work on its own, as it could be seen as an example of working with mega-corpora and big amounts of linguistic data and the potential problems one may come across.

This paper is structured as follows: the next (second) section contains a detailed description of methodology and the pitfalls of working with large amounts of data; the third section presents the results and describes the tendencies that we see by comparing the different datasets; the fourth part discusses the potential explanations between the datasets.

2. Methodology

The corpora used for this study are i) the British National Corpus (BNC)⁶; and; ii) the Corpus of Contemporary American English (COCA)⁷. Whereas COCA, with its (more than) one billion words of text is often given as an example of a *mega-corpus*, some doubts may arise as to whether we can refer to the BNC as to a mega-corpus. Weisser (2016: 30) suggests a threshold of 100 million words of text and according to his classification (2016: 20, Table 2.6) both COCA and the BNC can be termed as modern mega corpora.

The present section contains particular steps, which are taken in order to obtain the picture of how frequent the studied phrases are, which variants happen to be the most popular ones and if there are any differences with regard to the use of singular and plural *they* between the two regional varieties.

⁶ The BNC contains approximately 100 million words of text texts from different genres (e.g. fiction, newspapers, magazines, spoken and academic) and is available online at <https://www.english-corpora.org/bnc/>.

⁷ COCA contains more than one billion words of text from different genres (e.g. fiction, popular magazines, spoken, newspapers, academic texts) and is available online at <https://www.english-corpora.org/coca/>.

- 1) For both the BNC and COCA, the search engine at <https://www.english-corpora.org> is used.
- 2) For *lose one's life*, two queries are conducted, namely i) [lose] _app* life and ii) [lose] _app* lives. The [lose] part encompasses all inflected forms of the verb to lose and _app* refers to all possessive pronouns.
- 3) For *lose one's job*, one query is conducted, namely 1) [lose] _app* [job]. The query is in principle the same as with *lose one's life*, but since the plural of *job* is regular, one could use the square brackets to see both the singular and plural contexts (with *life* and *lives* the use of square brackets does not seem to work as smoothly, so, in order to stay on the safe side, two queries are conducted to account for all of the instances).
- 4) Two datasets are created – one for the BNC and one for COCA – both of them contain information on all the detected instances of the two types. For the BNC the dataset contains 632 observations, for COCA the number is sixteen times higher – 10144.
- 5) In the course of data collection, manual qualitative assessment of the data is conducted for results containing the possessive pronoun *their*, such as *lost their life* or *losing their jobs*. The goal is to differentiate between the cases in which *their* was used to refer to third person singular – with pronouns such as *anybody*, *somebody*, *everybody*, and the cases in which there was a singular object for a plural subject.
- 6) A careful visual exploration of the data is conducted, as a result of which certain phrases, which, at first sight, may seem to instantiate the constructions in focus, but turn out to be “false positives”, can be detected. One such example is provided by cases containing the noun phrase *life savings* and not just *life* (37 of them found in COCA; 5 found in the BNC), such as *lost their life savings* or *lost her life savings*. These cases are removed from the datasets.
- 7) The programme of choice for the data exploration and analysis is R⁸, with its integrated development environment RStudio⁹; the graphics are created with the use of the ggplot2¹⁰ package. Thanks to the careful visual exploration and qualitative assessment of the datasets (described

⁸ R Core Team (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <http://www.R-project.org/>.

⁹ RStudio Team. (2020). RStudio: Integrated Development for R. RStudio, PBC, Boston, MA URL <http://www.rstudio.com/>.

¹⁰ Wickham, H. (2016). ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York. ISBN 978-3-319-24277-4, URL <https://ggplot2.tidyverse.org>.

in Steps 6 and 7 above), it is hoped that the detection precision is very high.

3. Results

In the BNC dataset there are 156 instances representing the type *lose one's life*. They can further be divided among sixteen variants of the *to lose one's life* construction. Figure 1 presents them all with regard to the number of instances, which were detected.

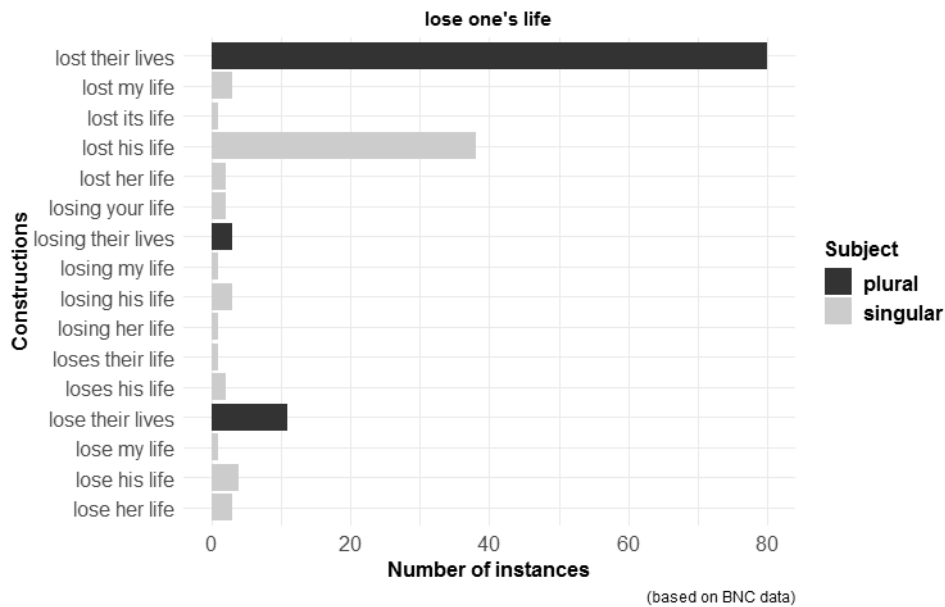


Fig. 1: Sixteen variants of the *lose one's life* construction extracted from the BNC

As we can see, the most frequent variant is *lost their lives* – there are eighty instances found in the corpus. It is followed by *lost his life*, of which there are 38 instances. The third place is taken by *lose their lives*, with 11 instances. The colour of the bars in Fig. 1 refers to the number of the subject in a given variant.

The second construction in our focus, namely *lose one's job*, is roughly three times more frequent than *lose one's life*. There are 476 instances of the construction and twenty-eight different variants. Figure 2 illustrates the frequency of particular variants.

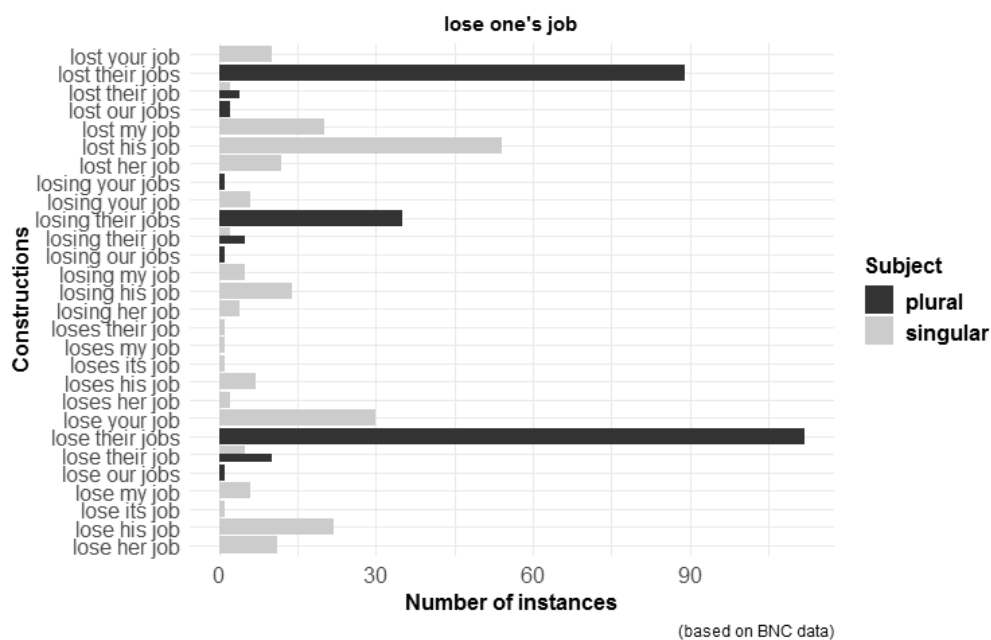


Fig. 2: Twenty-eight variants of the *lose one's job* construction extracted from the BNC

The most frequent one is *lose their jobs*, with 112 instances, the second most frequent variant is *lost their jobs* with 89 instances. The third most frequent is *lost his job* – there are 54 instances of it.

As has already been stated in the Methodology, all the variants containing *their*, have been assessed manually so as to qualify them as either instances of singular or plural subjects.

For *lose one's life*, only 1 instantiation of singular *their* has been detected, which is exemplified by (4). For *lose one's job*, the ten cases of singular *they* discovered belong to the following variants: *lose their job* (5 instances), *losing their job* (2 instances), *loses their job* (1 instance) and *lost their job* (2 instances) they can be seen below – see (5) and (6). All cases containing a singular *they* found in the BNC feature a singular object.

(4) And it's the attitude I think that very often causes everything from a major accident like that where someone loses their life (...). (BNC; 1993)

(5) If you can imagine someone losing their job, the depression that actually causes, perhaps both within them (...). (BNC; 1985-1994)

(6) No one dared refuse for fear of losing their job. (BNC; 1995)

In COCA, there are 2326 instances of *lose one's life* construction. These cases are divided between thirty-one variants, the most popular of them being *lost their lives* (1109), *lost his life* (297), *lose their lives* (164), *losing their lives* (101) and *lost her life* (96). Figure 3 presents a graphical representation of these results.

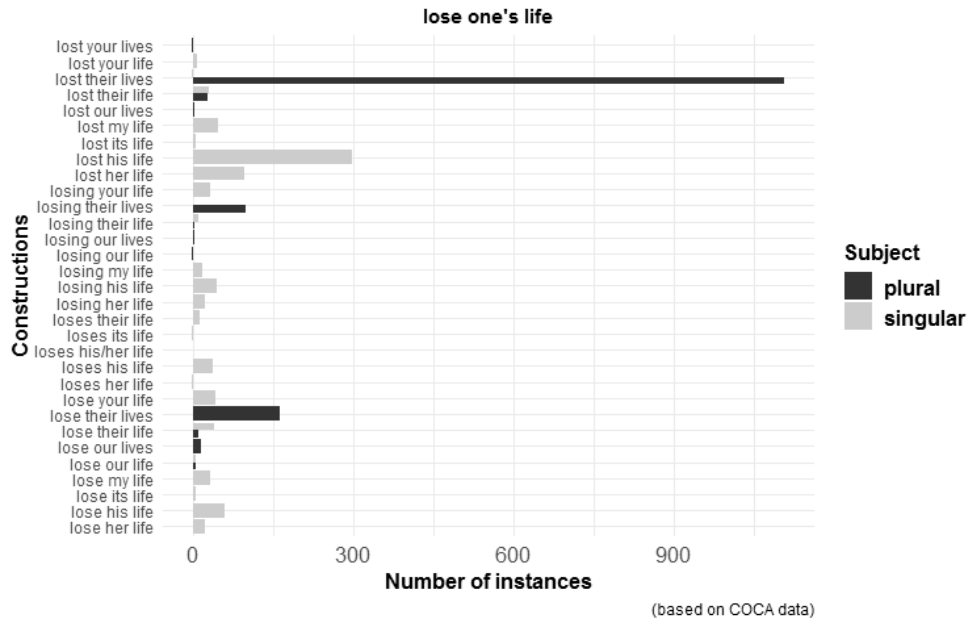


Fig. 3: Thirty-one variants of the *lose one's life* construction extracted from COCA

Interestingly, in spite of the difference in the sheer amount of data available in COCA (COCA containing almost fifteen times more instances of *lose one's life* construction than the BNC) as compared to the BNC, the three most frequent constructions are the same for both the former and the latter corpus. The situation is almost identical for the *lose one's job* construction – the three most frequent constructions are the same for both of the corpora, however, the frequency order differs slightly. The most popular variant for this type in COCA is *lost his job* (1141 instances), the second most frequent is *lost their jobs* (1085), and the third most frequent, *lose their jobs* (1064). Figure 4 illustrates the results.

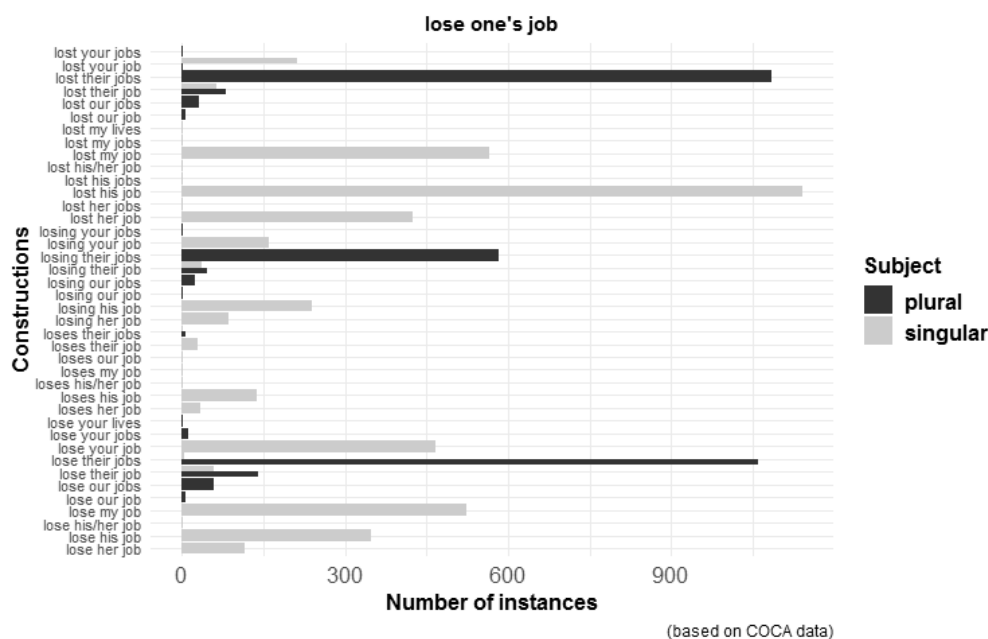


Fig. 4: Thirty-nine variants of the *lose one's job* construction extracted from COCA

As was the case for the BNC data, a manual exploration of the instances of variants containing *their* is conducted. During this analysis, sentences containing the particular possessive pronoun are, one by one, read and qualified as being cases of plural or singular subjects.

For the *lose one's life* type, there are ninety-five instances of singular use of *they* in the dataset of cases with a singular object; they belong to the variants *lost their life* (58 instances), *lose their life* (50 instances), *losing their life* (13 instances), *loses their life* (14 instances). Two cases of a singular subject and a plural object have been detected (variant *lost their lives*). In the case of *lose one's job*, the whole dataset contains 189 instances of singular *they*, belonging to the variants *lost their job* (64 instances), *loses their job* (29 instances), *losing their job* (37 instances), *lose their job* (59 instances). There are six instances of singular *they* with a plural object – variants *lose their jobs* (4 instances) and *loses their jobs* (2 instances).

Table 1 sums up the results obtained for each of the variants in the two corpora; Figures 5 and 6 provide illustrations. The cases with singular subjects and plural objects turned out to be very scarce, which is, however, not surprising, given the fact that in English there is a strong tendency towards distributive correlation (Quirk et al. 1985: 768) – plural subjects typically having plural objects.

Table 1: Sum-up of the results obtained from the BNC and COCA

Variants	BNC		COCA	
	singular <i>they</i>	plural <i>they</i>	singular <i>they</i>	plural <i>they</i>
<i>lost their life</i>	0	0	31	27
<i>lose their life</i>	0	0	40	10
<i>losing their life</i>	0	0	10	3
<i>loses their life</i>	1	0	14	0
Total for <i>life</i>	1	0	95	40
<i>lose their job</i>	5	10	59	141
<i>losing their job</i>	2	5	37	46
<i>loses their job</i>	1	0	29	0
<i>lost their job</i>	2	4	64	81
Total for <i>job</i>	10	19	189	268
Total for <i>lose one's life & lose one's job</i>	11	19	284	308

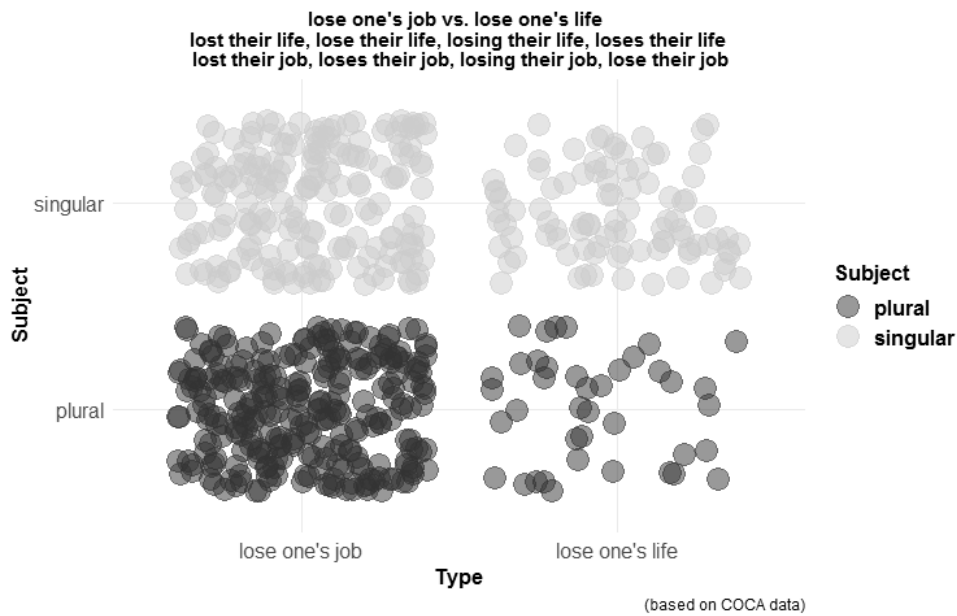


Fig. 5: Cases of singular *they* for the both types – *lose one's life* (right) and *lose one's job* (left); the particular variants are listed in the plot title. Based on COCA data

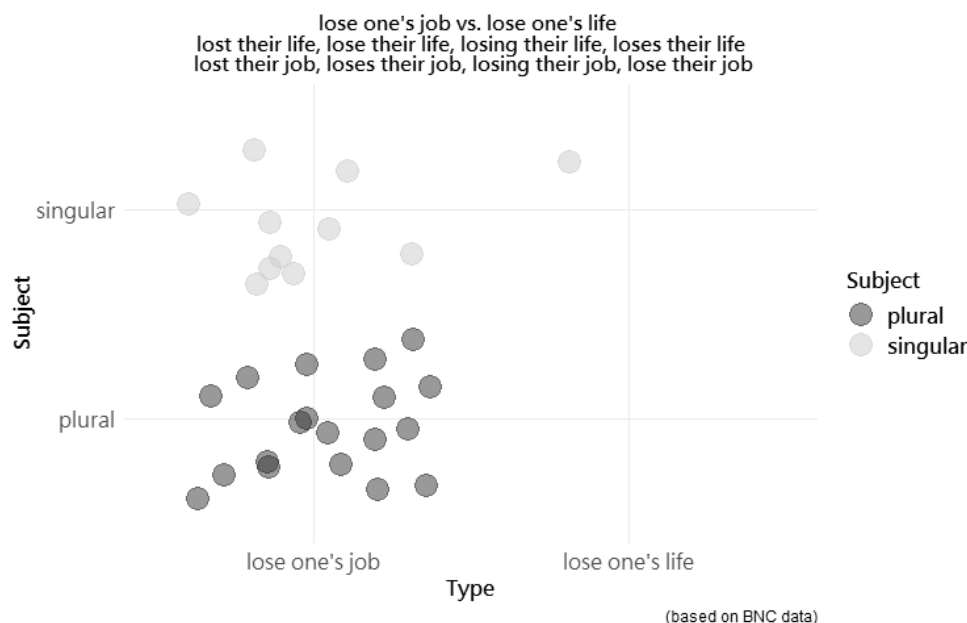


Fig. 6: Cases of singular *they* for the both types – *lose one's life* (right) and *lose one's job* (left); the particular variants are listed in the plot title. Based on BNC data

4. Discussion

In the COCA dataset, there are, obviously, more instances of singular *they* than in the BNC. To provide the clearest explanation, one has to say that COCA itself contains more than one billion words of text, whereas the size of the BNC amounts to 100 million words of text. For the *lose one's life* type, fifteen times more instances were found in COCA, than in the BNC (all variants counted up); for *lose one's job*, the difference was ~16.4 times in favour of COCA.

It is a bit difficult to directly compare both datasets, since COCA is around ten times (to be precise 10.41 times) larger than the BNC, with regard to the number of words (1 001 610 938¹¹ words of text in COCA vs. 96 263 399¹² words in the BNC). Therefore, to make our datasets comparable in size, we multiply the numbers relevant for the BNC by 10.41. The question

¹¹ Information retrieved from the webpage at <https://www.english-corpora.org/coca/> on July, 30, 2020.

¹² Information retrieved from the webpage at <https://www.english-corpora.org/bnc/> on July, 30, 2020.

this procedure of extrapolation could answer is “What would the rough numbers be, if the BNC dataset was as big as the COCA dataset”. In order to arrive at an answer, we multiply $11 * 10.41$ (to get the result for singular *they*) and $19 * 10.41$ (for plural *they*). The results we arrive at, are 114.5 for the singular *they* and 197.79 for plural *they* for the BNC dataset. Table 2 sums up the results. In terms of percentages, this procedure does not change anything, as the proportion of the cases stays the same.

Table 2: Singular and plural *they* in the BNC and COCA – results after extrapolation

Variants	BNC		COCA	
	singular <i>they</i>	plural <i>they</i>	singular <i>they</i>	plural <i>they</i>
<i>Lose one’s life</i> and <i>lose one’s job</i> – total results	114.51 (after extrapolation)	197.79 (after extrapolation)	284	308
<i>Lose one’s life</i> and <i>lose one’s job</i> – total results in percentages	36.66%	63.34%	47.97%	52.03%

Nevertheless, for both the singular *they* and plural *they*, the BNC results after the renormalization are lower than the values obtained for the COCA dataset. Interestingly, the figure for singular *they* is much lower – 114.5 as compared to 284. This observation might suggest that singular *they*, at least used with the two constructions in question, seems to be more widely used in the American variety of English. Further support for this claim might be offered by a look at Table 1 and Fig. 5. Strikingly, in the BNC data, there were almost no cases of singular *they* with the *lose one’s life* construction.

This conclusion goes in line with what Quirk et al. (1985: 770) claim on the topic of singular *they*, namely “At one time restricted to informal usage, it is now increasingly accepted even in formal usage, especially in AmE [American English]”.

Grammatical changes seem to happen at a rate much slower than it is usual for lexical changes (Mair 2006; McCarthy 2020). Out of the two varieties, the present work looks at, American English is the one which, according to Mair (2006: 188), is “often assumed to lead in the change towards more informal modes of expression in writing”. Given this tendency, the fact that the

singular *they* seems to be somewhat more frequent in written American English, should not surprise us.

As a counterargument, however, one might raise the fact that COCA contains “younger” material – texts stem from almost three decades between 1990 and 2019, whereas the texts included in the BNC are from the 1980s and early 1990s. In the best-case scenario one could imagine having two corpora containing texts from, for instance, the same decade.

REFERENCES

- Baron, D. (2018). A brief history of singular 'they'. Oxford English Dictionary: The definitive record of the English language. Accessed on July, 20, 2020, available online at <https://public.oed.com/blog/a-brief-history-of-singular-they/>
- Baron, D. (2020). *What's your pronoun? Beyond He and She*. W. W. Norton.
- Bradley, E. D. (2020). The influence of linguistic and social attitudes on grammaticality judgements of singular 'they'. *Language Sciences* 78.
- Gernsbacher, M. A. (1997). Generic pronominal anaphora: The case of the English singular they. *Verbum* 19 (1-2), 67-8.
- Hilpert, M. (2014). *Construction Grammar and its application to English*. Edinburgh: Edinburgh University Press.
- Hoffmann, T. & Trousdale, G. (2013). Construction Grammar: Introduction. In: *The Oxford Handbook of Construction Grammar*, T. Hoffmann & G. Trousdale (eds.).
- McCarthy, M. (2020). *Innovations and Challenges in Grammar*. Routledge: Taylor & Francis Group.
- Mair, Ch. (2006). *Twentieth-century English: History, variation and standardization*. Cambridge: Cambridge University Press.
- Quirk, R., Greenbaum S., Leech G. & Svartvik J. (1985). *A Comprehensive Grammar of the English Language*. Edinburgh: Longman.
- Solomon, A. N. (2019). The case for singular they and implications for English Language Teaching. *TESOL Working Paper Series* 17, 69-79.
- Weisser, M. (2016). *Practical Corpus Linguistics: an Introduction to Corpus-Based Language Analysis*. Malden, MA & Oxford: Wiley-Blackwell.

Language corpora & Dictionaries

- Davies, M. (2004-) *British National Corpus* (from Oxford University Press). Available online at <https://www.english-corpora.org/bnc/>
- Davies, M. (2008-) *The Corpus of Contemporary American English (COCA): One billion words, 1990-2019*. Available online at <https://www.english-corpora.org/coca/>
- Cambridge Online Dictionary, s.v. *lose your life*, retrieved on May 17, 2020. Available at <https://dictionary.cambridge.org/dictionary/english/lose-your-life>
- Merriam-Webster, s.v. *lose one's job*, retrieved on May 20, 2020. Available at <https://www.merriam-webster.com/dictionary/lose%20one%27s%20job>

Software

- R Core Team (2013). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL <http://www.R-project.org/>
- RStudio Team. (2020). *RStudio: Integrated Development for R*. RStudio, PBC, Boston, MA URL <http://www.rstudio.com/>
- Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. ISBN 978-3-319-24277-4, URL <https://ggplot2.tidyverse.org>