



Andreea S. Calude*, Maebh Long and Jessie Burnette

#AreHashtagsWords? Structure, position, and syntactic integration of hashtags in (English) tweets

<https://doi.org/10.1515/lingvan-2023-0044>

Received March 17, 2023; accepted February 22, 2024; published online ■■■

Abstract: As social media use continues to increase its presence in our lives, so does the language of such platforms. One of the most salient features of social media discourse is the hashtag. Starting its life on Twitter (X) about 15 years ago, the hashtag has seeped from online to offline communication. Yet, it is not clear whether hashtags are words, tags, or something else altogether, nor is it clear what morphological process gives rise to them. This study presents an extensive analysis of 3,423 hashtags from 1,216 English-language tweets, each manually coded for various linguistics features, including position in the tweet, grammatical function, and syntactic integration. Our findings suggest that hashtags are extremely varied and we propose that they are indeed words, arising through a process of hashtagging (which is distinct from compounding). We also argue that some hashtags are syntactically integrated while others constitute parenthetical material.

Keywords: hashtags; word-formation; syntax; social media language; Twitter (X)

1 What are hashtags and how do they arise?

Few language elements allow researchers the luxury of being able to scrutinize in full their birth and evolution through time. The hashtag is one of these.

The first hashtag was used in 2007 by Chris Messina (Figure 1), proposed specifically for tagging Twitter content, mimicking the “channel” function of Internet Relay Chat (Messina 2007; Scott 2015). Although originally devised for Twitter (now X), hashtags have spread to other social media platforms as well as other language mediums, including spoken conversation (Scott 2018).

It is remarkable just how unremarkable the *#barcamp* hashtag is; many people do not know this was the first hashtag used. Despite the increased attention that hashtags have been receiving (Caleffi 2015; Page 2012; Zappavigna 2011, 2012, 2015), questions remain regarding their linguistic characteristics, from what they are (words, tags, or something else), to how and where they are used in discourse. Answering these questions is crucial to understanding social media language.

The status of hashtags vis-à-vis word-hood is more than a theoretical quibble. If hashtags are indeed words, questions arise about how they fit into the discourse they are part of. Previous work suggests that at least *some* hashtags play a role in the syntax and semantics of the tweets in which they appear (Section 3). So, for at least these hashtags, syntactic theory must take into account their existence and specify whether they are integrated as syntactic constituents. To our knowledge, no previous study has addressed the issue of syntactic integration of hashtags.

*Corresponding author: Andreea S. Calude, University of Waikato, Hamilton, New Zealand, E-mail: andreea.calude@waikato.ac.nz
Maebh Long and Jessie Burnette, University of Waikato, Hamilton, New Zealand





Figure 1: First ever use of a hashtag.



2 Twitter language and the Covid19 NZ Vaccine corpus

With more than 300 million registered users (Statista 2020), the microblogging site Twitter (now X, <https://twitter.com>) is a well-known social media platform in the English-speaking world. The language of Twitter posts straddles the spoken–written continuum (Trye et al. 2020). On the one hand, tweets are often informal, are produced spontaneously, and involve slang, non-standard forms, and simplified grammar, especially when posted by individuals (as opposed to organizations), suggesting they are best placed at the spoken end of the continuum. On the other hand, tweets are invariably written and read, and thus associated with written language, and depending on the type of Twitter account, they are also editable, which brings them towards the written end of the continuum.

Twitter language is both challenging and appealing. It is challenging because it can be messy and unpredictable. It is appealing because it is dynamic and emergent and, like speech, it is often associated with rampant language innovation.

In previous research (Burnette and Calude 2022; Burnette and Long 2022), we have scrutinized how users engaged with social media platforms to discuss their experiences in relation to COVID-19. Building on this work, we created a new corpus on the topic of vaccines (the Covid19 NZ Vaccine corpus). This corpus consists of 4,701 English-language tweets containing both the hashtag *#Covid19NZ* and its variations (e.g. *#covid19nz*) and the lemmas *VACCINE* (*vaccinate*, *vaccines*, *vaccination*, and so on, as well as hashtags containing these words, such as *#vaccinemandate*, *#vaccine_passes*, *#vaccinecertificate*) or *VAX* (*vaxxed*, *vaxxing*, and related hashtags, e.g. *#vaxathon*, *#vaxthenation*). The tweets, which were posted between August and November 2021, were extracted using Python code, written by one of our collaborators, and the Twitter API (<https://developer.twitter.com/en/docs/twitter-api>). Although we did not collect this data specifically for the purpose of analysing hashtags, an initial keyword analysis suggested that hashtags were particularly salient in our corpus (see Long et al. In press). We therefore used this opportunity to extend current research on hashtags.

Owing to our interest in analysing the syntax of hashtags, we reduced the number of tweets to a manageable set which could be inspected manually, while still retaining sufficient data to identify trends. Thus, a subset of 1,216 tweets was manually analysed. Tweets were selected from the larger corpus from each month with reference to key events, for example changes in vaccination policy, as these events generated larger numbers of *#Covid19nz* tweets. After choosing a range of dates, tweets were randomly and blindly selected from this range.

While Twitter's privacy policy (<https://twitter.com/en/privacy>) allows for the use of publicly available content for research purposes, we have kept examples cited here anonymous in order to protect users' identities (Fiesler and Proferes 2018).

3 Previous linguistics analyses of hashtags

Since their inception, hashtags have received ample scholarly attention, particularly from computational linguists performing sentiment analyses of tweets (e.g. Venkit et al. 2021; Wagh and Punde 2018; inter alia). However, analyses of hashtags from a theoretical linguistics perspective whose goal is not data mining but

explaining language phenomena remain comparatively less prevalent. Within this body of research, two main themes have emerged: (i) studies which analyse the function of hashtags in discourse (typically from tweets), and (ii) studies which probe the morphological process that gives rise to them. As both strands of research relate to this study, we summarize some main claims in what follows.

The key functions of hashtags are searchability, discoverability (Caleffi 2015), and “affiliation” of posts to various communities or topics (Zappavigna 2011, 2012, 2015), where “affiliation” refers to “the process by which people involve themselves in social bonds” (Zappavigna 2012: 795). This type of function relates to technological affordability arising from the links facilitated by social media platforms. A different type of function, and one more closely related to what traditional linguists link to discourse function, is the use of hashtags for self-promotion and “self-branding” (Page 2012). Other discourse uses include interpersonal experiential functions, expressing stance, emotions, and evaluations. See (1) for some examples: (1a), (1c), and (1d) are from Zappavigna (2015: 275), while (1b) is from Page (2012: 191).

- | | |
|-----|---|
| (1) | <ul style="list-style-type: none"> a. Affiliation <i>RT @User: Today's knitting technique: Butterfly edging. #knitting [link]</i> b. Self-branding <i>I don't think #eurovision was as good as last year. The show, songs and performances were far bigger and better in Moscow.</i> c. Emotion <i>Tattoo no. 5 booked #excited</i> d. Stance, evaluation <i>When you find out things you really wish you didn't #upset #seriously</i> |
|-----|---|

Looking beyond social media, Scott (2018) illustrates the use of hashtags in spoken conversation and shows that their use is often playful and humorous, mocking online behaviour or referencing online memes; see (2) from Scott (2018: 62):

- (2) *#I can't date you if ... you can't drive*

It is interesting to consider Scott's spelling in (2). There seems to be a mismatch between how the hashtag might have been spelled on social media (i.e. without spaces) and how it is transcribed from spoken conversation (with spaces). The spaces suggest the hashtag contributes to the meaning of the utterance and is accommodated within its structure.

Turning to the morphology of hashtags, there are two main positions regarding the process that gives rise to them. Caleffi (2015: 52) proposes a novel process called “hashtagging”, which she claims to be a new mechanism for “producing items, called hashtags, whose linguistic nature may be difficult to identify and relate to any traditional part of speech”. Analysing English and Italian hashtags, Caleffi (2015: 67) concludes that hashtagging produces “‘linguistic items’, which are both words and yet not words”. She does not, however, define what she means by a “word”.

A contrasting view proposes that hashtags are compounds, coined through traditional lexical compounding (Maity et al. 2016). The authors state that hashtag compounds involve similar parts-of-speech combinations to traditional lexical compounds, such as noun-noun (*bus stop*) or adjective-verb (*dry-clean*) (2016: 3). They do not define the term compound formally but note “differences in the mechanisms of compound formation in ordinary language and in social media” (2016: 3). However, they do not dwell on the nature of these differences.

It is not straightforwardly obvious how a compounding analysis can accommodate hashtags which involve full clauses, such as *#WeAreTheBest* or *#MentionPeopleYouReallyLove*,¹ with some arguing against a compound analysis in such cases (Trye et al. 2020: 14). Both Page (2012) and Caleffi (2015) provide clausal hashtag examples but do not discuss problems in their categorization. This article aims to contribute to debates surrounding the status of hashtags as words, compounds, or something else.

¹ One reviewer points out that compounds spanning entire clauses are possible in spoken language (e.g. an *I-told-you-so* tone); however, these examples are specialized uses in this language medium. While it may be the case that hashtags follow spoken language compounding, this needs to be established through analysis first.

4 Linguistics of hashtags in Twitter

We extracted 3,423 hashtags from our corpus, which contained 518 unique hashtags, three of which were non-English hashtags (28 tweets were eliminated from the original data because they consisted solely of hashtags and links, with no other text). Most tweets exhibited multiple hashtags, on average three, with a range of 1–32. Among these, 472 tweets contained one hashtag, *#Covid19NZ* (or variations of it), while the remaining 744 tweets contained a mix of hashtags. Figure 2 provides a frequency distribution of the hashtags, showing that, typically, tweets contain between one and four hashtags and seldom more than 10 (this and all subsequent plots are created with RStudio [RStudio Team 2020] and the *ggplot2* package [Wickham 2016]).

Table 1 gives all hashtags that occurred at least 10 times and the number of times each appeared. We can measure the length of hashtags by counting the number of components involved, including words, numbers, and emojis (we use spaces in written language spelling conventions as indicators of length, so that “New Zealand” has length 2, while “NZ” has length 1). Inspecting the most frequent hashtags in Table 2 suggests that hashtag length is not a strong predictor of frequency. The mean, mode, and median for hashtag length are all two words. Table 2 provides the frequency of the various types of hashtags in the sub-corpus. As noted earlier, in addition to words and numbers (including those used as homophones to denote words, e.g. *#vax2themax*), we also find hashtags containing emojis, such as *#getcovidvaccinated* 🍷 and *#vaccinate* 💉.

Tables 1 and 2 show that the internal structure of the hashtags includes single words (*#curfew*), complex words involving affixes (*#racist*, *#losers*), compounds (*#PfizerVaccine*, *#NewZealand*), multi-word phrases (*#plentyofreasons*), and whole clauses (*#LeaveOurKidsAlone*).

While hashtags can involve affixes within them, some hashtags exhibit affixation which is positioned outside the hashtag itself. In (3), the morpheme *anti* is used as a prefix of the hashtag *#Covid19NZ*, although, unlike regular roots, it is placed outside the hashtag and separated from the root by a hyphen and the hashtag symbol. However, while the spelling conventions might differ, the meaning and function of the hashtag matches regular roots and derivation processes in English.

- (3) *People getting anti-#Covid19NZ vaccinations at Wellington’s Sky Stadium. Bookings for Sky Stadium can be made online at <http://BookMyVaccine.nz> (enter Pipitea in the location box and select ‘Sky Stadium Drive-Thru’) or freephone 0800 28 29 26. Maybe space on Sunday too?*

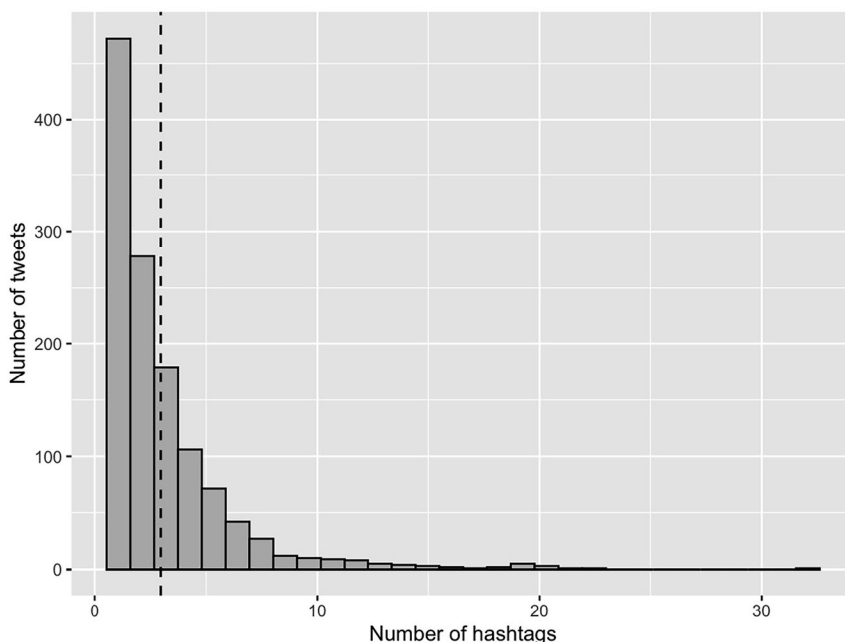


Figure 2: Frequency distribution of hashtags in the Covid19 NZ Vaccine sub-corpus (3,423 hashtags within 1,216 tweets).

Table 1: Hashtags which occurred 10 or more times in the Covid19 NZ Vaccine sub-corpus.

Rank	Hashtag	No. of occurrences
1	#COVID19nz	741
2	#nzpol	200
3	#COVID19	137
4	#SuperSunday	76
5	#lockdownnz	60
6	#vaxathon	57
7	#GetVaccinated	55
8	#COVID	53
9	#nz	50
10	#CovidNZ	43
11	#auspol	37
12	#COVID19Aus	36
13	#NoVaccinePassports	35
14	#vaccine	35
15	#vaccination	30
16	#COVID19nsw	29
17	#coronavirus	28
18	#NewZealand	27
19	#NoVaccineMandates	27
20	#Maori	25
21	#Covid_19	24
22	#COVID19Vic	24
23	#CovidVaccine	24
24	#hewakaekenoa ^a	24
25	#lockdown	24
26	#Corona	23
27	#ResignJacinda	23
28	#vaccinated	23
29	#waateanews	23
30	#covid19qld	16
31	#rollupyoursleevesnz	16
32	#vaccines	16
33	#Pfizer	15
34	#vaccineSideEffects	15
35	#vaccinate	14
36	#Auckland	12
37	#CovidIsNotOver	12
38	#CovidVic	12
39	#DeltaVariant	12
40	#GetVaccinatedNow	11
41	#VaccinePassports	11
42	#Booster	10
43	#COVIDIOTS	10
44	#covidnsw	10
45	#TeamOf5Million	10
46	#VaccinesWork	10

^aThis hashtag references a slogan in the Māori language (the Indigenous language of Aotearoa New Zealand) linked with COVID-19, *he waka eke noa* ‘we are all in this together’.

We now turn to the grammatical function of hashtags. Although in principle hashtags can function as any word category, they are largely (70 %) used as (common or proper) nouns or as tags, as seen in (4); tags are labels which categorize a tweet, affiliating it to a particular topic (COVID-19) or a particular group (those interested in NZ politics).

Table 2: Overview of the internal structure of hashtags.

Length (in words/emojis)	No. of hashtags	Percentage of total	Examples
1	172	33.2 %	#antivaxxers, #BREAKING, #curfew, #elimination, #hesitancy, #humanity, #hypocrite, #medicine, #lockdown, #losers, #racist, #smallpox, #vaxathon, #vaccination, #vaccines
2	216	41.7 %	#BigPharma, #boostershots, #COVID19nz, #DeltaVariant, #GladysFail, #jacindaardern, #MaskUp, #nzlockdown, #nzpol, #NewZealand, #PfizerVaccine, #ResignJacinda, #ShotBro, #VaccinePassports, #90percent
3	93	18.0 %	#covidtracerapp, #doingmypart, #doyourduy, #GotYaDot, #GymLife ✓, #level4lockdown, #NoMedicalApartheid, #plentyofreasons, #RightToDie, #TheAllVaxxed
4	23	4.4 %	#CovidIsNotOver, #NoVaccinePassportsAnywhere, #TeamOf5Million, #birdoftheyear, #covidisalie, #getcovidvaccinated 🧪, #GladysTheSuperSpreader, #LeaveOurKidsAlone, #NZCOVIDTracerApp, #ScottyFromDamageControl, #vax2themax
5	4	0.8 %	#rollupyoursleevesnz, #birdoftheyear2021, #doitforthekids, #VaccinateToSaveSouthAfrica
6	–	–	–
7	1	0.2 %	#rollupyoursleevehatsecondtime
Non-English	3	0.6 %	#海外生活
Unclear	6	1.2 %	#coronatoegangsbewijs
Total	518	100 %	

- (4) *#nzpol² Isn't it telling how our street gangs can put aside their politics to work together to reduce the spread of #COVID19NZ and yet right wing politicians like Judith Collins and David Seymour can't and won't. Gang leaders urge whānau³ to get vaccinated [link]*

However, our data also contains examples where hashtags are used as directives (13 %), such as in (5a); as adjectives (7 %), as in (5b); and as clausal units or slogans (5 %), seen in (5c). There are also two cases where they denote numerals, such as in (5d).

- (5) a. *Terrible. Pull your finger out #NZ. #GetABloodyVaccine #Covid19NZ*
 b. *So, it turns out I have to order myself a new birth certificate before I can sign up for the stupid vaccine certificate! Don't see why the kiwi access card can't be used as ID too 🙄🙄 #covid19nz #COVID19nz #frustrated*
 c. *So proud of each and every New Zealander who took up the challenge and got vaccinated today. #100000 #COVID19nz #freedomiscoming #smashitforsummer*
 d. *Time for jab #2. Goooo being vaccinated! #covid19nz*

Next, we consider the position of hashtags within tweets, and here too we find variation. Some tweets contain hashtags at the beginning, most tweets have hashtags at the end, and some exhibit hashtags within the body of the tweet. There are also tweets with hashtags in a mixture of positions (e.g. example [5d]), sometimes in all three (beginning, middle, and end). Figure 3 provides a quantitative overview summarizing the distribution observed. Most hashtags tend to be either tweet-final or tweet-medial, or both within the tweet and at the end of it. The least preferred position is tweet-initial. This suggests that users are not immediately focused on hashtags; they begin with the message of the tweet and only add hashtags along the way or at the very end (which could also be the result of following what others are doing).

Inspecting the syntax of tweets containing hashtags, we find that some hashtags are not syntactically integrated within the text containing them, behaving in a similar manner to parentheticals (Dehé and Kavalova 2007), while others are. Like parentheticals (Dehé and Kavalova 2007: 4), syntactically non-integrated hashtags fail tests

² #nzpol references NZ politics.

³ Whānau is the Māori word for extended family.

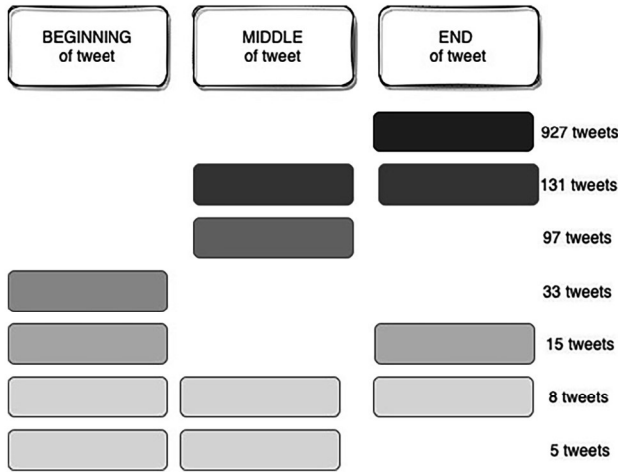


Figure 3: Position of hashtags in tweets.

of constituency, such as clefting, movement, and question-answer. In contrast, integrated hashtags are amenable to constituency tests. An analysis of the first two hashtags in example (5a) illustrates this difference, as seen in Table 3.

Punctuation can sometimes be an indicator of integratedness, for example, the hashtags #GetABloodyVaccine and #Covid19NZ (in example [5a]) occur after a full stop, which signals a separation between the sentence and the hashtags. Punctuation is also aligned with constituency in (6), where the question mark following the second hashtag helps to position it as a constituent in the sentence containing it. The syntactic integration of both hashtags in this example is also evident from the fact that without them, the sentence would be incomplete and ungrammatical.

- (6) *Whānau Ora probably have good intentions but will Māori people unvaccinated against #Covid19NZ co-operate with them? Will they trust the system in future? Or will this backfire and people be even less likely to #GetVaxxed?*

Tweet-final and tweet-initial positions can feature both integrated and non-integrated hashtags; tweet-medial hashtags tend to be syntactically integrated. However, there are also exceptions, as some tweet-medial hashtags interrupt the flow of the tweet and break up the sentence structure. Their contribution is affiliative and not semantic, and their position is at odds with their lack of syntactic integration. In (7), the first three hashtags are unintegrated, whereas the fourth hashtag is integrated.

- (7) *Early to rise, then early to bed, holding pattern Day 45 in Auckland’s Lockdown #lockdown #COVID19nz #pandemic #GetVaccinated to win back our freedom* 🇳🇿 🙏 🇺🇸

Within our hashtag data, we find that affiliative tag-like hashtags are usually not syntactically integrated. The distribution of syntactically integrated hashtags is given in Figure 4. The figure shows a relationship between integration and tweet position: tweet-final hashtags tend to be unintegrated, while tweet-medial hashtags are overwhelmingly syntactically integrated. Tweet-initial hashtags are similar to tweet-final hashtags, generally

Table 3: Analysis of the first two hashtags in (5a) in terms of constituency tests.

Constituency test	Hashtag #NZ	Hashtag #GetABloodyVaccine
Clefting	<i>It is #NZ that [should] pull its [your] finger out!</i>	<i>*It is #GetABloodyVaccine that pull your finger out #NZ!</i>
Movement	<i>#NZ pull your finger out!</i>	<i>#GetABloodyVaccine pull your finger out #NZ.</i>
Question-answer	<i>Who should pull its finger out? #NZ</i>	<i>??</i>

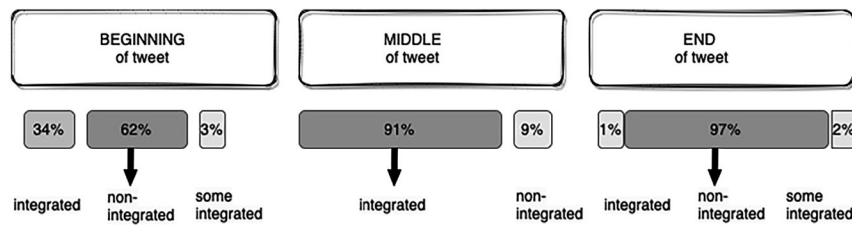


Figure 4: Syntactic integration and position within tweet.

being syntactically unintegrated, however are more likely to be integrated than tweet-final hashtags. Although rare, some tweets contain a group of hashtags, in which one (or more) is integrated and a following one is unintegrated; see (8).

- (8) *#COVID19nz* 📌🤔 *#DeltaOutbreak: #NZ to move to #traffilight system on Friday next week, #Auckland hairdressers open from this Thursday - PM #JacindaArdern #pakustv #NYC #globalhealth #lockdownnz #covidisnotover #GetVaccinated* [link]

Hashtags that are syntactically integrated typically function as any regular argument might in written language, the # symbol notwithstanding. For example, hashtags can function as heads of their grammatical phrases, typically noun phrases:

- (9) *Will foreigners who have received low efficacy vaccines be allowed into #NZ in April 2022? Not all #vaccines are as effective as the #Pfizer vaccine which #Kiwis have taken. I guess we will have to wait for more details. #nzpol #Covid19nz*

However, one difference between arguments and integrated hashtags (e.g. #NZ, #vaccines, #Pfizer, and #Kiwis in example [9]) is that hashtags have a dual role to play, acting as constituents of the phrases and clauses in which they occur but also fulfilling an affiliative function enabling discoverability. Hashtags can function as heads of phrases, as #vaccines does in (9), but also as modifiers within these, as #Pfizer does in (9) and #covid19nz in (10). In (10), the phrase *ur [your] #covid19nz risks* contains the head *risks* and two modifiers, one of which is a hashtag. Similarly, the head #GoldenTicket also has two modifiers (*an* and #ironclad), one of which is a hashtag.

- (10) *Name calling aside, here's an actual accurate ranking of ur #covid19nz risks Scott:*

- (1) *#nzpol Ardern Govt drops #elimination*
- (2) 📌📌 ed Aucklanders crossed the border
- (3) 📌📌 ed locals hanging with Aucklanders, all assuming they have an #ironclad #GoldenTicket 😞
- (4) *unvaxxed*

Similarly, in (11), the noun phrase *the #NewZealand Government* contains the head noun *Government* and the modifier #NewZealand. The phrase *the #NewZealand Government* is problematic because it is unclear whether it ought to be analysed as a compound or as a noun phrase, headed by the noun *government* (for a discussion of ambiguity between phrases and compounds, see Bauer 2017). Another ambiguous example is given in (12).

- (11) *@AmnestyNZ is urging the #NewZealand Government to rethink its approach to passing the COVID-19 Response (Vaccination) Legislation Bill due to serious concerns about the lack of opportunity for public consultation and scrutiny. #covid19nz*
- (12) *Double #vaxxed & a shirt with collar for #formalfriday #COVID19nz*

5 Discussion and conclusion

In this paper, we investigate linguistic properties of hashtags in the 2021 Covid19 NZ Vaccine corpus of Twitter data. In keeping with findings from other scholarly work (Caleffi 2015; Page 2012), we also found great variety in the internal structure of the hashtags analysed. In particular, we note the presence of simple as well as complex hashtags, some spanning entire clauses (*#CovidIsNotOver*) and directives (*#rollupyoursleevesnz*).

So, are hashtags words? The notion of ‘word’ is itself problematic (Bauer 2017; Haspelmath 2011), yet it is ubiquitous and useful, and abandoning it is impractical (Haspelmath 2023). Adopting Haspelmath’s (2023: 2) latest definition, we take a word to be “(i) a free morph, (ii) a clitic, or (iii) a root or compound plus possibly nonrequired affixes plus required affixes if there are any”, an important consideration being a move away from spelling conventions (spaces are not helpful; see also Bauer 2017: 6).

Hashtags exhibit similar characteristics to free morphs. Syntactically integrated hashtags function as constituents in the clauses they are part of, they can act as heads of phrases or as modifiers, and they can even take affixes (e.g. *anti-#Covid19NZ*). These are all in keeping with category (i) of the definition. Unintegrated hashtags share the ability to stand alone, like free morphs. We do not see the presence of clausal or directive hashtags as problematic to word-hood because as a unit, they encompass a particular meaning: for instance, *#rollupyoursleevesnz* represents a recurrent (and perhaps formulaic) call to action. As with other longer words in spoken language (see footnote 1), tweeters have moulded these longer, formulaic hashtags into units which function as free morphs. Furthermore, their use beyond social media in spoken conversation (Scott 2018) shows that speakers engage with these elements in a similar fashion to how they do with words (compare the lack of hyperlinks or “@” mentions in other language genres).

A second question is whether the process which gives rise to hashtags can be said to be compounding. Taking a compound to be “a combination of two roots that must occur next to each other and [which] cannot be expanded by nominal or adjectival modifiers” (Haspelmath 2023: 4), we see that this is violated by the fact that hashtags sometimes expand on existing versions, as seen for example with *#getvaccinated* and *#getcovidvaccinated*. Added to that, in order to interpret their meaning, users need to split them into component words and parse their internal structure. For those English compounds which are headed, their head occurs to the right (e.g. the head of *lipstick* is *stick*, the head of *bus stop* is *stop*), yet many hashtags seem to have their head to the left (e.g. *#getvaccinated*). This is not in keeping with English compounding rules. For these reasons, we support Caleffi’s (2015) term *hashtagging* and reject a compound analysis.

We propose that hashtags fit the notion of “word-hood” and that the morphological process which gives rise to them is distinct from lexical compounding (we support the notion of *hashtagging*). Like other word phenomena, we also find that some hashtags fit the definition of word-hood better than others, particularly those that are syntactically integrated. At the other end of the continuum, we find hashtags used tweet-finally which are syntactically unintegrated and which function as a peripheral construction aptly described as a “conversational aside” (Zappavigna 2015: 278).

Cross-linguistic studies of hashtags (e.g. Caleffi 2015) and studies from other platforms and genres are needed to further increase understanding of this new language element. Given the salience of hashtags and their increased use over time, future research is set to tackle new questions in this area of rich lexical innovation.

Acknowledgements: We are grateful to the University of Waikato ALPSS Research Fund, to David Trye for comments on the draft and for his Python code and help in extracting the data, and to the anonymous referees for their helpful comments. All remaining errors are of course our own.

References

- Bauer, Laurie. 2017. *Compounds and compounding*. Cambridge: Cambridge University Press.
- Burnette, Jessie & Andreea S. Calude. 2022. Wake up New Zealand! Directives, politeness and stance in Twitter #Covid19NZ posts. *Journal of Pragmatics* 196. 6–23.

- Burnette, Jessie & Maebh Long. 2022. Bubbles and lockdown in Aotearoa New Zealand: The language of self-isolation in #Covid19nz tweets. *Medical Humanities* 49. 93–104.
- Caleffi, Paola-Maria. 2015. The “hashtag”: A new word or a new rule? *SKASE Journal of Theoretical Linguistics* 12(2). 46–69.
- Dehé, Nicole & Yordanka Kavalova. 2007. An introduction. In Nicole Dehé & Yordanka Kavalova (eds.), *Parentheticals*, 1–25. Amsterdam: John Benjamins.
- Fiesler, Casey & Nicholas Proferes. 2018. Participants’ perceptions of Twitter research ethics. *Social Media + Society* 4(1). 1–14.
- Haspelmath, Martin. 2011. The indeterminacy of word segmentation and the nature of morphology and syntax. *Folia Linguistica* 45. 31–80.
- Haspelmath, Martin. 2023. Defining the word. Unpublished manuscript. https://www.academia.edu/download/95034120/Defining_the_word.pdf (accessed 15 May 2023).
- Long, Maebh, Andreea S. Calude & Jessie Burnette. In press. Perceptions of vaccination hazards and pandemic risk in #Covid19NZ tweets. *Medical Humanities*.
- Maity, Suman Kalyan, Ritvik Saraf & Animesh Mukherjee. 2016. #Bieber+#Blast=#Bieberblast: Early prediction of popular hashtag compounds. In *Proceedings of the 19th ACM conference on computer-supported cooperative work & social computing*, 50–63. New York: ACM.
- Messina, Chris. 2007. Groups for Twitter; or a proposal for Twitter tag channels. <https://factoryjoe.com/2007/08/25/groups-for-twitter-or-a-proposal-for-twitter-tag-channels/> (accessed 19 May 2022).
- Page, Ruth. 2012. The linguistics of self-branding and micro-celebrity in Twitter: The role of hashtags. *Discourse & Communication* 6(2). 181–201.
- RStudio Team. 2020. *RStudio: Integrated development for R*. Boston: RStudio. Available at: <http://www.rstudio.com/>.
- Scott, Kate. 2015. The pragmatics of hashtags: Inference and conversational style on Twitter. *Journal of Pragmatics* 81. 8–20.
- Scott, Kate. 2018. “Hashtags work everywhere”: The pragmatic functions of spoken hashtags. *Discourse, Context & Media* 22. 57–64.
- Statista. 2020. LinkedIn statistics. <https://www.statista.com/topics/951/linkedin/#dossierKeyfigures> (accessed 7 March 2022).
- Trye, David, Andreea S. Calude, Felipe Bravo-Márquez & Te Taka Keegan. 2020. Hybrid hashtags – #YouKnowYoureAKiwiWhen your tweet contains English and Māori. *Frontiers in Artificial Intelligence* 3. <https://doi.org/10.3389/frai.2020.00015>.
- Venkit, Pranav, Zeba Karishma, Chi-Yang Hsu, Rahul Katiki, Kenneth Huang, Shomir Wilson & Patrick Dudas. 2021. A “sourceful” twist: Emoji prediction based on sentiment, hashtags and application source. *arXiv.org*. Available at: <http://arxiv.org/abs/2103.07833>.
- Wagh, Rashika & Payal Punde. 2018. Survey on sentiment analysis using Twitter dataset. In *2018 Second international conference on electronics, communication and aerospace technology*, 208–211 (accessed 23 May 2023).
- Wickham, Hadley. 2016. *Ggplot2: Elegant graphics for data analysis*. New York: Springer. Available at: <https://ggplot2.tidyverse.org>.
- Zappavigna, Michele. 2011. Ambient affiliation: A linguistic perspective on Twitter. *New Media and Society* 13. 788–806.
- Zappavigna, Michele. 2012. *Discourse of Twitter and social media: How we use language to create affiliation on the web*. London: Bloomsbury.
- Zappavigna, Michele. 2015. Searchable talk: The linguistic functions of hashtags. *Social Semiotics* 25(3). 274–291.

Q5

Q6

Q7