*Significances (Signifying)*, *5*(1), 128-162.
DOI : https://doi.org/10.52497/significances.v5i1.314

# Relationship with a semioreferential object as a predictor variable: The case of insiders' and outsiders' uses of "very New York" on Twitter

Gabriel Frazer-McKee[1] & Bruno Courbon[2] [3]

## Résumé

*Nous examinons ici plus d'un millier d'occurrences de la construction very New York utilisée sur Twitter par des personnes qui appartiennent ou non à la communauté new-yorkaise. L'analyse s'appuie sur des caractéristiques constructionnelles, sémantico-pragmatiques et métalinguistiques. Nous faisons entre autres ressortir la valeur majoritairement positive de la construction, ainsi que la richesse des représentations convoquées et la récurrence de leurs relations (p. ex. la mode, l'alimentation). Nous mettons en évidence, en recourant à la régression multinomiale, que l'appartenance à la communauté new-yorkaise prédit plusieurs caractéristiques linguistiques de la construction, même lorsque sont prises en considération les variables sociodémographiques disponibles. Les résultats montrent que la relation des utilisateurs avec l'objet sémioréférentiel « New York » influence très probablement leur façon de donner du sens à la construction. L'étude souligne l'importance, pour étudier des phénomènes sémantiques, de prendre en compte de façon détaillée les caractéristiques relatives aux participants des médias sociaux*

**Mots-clés** : linguistique sociocognitive ; sémantique ; corpus de réseau social ; construction lexicale

## Abstract

*We investigate more than a thousand instances of the construction very New York produced by insiders and outsiders on Twitter. The analysis bears on constructional, semantico-pragmatic, and metalinguistic features. Amongst other things, we find that the construction has typically positive valence and is used to exploit a rich but recurring set of mental representations (e.g. fashion, food). Crucially, we show using multinomial regression that insider-outsider status predicts several of the construction's linguistic features, even when considering users' sociodemographic characteristics. These results suggest that users' relationship with the semioreferential object "New York" likely influences their ways of imbuing the construction with meaning. The study underscores the relevance of characterizing online corpora's contributors in detail when investigating semantic phenomena.*

**Keywords:** social cognitive linguistics; semantics; social media corpus; lexical construction

[1] Université Laval (Canada). 0000-0002-0860-6192

[2] Université Laval & CRIFUQ (Canada). 0000-0001-8201-1237

[3] Corresponding author: bruno.courbon@lli.ulaval.ca.

## 1. Introduction

"*I'm about to make a very New York tweet. Something only those who live in The Big Apple, as we New Yorkers call it, will understand*"[4] declares a Twitter user, thereby indicating both their affiliation with and intimate knowledge of New York City. Comparatively, another Twitter user –a self-identified outsider— writes: "*A very New York moment(?): our very reserved [Haitian] uber driver completely lit up upon learning we were Canadians and could speak a little (a LITTLE) French*". Underlying these uses of the construction "very New York" is a rich semantic network (What is *New York* about *a* <u>*New York tweet*</u> or *a* <u>*New York moment*</u>?) leveraged by persons who have (presumably) dissimilar experience-derived mental representations of the "same" object (New York City). The first tweet, for instance, hints at strong place attachment as well as strong group affiliation and insider-type knowledge. On the other hand, the second tweet –per the question mark parenthetically appended to the noun *moment*— suggests tentativeness with regards to the linguistic characterization of the experience that likely stems from her outsider-type knowledge.

Intuitively, individuals' experiences with a semioreferential object[5] are strongly associated with different individual or shared ways of talking about it (Courbon, 2018). However, it is not immediately obvious (1) how this speaker characteristic can be operationalized in corpus linguistics so that it might be fruitfully employed to investigate a research question at scale; and (2) whether an eventual operationalization of speaker-object experience is a significant predictor when considering other speaker characteristics (e.g. sociodemographic variables, such as sex, age, etc.). The questions before us are thus the following:

**Research question 1**: How can individuals' experiences with a semioreferential object be operationalized at scale?

**Research question 2**: When controlling for relevant sociodemographic variables (e.g. sex, age, educational attainment), is the operationalization of individuals' experiences with a semioreferential object a significant predictor of the interactional, semantico-pragmatic or other semiotic characteristics of a particular linguistic construction (here, *very New York*)?

This exploratory study is structured as follows. In section 2 (Methods), we justify our choice of research object ("degree adverb + city name"), and explain how *very New York* was selected as our specific research object. We then present our corpora (1200+ tweets containing uses of *very New York*, of which ≈600 tweets contain metalinguistic uses of *very New York*), provide detailed sociodemographic information regarding the study's Twitter users, and explain the manual annotation methods used to structure the data. Twitter users exploit the construction to make sense of and characterize their experiences and impressions relating to New York City. Particular attention is thus paid to the construction's formal, semantico-pragmatic and semioreferential features (Courbon, 2015, 2018), as well as to metalinguistic markers that frequently accompany it. Then, in Results (section 3), using multivariate statistical analyses, we show that speaker-object experience is –more so than any other speaker characteristic considered in this study— significantly associated with a variety of constructional, semantico-

[4] The data presented between quotes were produced in one or the other of the Twitter corpora. Original spelling was preserved.

[5] By "semioreferential object", we mean the relatively abstract mental object which is referred to via use of a particular sign (Courbon, 2020 : 19). Here, the semioreferential object is the mental object specifically associated with the sign *New York* (rather than, for instance, mental representations evocable by *Big Apple*).

pragmatic, and other semiotic characteristics. We conclude with a Discussion (section 4). This paper offers a fruitful (if embryonic) methodological framework for relating linguistic features with speakers' experiential characteristics, which is seldom done in corpus-based semantics.

## 2. Object and methods

In the following subsections, we justify our choice of object of investigation (the linguistic construction "degree adverb + proper name"), and the specific constructional components selected, namely the degree adverb *very*, and the proper name *New York*. We also describe our dataset's compilation and the selection of datapoints to be analyzed, as well as the types of qualitative and quantitative analyses carried out on these datapoints.

### 2.1. Object of investigation and dataset construction

#### *2.1.1. Construction to be investigated*

Our object of study, *very New York*, is a typical case of the under-investigated "degree adverb + proper name" construction found in Present-day English (PdE), a construction whose semantic-pragmatics has –at least in the published literature— previously only been the object of isolated comments (Bauer & Huddleston, 2002 : 1657; Zwicky, 2006; Paradis, 2008 : 318; Audring & Booij, 2016 : 632; Beltrama, 2016 : 229–232; Breban, 2018; Stratton, 2018; Aarts, 2018 : 283; Trousdale, 2018; Duffley, 2020 : 177–178) and non-quantitative empirical studies (Wee & Tan, 2008; Bylinina, 2011; Gonzálvez-García, 2014; Sant, 2018). In this type of construction, a proper name (e.g. *LA*; *San Francisco*; *Paris*) –a linguistic unit that denotes a single individual or entity within a speech community and that provides access to encyclopaedic information associated with the nominal referent (Langacker, 2008)— occurs in the scope of a degree adverb –a functional unit such as *quite*, *very* or *so* that expresses intensity (e.g. *so LA*; *very San Francisco*; *quite Paris*). This type of construction is an interesting study case for a sociosemantic investigation because it exhibits formal and qualitative properties that can potentially be exploited and semantized dissimilarly by different persons[6], including:

(1) Occurrence in several syntactic schemas (e.g. *How very LA*; *This is very LA; a very LA morning*; *very LA*) (Gonzálvez-García, 2014; Sant, 2018; Wee & Tan, 2008);
(2) Predication of a logical or grammatical subject via a variety of copular verbs (e.g. *This feels very LA*; *She talks quite LA*; *He is very LA*) (Frazer-McKee, 2020; Gonzálvez-García, 2014);
(3) Metonymic exploitation of the proper name's encyclopaedic potential (Bylinina, 2011; Duffley, 2020; Frazer-McKee, 2020; Frazer-McKee & Duffley, preprint), as in *Pizza is very New York* (metonymic access to the notions of food and pizza) or *Black clothing is very New York* (metonymic access to the notion of fashion and popular fashion colours);

[6] The individuals who exploit this construction do so in part for the purposes of self-presentation (cf. Beltrama, 2016). Moreover, their experiences with the referential object presented through instances of the construction should be considered, which makes them more multifaceted than would otherwise be abstract, idealized "speakers". This justifies the use of the term *persons* here rather than, for instance, "speakers" or "agents".

(4) Valence-loaded characterization of a logical or grammatical subject (Gonzálvez-García, 2014), as in *That's so LA; love that city!* (posive affect) or *Ugh, how so very LA* (negative affect);
(5) Amenability to various metalinguistic manipulations (as we shall see in this study) to achieve different expressed meanings, such as intensification upscaling (e.g. *very very LA; *very* LA*), and various orthographic means of expressing emphasis (e.g. *VERY LA*; *very *LA**; **very LA**) or distancing (e.g. *very "LA"*), as well as full or partial occurrence as reported speech (e.g. *She said she was "very LA"*).

### *2.1.2. Data source selection*

The social media platform Twitter –the "fruitfly" of the social sciences (Tufekci, 2014)— was selected as the study's data source over other possible data sources, such as existing English-language mega-corpora (e.g. COCA), other social media platforms (e.g. Facebook) or data identified via the Google search engine. As a "communicational framework" (Nicolaï, 2022, this issue), Twitter presented a threefold advantage: (1) the construction "*very* + large Northern American city name" occurs tens of thousands of times on this platform (unlike in existing mega-corpora; Frazer-McKee, 2020 : 121-122); (2) data collection can target specific strings and be automated using officially-provided, easy-to-leverage software tools (unlike Facebook); and (3) unlike the oft-anonymous data gathered from Google, socio-demographic information is typically found in Twitter users' online profiles (e.g. user's self-declared geographic location, sex, ethnicity, etc.; Mislove *et al*., 2011). All data reported hereafter were collected by searching for specific string combinations via Twitter's API v2 in August 2021 (date span: 2010-2020).

### *2.1.3. Degree adverb selection*

Whereas degree adverbs often exhibit selectional preferences with gradable adjectives (e.g. *##very perfect* vs *completely perfect*; Paradis, 2001)[7], there are no known restrictions on the type of degree adverb that can participate in the "degree adverb + proper name" construction (e.g. *quite/very/so/completely Montreal*) (Gonzálvez-García, 2014). *Very* –the "prototypical" degree adverb (Buchstaller & Traugott, 2006 : 348)— was selected for the purposes of this study because it is not only amongst the most productive degree adverbs in PdE (Breban & Davidse, 2016), but also because it occurs in common syntactic configurations (e.g. Det + degree adverb + NN compound) not available to many other boosters and maximizers that also participate in the "degree adverb + proper name" construction (e.g. *##a so New York moment*), and because it is associated with less corpus noise than the more-studied *so* (sometimes called "Gen[eration]-X *so*"; e.g. Stange, 2020), which shares form with a conjunction that commonly collocates with proper names (e.g. *<u>So</u> Lady Gaga is in town...*). Lexically, *very* is a degree adverb that is commonly considered to be "bleached", as it still expresses a notion of high intensity but has lost the notion of truth it once denoted in Middle English (Stoffel, 1901 : 33; Bolinger, 1972 : 28; Israel, 2002 : 424; Lorenz, 2002 : 146; Breban & Davidse, 2016 : 239; for a rare dissenting view, see Bordet, 2017). In terms of scale semantics, the (in)felicity achieved by coordinating schemas (King, 2016) shows that *very* expresses greater intensity than a booster such as *quite* (e.g. *It's quite Montreal, but not very Montreal*) but lower intensity than a maximiser such as *completely* (*##It's completely Montreal, but not very Montreal*).

[7] "##" indicates infelicitous or zeugmatic linguistic combinations.

### *2.1.4. City name selection*

It was decided to investigate the name of a city, because user-provided location information could be readily exploited to divide the dataset into two broad groups – insiders and outsiders – constituted of persons whose different semioreferential relationships with NYC "colour" their use of the construction. Given the quantity of datapoints required to compare outsiders' and insiders' discourse while controlling for sociodemographic variables, we limited our investigation to a single "degree adverb + city name" sequence; the precise city name was to be determined mainly on the grounds of productivity but also taking into account insider-outsider ratio.

We gathered a "disposable corpus" of over 232, 000 tweets (2.2 million words) in which the degree adverb *very* was followed by the names (e.g. *Montreal*), acronyms (e.g. *Los Angeles → LA*), and diminutives (e.g. *Philadelphia → Philly*) of large Northern American cities (n=57) situated in urban areas with populations exceeding 1 million people (Demographia, 2021 : 19–30). Tweets with the following characteristics were excluded using the regex module for Python: retweets with no new relevant content, suffixed instances (e.g. *very Montrealish*), adjectival compounds (e.g. *very Montreal chic*), out of scope instances (e.g. *Yes, very. LA is the best*), and instances in which the city name was right-branching or part of a longer proper name (e.g. *very LA Times; very LA in the 1970s*). The cleansed tweet counts of a selection of these cities are reported in Figure 1.

*New York* was chosen as the city name to be investigated owing to the following considerations: its productivity (see Figure 1), the token's advantageous insider/outsider ratio (roughly 1:1), and the wide geographic distribution of its outsiders. From a less immediately quantifiable perspective, the city name *New York* also offers several other advantages for a sociosemantic study, most notably a rich convergence of sociolinguistic (Newman, 2014), sociocultural (Erenberg, 1984; Rantisi, 2004; Schloss, 2009), socioeconomic (Currid-Halkett, 2009), historical (Ellis, 2004), technological (Landau & Condit, 1999), and political dimensions. One of the few cities on the planet that can be called a *global* or *world city* (Abu-Lughod, 2000), many of these dimensions of New York have –crucially— been partially accessible to outsiders for decades via indirect means, such as television or films (e.g. Sadler & Haskins, 2005), and more direct means, such as tourism (e.g. Gilbert & Hancock, 2006; Phillips & Jang, 2010).

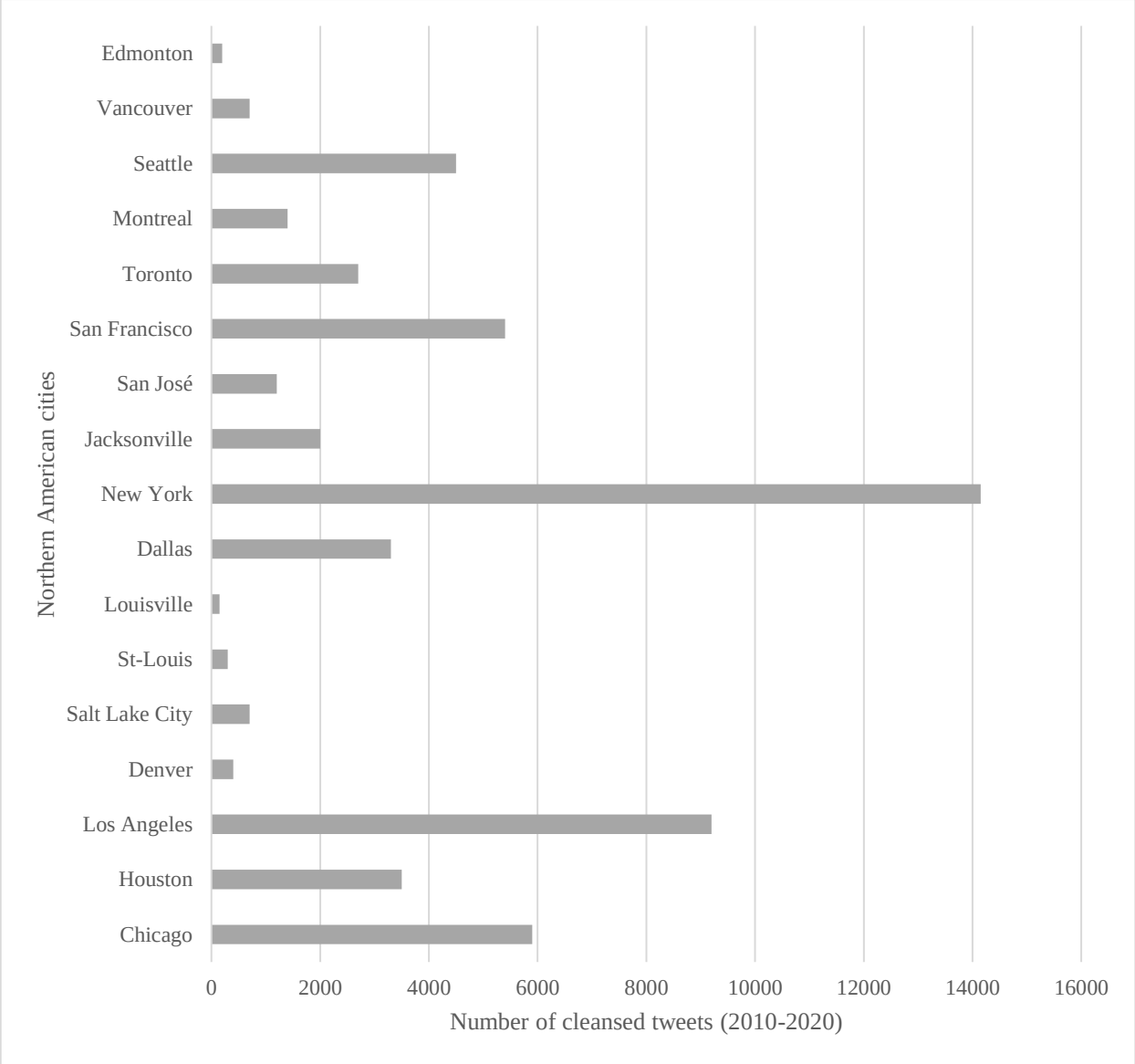


**Figure 1. Number of cleansed tweets for select city names in the "*very* + large Northern American city name" construction**

### *2.1.5. Datapoint selection*

Two partially overlapping datasets (i.e. Corpus1 and Corpus2) were exploited in this study.

Corpus1 served to investigate constructional information, various semantico-pragmatic dimensions, and sentiments expressed in tweets containing the lexical construction *very New York*. This first corpus was comprised of 600+ unique datapoints from the year 2017. The year 2017 was chosen because of its relative recency and also because synchronously-collected user location data was available for that specific year from an unpublished project.

The second dataset (Corpus2) was comprised of tweets containing specific metalinguistic features that occasionally accompany *very New York*, such as caps lock (e.g. *VERY New York*), diacritics (e.g. *very *New York**), and quote marks (e.g. "*Call and tell him he looks very "New*

*York" today!!*" or "*Trying to think of a very "New York" place to have dinner tonight.*"). Owing to the relative rarity of these metalinguistic features (around 14% of all tweets), Corpus2 was comprised of 594 tweets from the years 2010-2019. All metalinguistic tweets were identified using the regex module for Python.

While available and more recent, data from 2020 and 2021 were not considered for this study because a preliminary investigation suggested that the tweets' themes had been both directly (e.g. "*Wearing masks is very New York*") and indirectly (e.g. "*25 very New York things you can do during lockdown*") influenced by the COVID-19 pandemic, resulting in data that would likely have been thematically skewed compared to pre-pandemic tweets.

#### *2.1.6. Description of the corpora's Twitter users*

The study's two corpora are constituted of 1244 tweets produced by 1138 unique Twitter users' profiles; 95% (n=1083) of these users' profiles were personal accounts (see Tables 1 and 2 in Annex). Initial insider-outsider status was assigned based on the self-declared geolocation of the account (e.g. country, city, state). This initial classification could, however, be subsequently revised considering data from three different sources:

(1) Data found in the tweet itself (e.g. *"I'm from Brooklyn and this is very New York"*);
(2) Data found in the user's profile (e.g. "Born and bred in NYC!");
(3) Data found by *off-Twitter-matching*[8] the Twitter account to other means of self-presentation such as a LinkedIn résumé or a personal professional website.

144 users' initial insider-outsider status was recoded as a result of meticulously verifying each tweet for the above.

We also collected the following sociodemographic features for each of the corpus' 1138 personal accounts, when available:

(1) The apparent sex of the Twitter user, using a simple "male/female" coding scheme to classify their profile picture or biographical information found in their profile description, including preferred personal pronouns (e.g. "she/her"), parental status (e.g. "proud mom"), job description (e.g. "tech guy"), and gender-marked first names (e.g. "Kevin" or "Ana");
(2) The approximate age of the user (i.e. 20s, 30s, 40s, 50s, 60s, unknown), based on a visual assessment of their Twitter profile picture or their profile picture(s) found on other social media platforms. Apparent age estimation via visual inspection of photographs yields fairly accurate results, though it is vulnerable to some systematic biases, including "viewer's recent and normative exposure to faces" (Clifford *et al.*, 2018, online) as well as an "own age" bias (Rhodes & Anastasi, 2012);
(3) The apparent ethnicity of the account owner, using the following coding scheme adapted from Kapidzic and Herring (2014) to classify their profile picture (White; Black or African American; Hispanic or Latino (requirement: a Latino or Hispanic surname); South-East Asian; East Asian; Arab; No profile picture or other). Ethnicity

[8] Per Twitter's (2022) developer policy, off-Twitter-matching is permitted for research purposes on condition that the matching process rely only on public data (e.g. username, information found in public tweets or the user's profile, etc.). This condition was respected: no Twitter user in our corpus was contacted directly or indirectly by the authors to collect personal information for this study.

identification via visual inspection of photographs is a reasonably reliable technique, though it is vulnerable to certain biases, not least the well-documented "own-race-bias" (Meissner & Brigham, 2001);

(4) Whether the user had a university-level education, per their Twitter profile (e.g. "UCLA alumni") or their online professional résumé (e.g. LinkedIn).

Assessing the impact of these variables is especially important for a sociosemantic study such as ours because life experiences and interests –those things which tend to be characterized by this study's construction— are not solely predicted by insider-outsider status (if indeed they are at all), but are also associated with sociodemographic features, including sex/gender (e.g. male predilection for things, female predilection for people; Su *et al*., 2009), ethnicity (e.g. musical or sports interests; see Ogden & Hilt, 2003), educational attainment (e.g. active citizen behaviours; Hoskins *et al*., 2006), and age (e.g. time allotted to leisure activities; Verbrugge *et al*., 1996).

The users' sociodemographic characteristics are presented in Tables 1 and 2 in Annex. As the corpora are treated separately (see 3. Results), we present the users' characteristics in separate tables.

A notable characteristic of outsiders in both corpora is that they are largely located in the so-called "inner-circle" countries (Kachru, 1985) of the English-speaking world (i.e. United States, United Kingdom, Canada, Ireland, New Zealand, and Australia), while a key feature of insiders is that they are overwhelmingly –as expected— located in NYC itself.

Insider and outsider groups in both corpora also share several characteristics. In both corpora, insiders are more educated overall than are outsiders, a significantly greater percentage of insiders than outsiders were off-Twitter-matched to a reliable work experience source such as LinkedIn (Guillory & Hancock, 2012), a majority of both insiders and outsiders declared an American location in their profile, and a majority of insiders and outsiders were (a) White and (b) young adults in their 20s or 30s.

As the tweets of the two corpora are produced by almost entirely different sets of users, it is not surprising that the corpora's users also exhibit some important sociodemographic dissimilarities. In Corpus1, for instance, sex is predictive of insider-outsider status ($p<0.05$) according to a Fisher exact test, a majority of insiders being women, and a majority of outsiders being men. While the male-to-female ratio does not diverge significantly in the metalinguistic corpus (Corpus2), ethnicity is not as well balanced as in Corpus1: African American users are significantly more likely to be insiders in Corpus2 ($p<0.001$), whereas East Asian users are more likely to be outsiders ($p<0.001$); ethnicity is not predictive of group membership in Corpus1. However, unlike Corpus2, Corpus1's groups exhibit significant age differences: more insiders were estimated to be in their 20s ($p<0.001$), while more outsiders were estimated to be in their 30s ($p<0.001$).

### 2.2. Analyses

In the following subsection, we describe the dimensions of the "behavioural analysis" conducted on the study's tweets containing the construction *very New York* (henceforth, *the Cx*), namely (1) constructional analysis; (2) semantico-pragmatic analysis; (3) sentiment analysis; and (4) metalinguistic analysis. Microsoft Excel (Microsoft Corporation, 2018) was used to annotate each datapoint. Full definitions and examples of the study's annotated

variables are available as supplementary material in a 20+-page document hosted on www.protocols.io (see Frazer-McKee & Courbon, 2022). GFM conducted the constructional, semantic, part of the metalinguistic, and sentiment analyses independently, while BC conducted the metalinguistic, and part of the semantic analyses independently. BC independently verified GFM's constructional, semantic, and sentiment analyses, and GFM independently verified BC's metalinguistic analyses. Disagreements were resolved via consensus.

#### *2.2.1. Behavioural profiling*

Behavioural[9] profiling is a corpus-based semantic method that involves manually annotating each datapoint of a corpus for both formal and qualitative features in view of conducting statistical analyses to identify important patterns in the dataset (see Gries & Divjak, 2009). Behavioural profiling was used to characterize the constructional and semantic dimensions of each tweet.

The constructional analysis sought to characterize the Cx mainly on the level of linguistic form (though we recognize that there are form-meaning correlations involved with the Cx, so that there is inevitably overlap between syntactic and semantic levels of analysis –e.g. negation, NN compounds). The following variables were coded for:

(1) Whether the Cx' subject was expressed lexically (e.g. *This guy is very New York!* vs *very New York!*);
(2) Whether the construction's sentence contained a negation (e.g. *That is not very New York*);
(3) The lemmatized form of the verb, if any, found in the sentence that contained the Cx (e.g. *looks very New York to me!*), and whether the verb was a perception verb (e.g. *seem* or *look*);
(4) Whether the Cx occurred in a broader paratactic construction (e.g. "*very college educated, very left-lea*[*n*]*ing, very New York, very media-savvy, very middlebrow*") and what the other elements in the paradigmatic set were (e.g. "*very college educated, very left-lea*[*n*]*ing, very New York, very media-savvy, very middlebrow*");
(5) Whether the Cx' degree adverb was preceded by another degree device (e.g. "*traffic is so very New York!*", "*She would know. She's very very New York.*");
(6) The type of syntactic schema in which the construction occurs (see Table 3).

[9] While we endorse Behavioural Profiling as a method of corpus investigation, the term "behavioural" is unfortunate to our minds, as it not only denies the contribution of the annotator's perceptions and theories to data analysis but also unwittingly suggests that the (linguistic) object of investigation has agency.

| Schema | Illustrative examples |
|---|---|
| Cx | (a) *Very New York!*<br>(b) *Very New York of you!*<br>(c) *Very New York, isn't it?* |
| (Det) Cx NP | (a) *Very New York tweet*<br>(b) *a very New York day*<br>(c) *some very New York guy* |
| (X) VP Cx | (a) *Feels very New York*<br>(b) *This feels very New York*<br>(c) *It's very New York of them* |
| *How* Cx | (a) *How very New York!*<br>(b) *How very New York of him!*<br>(c) *How very New York to lose like that...* |

**Table 3**. **The main syntactic schemas in which the Cx occurs**

The semantic analysis, on the other hand, sought to characterize the semantico-pragmatic dimensions of the Cx. This analysis bore upon (1) the types of referential objects found within the scope of the Cx (e.g. animates, events), and (2) a selection of broader situation features (e.g. objects, participants) evoked in the tweet containing the Cx that were relevant to its expressed meaning.

As previously discussed, the "degree adverb + proper name" construction is used to characterize another linguistic item (e.g. "*a very New York <u>thing</u>*"). The element found in the referential scope of the Cx was therefore annotated for the following:

(1) Whether it was (a) linguistically-expressed in the same syntactic schema as the Cx (e.g. "*Venus on her "very New York" <u>dress</u>*"), (b) linguistically-expressed in a different syntactic schema than that in which the Cx occurred (e.g. *Very New York. <u>Her dress</u>, I mean*) or (c) whether the element characterized by the Cx had to be inferred (for example, from a photograph);
(2) How the lexical item in the Cx' referential scope was construed *in situ* (e.g. in "*a very New York <u>opening</u>*", "*opening*" refers not to the opening in a wall but to the inauguration of a new subway line; "*thing*" in "*a very New York <u>thing</u> happened to me*" refers to a type of event, rather than to a physical object);
(3) The metatype of the referential object found in the Cx' scope (see Table 4);
(4) The subtype of the referential object found within the Cx' scope (see Table 5).

| **Referential object metatype** | **Common subtypes** |
|---|---|
| Animate | (a) Human<br>(b) Animal<br>(c) Other animate |
| Event | (a) Speaker involved as agent<br>(b) Speaker involved as experiencer<br>(c) Speaker involved as observer, either directly (e.g. witnessed the situation) or indirectly (e.g. heard about the situation from someone else) |
| Concrete thing | (a) Food<br>(b) Place<br>(c) Fashion item<br>(d) Other concrete thing |
| Abstract thing | (a) Behaviour<br>(b) Impression (of a person, a place, etc.)<br>(c) Internal state (e.g. attitudes, opinions, thoughts, emotions)<br>(d) Other abstract thing |
| Temporal expression | (a) Moment<br>(b) Period of day (e.g. "*this evening*", "*this afternoon*")<br>(c) Period of week (e.g. "*today*", "*tomorrow*", "*last weekend*", "*Monday*")<br>(d) Other temporal expression (e.g. seasons) |
| Semiological object | (a) Linguistic feature of New York English (e.g. words, accent)<br>(b) Audio, visual or audiovisual object (e.g. photograph, video)<br>(c) Linguistically-expressed content (e.g. a novel, a story, a sentence)<br>(d) Other semiological object |

**Table 4**. **The metatype and common subtypes of generic conceptual categories found within the Cx' referential scope**

Given that the Cx is simultaneously motivated by and used to characterize some aspect(s) of a real-world or imagined situation, information regarding the situation implied or described in each tweet was also coded for. Following classic approaches to the ontology of situations/events (Fillmore, 2006), situations were considered to be constituted of one or more of the following dimensions:

(1) Participants (i.e. the animates involved, including animals);

(2) Time (e.g. "*tonight*", "*before I was out of diapers*");
(3) Place (e.g. "*Times Square*", "*on the back of the ferry*", "*on the subway platform*", "*in the streets of NY*").

These dimensions could be either explicitly coded or implicitly evoked or referred to, in, for instance, a photograph (labels: Lexically explicit time; Implicit time; Explicit locations; Implicit locations).

To better capture the specificities of situations evoked by the Twitter users when using *very New York*, we also annotated each tweet for the following:

(5) Whether the Twitter user was directly involved in the situation (e.g. "*Today in very New York things: Dragged myself back into city after work, ate late evening pasta* […] *next to Alec Baldwin*");
(6) Salient situation objects (e.g. "*i ate a slice of pizza very late in the street*");
(7) Whether there were any ethnocultural properties present in the situation, such as references to immigration, ethnicity, foods (e.g. bagels, pizza) or music with strong ethnic associations (e.g. rap, hip-hop, Latin music), ethnic diversity or ethnic or ethnocultural mingling ("melting pot" phenomena).

All told, each tweet from Corpus1 was annotated for 20 constructional and semantico-pragmatic variables that had a combined total of 55 predefined levels.

#### *2.2.2 Sentiment and humour analysis*

In addition to the constructional and semantic variables mentioned in 2.2.1, Corpus1's tweets were manually assigned one of the following sentiment tags:

(1) Positive;
(2) Negative;
(3) Very negative;
(4) Neutral;
(5) Indeterminate.

This classification bore on the Tweet's overall valence and rested on (1) valence-loaded lexical items (e.g. "*Great value, fast service, large portions, an old fashion deli style restaurant, very New York.*" → positive; "*How very New York. Teach those poor misguided kids how to hate life*" → negative; and (2) valence-indicative constructional elements, such as punctuation, caps lock and the syntactic schema "How Cx" (e.g. "*VERY New York!*" → positive; "*How* Cx + full stop" → negative). Very negative valence was assigned to tweets that contained especially marked lexical items or personal attacks. Additionally, tweets were assessed for explicit (e.g. "*the experience was VERY new york LOL*") and implicit indications of humour (e.g. evocation of an absurd scenario, such as being told one possesses "*very New York eyes*").

#### *2.2.3. Metalinguistic marker analysis*

The metalinguistic corpus' tweets (Corpus2) were annotated for the following:

(1) Focus on a specific segment within the construction *very New York*: (a) Degree adverb (e.g. *"Whatever you think of him, he's VERY New York"*; **very* New York*); (b) City name (e.g. "*im feeling very NEW YORK right now!*"); (c) Degree adverb + city name (e.g. "*Listening to the VERY NEW YORK 'with or without you' u2 song*");
(2) The material means of achieving the focus which contains *very New York*: (a) typographical character (e.g. *very *New York**); (b) caps lock (e.g. *very NEW YORK*); (c) quote marks (e.g. "*It is very "New York" and def a must visit!!!*"), which may indicate a distancing effect (Martin & White, 2005 : 213), or at least polyphonic effects (see "*hétérogénéité(s) énonciative(s)*" in Authier-Revuz, 1984);
(3) Whether the Cx was used as direct reported speech (DRS) between quote marks: DRS is found between proper quotation marks (e.g. "*I just got told I had "a very New York look"*"), with or without an explicitly identified enunciative source;
(4) Whether there was expressive punctuation after the Cx, such as exclamation marks or ellipsis (e.g. "*VERY new york CLASSIC!!*");
(5) Whether there were emoticons in the tweet (e.g. "*It's massive, it's very 'New York' in some ways but still very 'Toronto' idk how to describe it aha just come and see :)*".

Additionally, Corpus2's tweets were analyzed for select semantic, constructional and pragmatic features:

(1) The type of syntactic schema in which the Cx occurs (see Table 3);
(2) The type of verb, if any, used to predicate the Cx of a logical or grammatical subject (e.g. *be*, *dress*, *feel*);
(3) The metatype and subtype of the referential object characterized by the Cx (Table 4);
(4) The tweet's overall valence, and the presence/absence of humour (see 2.2.2);
(5) Whether the Twitter user was present in the tweet;
(6) Whether the tweet was explicitly addressed to someone in particular.

## 3. Results

In the following subsections, we present the results of our analyses. We begin by presenting the study's overall findings, and then present individual analyses in detail.

### 3.1. Overall findings

The aim of the statistical analyses was to determine whether insider-outsider status is a significant predictor of a linguistic outcome (e.g. type of syntactic schema used, use of humour) while simultaneously controlling for the influence of sociodemographic predictors (e.g. age, sex); assessing other predictors' potential influence was necessary, as it is possible that insider-outsider status (a novel variable) is not a significant predictor when simultaneously considering the impact of other predictors. Binomial regression was used when the dependent variable had only two levels (e.g. interaction or lack thereof with another user). Multinomial regression was employed when the dependent variable (e.g. type of object characterized) had three or more levels (e.g. animate, concrete thing, abstract thing). Multinomial regression is a currently "underutilized" (Gries, 2021 : 343) statistical technique, but it holds considerable interest for

corpus-based semantics, because it can handle nominal data while making no assumptions about the data's normality, linearity or homoscedasticity. Nagelkerke's $R^2$, a measure of the extent to which the predictors account for variation in the data, was computed for both types of statistical tests; Nagelkerke's $R^2$ is an index between 0 (no variation explained) and 1 (all variation explained).

Binomial and multinomial regression analyses with various linguistic features (e.g. types of syntactic schema, referential object metatype) as dependent variables were performed with 5 predictors as covariates:

(1) Age (20s; 30+; binary nominal variable);
(2) Sex (male/female; binary nominal variable);
(3) Ethnicity (white/racialized; binary nominal variable);
(4) Education (university-educated/other or unknown; binary nominal variable);
(5) Insider-outsider status (insider/outsider; binary nominal variable).

For statistical purposes, Age and Ethnicity had to be transformed into variables with fewer levels compared to how the data was collected and reported in Tables 1 and 2 in Annex, to avoid difficulties associated with data sparseness (e.g. very few Asian users; few users aged 40+). Variables were dummy-coded as follows.

| **Predictor** | **Level** | **Code** |
|---|---|---|
| Insider-outsider status | Outsider | 0 |
| | Insider | 1 |
| Age | 20s | 20 |
| | 30+ | 30 |
| Sex | Male | 0 |
| | Female | 1 |
| Ethnicity | White | 0 |
| | Racialized | 1 |
| Education | Unknown or high school | 0 |
| | University | 1 |

**Table 5. Predictors, and corresponding levels and codes**

All analyses were performed using SPSS v27 (IBM Corp, 2020). Per best practices, all predictors and variables with a count of less than 15 were omitted from the analyses (Tabachnik & Fidell, 2019). No collinearity was observed between the predictors.

It should be noted that while the two corpora were subjected to somewhat overlapping annotations (e.g. sentiment analysis), the datasets were not pooled, owing to the fundamentally different basis on which each corpus was gathered: datapoints in Corpus1 were selected on the basis of their date of publication (i.e. 2017), while the metalinguistic datapoints in Corpus2

were selected on the basis of form (e.g. use of quotation marks, use of emphasis markers such as caps lock).

A total of 25 types of dependent variables were selected for multivariate analysis (see Figure 2); variables that were not selected for statistical analysis were insufficiently frequent to be processed statistically (e.g. use of negation or question form). A p-value heatmap of the odd ratios for each predictor in relation with each of the dependent variables is reported in Figure 2 below; significant relationships are in dark grey, while near-significant ones are in light grey.

| | **Predictors** | | | | |
|---|---|---|---|---|---|
| **Types of dependent variables** | **Insider/Outsider status** | **University education** | **Sex** | **Age** | **Ethnicity** |
| Constructional variables | | | | | |
| Syntactic schema | p=.032 | p=.006 | p=.060 | | |
| Use of verb | | p=.110 | | | |
| Verb tense | p=.053 | | | | |
| Semantico-pragmatic variables | | | | | |
| Referential object metatype | p<.001 | | p<.001 | | p=.015 |
| Sentiment | | | | | p=.033 |
| Use of humour | | | | | |
| Ethnocultural dimensions | p=.005 | | p=.041 | | p<.001 |
| Involved in situation | p=.045 | | p=.001 | | |
| Talking with other users | p<.012 | p=.072 | | | p=.070 |
| Lexically explicit time | | | p=.020 | | |
| Implicit time | | | | | |
| Explicit locations | | | p=.046 | | |
| Implicit locations | | | | | |
| Salient situation objects | | | | | |
| Paradigmatic sets | p=.044 | | | | |
| Explicit participants | | | | p=.092 | p=.004 |
| Implicit participants | p=.047 | | | | p=.046 |
| Metalinguistic variables | | | | | |
| Focus | | | | | |
| Means | | p=.008 | | | |

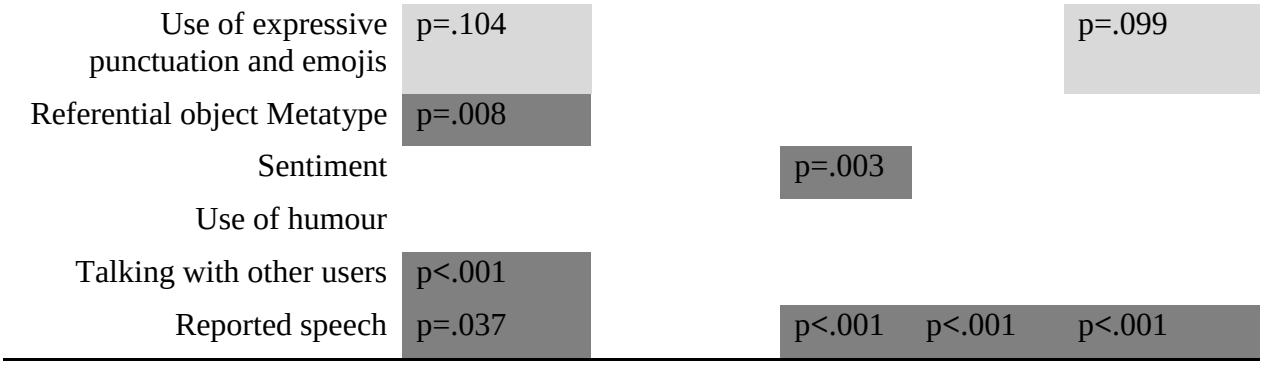


**Figure 2. Heatmap of the odd ratios for the relationship between predictors and types of dependent variables (α=5%)**

Insider-outsider status was found to predict several linguistic outcomes. For instance, insiders were found to combine a verb with the Cx more frequently and to more often be personally involved in the situation described than outsiders (e.g. "*I write accounts because I talk very New York*"). Outsiders' tweets, on the other hand, were more likely to contain ethnocultural dimensions (e.g. "*How very New York wake up for breakfast and encounter a (Vietnamese American) parade a block from my hotel.*"), and to feature interactions with other Twitter users (e.g. "*I know what you mean...that's a very New York feeling for me.*". Surprisingly, insider-outsider status did not predict tweet valence. While one might expect insiders to speak about NYC more positively than outsiders ("in-group bias") or outsiders to use *very New York* more negatively than insiders ("out-group bias"), only ethnicity (Corpus1) and sex (Corpus2) predicted tweet valence: white users were found to express more negative sentiments than racialized users, and women were found to express negativity more than men (Corpus2). Both insiders and outsiders spoke of NYC in very positive terms (53 % of the tweets in Corpus1 and 60 % in Corpus2 are positive, whereas only 24 % of the tweets in Corpus1 and 14 % in Corpus2 are negative or very negative). Nevertheless, per Figure 2, insider-outsider status is the most important predictor amongst the 5 predictors considered in this study, being a significant or near significant predictor of 12 of the 25 dependent variables. By way of contrast, sociodemographic variables are only predictive in a few cases. This is not a surprising result, given that, for instance, sex does not determine topic of conversation.

In the following subsections, we present these outcomes in detail, though we necessarily limit ourselves to variables we feel are especially important, owing to space considerations.

### 3.2 Constructional variables of interest

The construction is necessarily deployed in a syntactic schema, hence our interest in this constructional variable. Of particular interest are differences between insider and outsider uses of NN compounds –a syntactic schema that is most often used to denote a ready-made category (Plag, 2003), as in *a dog house* or *a New York thing*.

Per Table 5, certain nominal heads (e.g. *Christmas*, *tweet*) are used with roughly equal frequency by both insiders and outsiders, but there are also indications of conceptual differences between insiders and outsiders at work on two levels. Firstly, insiders use a greater variety of recurring nominal compounds (n=14) than do outsiders (n=9), suggesting that insiders may

associate a greater number of pre-made categories with the referential object than do outsiders. Secondly, insiders and outsiders do not exploit the same categories with equal frequency. Per a 2-by-3 Fisher exact test that compared the 3 most frequent of these nominal heads (i.e. *thing*, *moment*, and *accent*/*voice*) in insider and outsider discourse, insiders are more likely to use *moment* than outsiders ($p=0.014$), and outsiders are more likely than insiders ($p<0.0001$) to refer to a "*very New York accent*" or a "*very New York voice*".

Intriguingly, we noted (rare) cases in which the Cx' material form suggested apparent semiotization, through either the use of title case (e.g. "*This nonstop jackhammer in the queens public library is A Very New York Thing*"), or through the use of hyphens between the Cx' components (e.g. "*Thanks very-New-York lady loudly complaining about your kid's friends' Mom's actions at the bar*.").

It was found that university-educated users were more likely to prefer "Det Cx NP" over the reference category (i.e. "X VP Cx"). The association of university education with nominal compounds is not surprising, as greater use of nominal compounds is one of the hallmarks of academic writing (Biber & Gray, 2010 : 8).

| Insiders (n=148) | | | Outsiders (n=100) | | |
|---|---|---|---|---|---|
| NN head | n | % | NN head | n | % |
| **thing** | 34 | 22.97% | **accent / voice** | 12 | 12.00% |
| **moment** | 19 | 12.84% | **thing(s)** | 11 | 11.00% |
| day | 5 | 3.38% | **kind of / style of N** | 5 | 5.00% |
| **Christmas** | 5 | 3.38% | **tweet** | 4 | 4.00% |
| **tweet** | 5 | 3.38% | **Christmas** | 4 | 4.00% |
| story | 4 | 2.70% | feel | 3 | 3.00% |
| way of N | 4 | 2.70% | street fashion / style | 3 | 3.00% |
| feeling | 3 | 2.03% | couple | 2 | 2.00% |
| **kind of N** | 3 | 2.03% | **moment** | 2 | 2.00% |
| **voice / pronunciation** | 3 | 2.03% | | | |
| evening | 2 | 1.35% | | | |
| experience | 2 | 1.35% | | | |
| question | 2 | 1.35% | | | |
| scene | 2 | 1.35% | | | |

**Terms in bold are common to both insiders and outsiders

**Table 5. Ranking of NN compound heads that occurred 2 or more times in Corpus1**

While the odds ratio of insider-outsider status is a significant predictor of the syntactic schema dependent variable (Table 5), inspection of Table 6 below reveals that it is not actually

significantly associated with any particular syntactic schema (p>.10 in both cases**)**; this is because the odds ratio is a measure of a variable's *overall* effect rather than a characterization of its association with any *particular* dependent variable. Data sparsity may partly be in play here.

| **Dependent variable** | **Predictor** | **Co-efficient** | **p-value** | **Likelihood ratio p-value** |
|---|---|---|---|---|
| (Det) Cx NP | Insider-outsider status | .345 | .137 | **.032** |
| | Ethnicity | -.153 | .466 | .501 |
| | Sex | -.423 | .053 | .060 |
| | University education | .727 | **.002** | **.006** |
| | Age | .034 | .149 | .219 |
| Cx | Insider-outsider status | -.398 | .128 | **.032** |
| | Ethnicty | -.585 | .170 | .501 |
| | Sex | -.343 | .571 | .060 |
| | University education | .380 | .594 | **.006** |
| | Age | .043 | .158 | .219 |

Note: For all multinomial analyses, values in bold are significant at an alpha level of 5%

The reference category is "X VP Cx"; Nagelkerke $R^2$=.089

**Table 6. Multinomial analysis of predictors of syntactic schemas**

### 3.3. Semioreferential variables of interest

The Cx is used to characterize different kinds of referential objects, namely:

(1) **Animates**, especially humans (e.g. "*In case you ain't notice, I'm very New York*", or "*When people say someone is *very New York* what they mean is *they obviously have mental health problems**");

(2) **Semiological objects**, such as music, films, or any other fictional material (e.g. "*This is also a VERY New York story*.");

(3) **Concrete things**, such as food (e.g. "*im surprised you didn't know what an egg cream was they're a very new york/brooklyn thing*"), places (e.g. "*This bar is so very New York!*") or fashion items (e.g. "*I received the sweetest compliment yesterday from two ppl telling me my fashion was very 'New York'*.");

(4) **Events**, involving the Twitter user as observer (e.g. "*this seems like some very new york shit*"), as agent (e.g. "*I know it's my last day in NYC, but I DID do lots of very New York things yesterday and my feet hurt.*"), or as experiencer (e.g. "*I just had a great, very New York experience checking out a great gallery show*");

(5) **Abstract things**, such as behaviours (e.g. "*Apparently hailing an uber is "very New York" of me*") or impressions (e.g. "*Bought a dress for work and then left it on the bus! Had to catch a cab to catch the bus! I feel very New York City now.*");

(6) **Temporal expressions** (e.g. "*This week felt very "New York"*").

Two main rhetorical effects are produced via the Cx. The first effect is that of typicality, whereby the characterized object is placed into a relationship with something typically associated with New York (e.g. "*Pool parties after work is very New York*"). The second effect is one of comparison, whereby something associated with New York serves as a point of comparison for something external to New York, as in "*There is something very new york about san francisco*", "*What's weird about this place is it's very New York but the street outside is entirely quiet cos it's really Copenhagen.*", or "*A very New York Park*" (said of a park in Toronto, Canada).

The absolute frequency of the object metatypes and subtypes are found in Table 7 below.

| | **Total** | **Insiders** | **Outsiders** |
|---|---|---|---|
| **Animate** | **90** | **42** | **48** |
| Human | 87 | 41 | 46 |
| Animal | 3 | 1 | 2 |
| **Event** | **77** | **47** | **30** |
| Involved as observer | 40 | 25 | 17 |
| Involved as agent | 9 | 6 | 3 |
| Involved as experiencer | 26 | 16 | 10 |
| **Concrete thing** | **53** | **25** | **29** |
| Food | 9 | 4 | 5 |
| Fashion item | 10 | 5 | 5 |
| Place | 25 | 12 | 13 |
| Other concrete thing | 10 | 4 | 6 |
| **Semiological object** | **110** | **48** | **62** |
| Linguistically-expressed message | 37 | 26 | 11 |
| Audio, visual, or audiovisual object | 46 | 15 | 31 |
| Feature of NYC English | 26 | 6 | 20 |
| Other semiological object | 1 | 1 | 0 |
| **Abstract thing** | **70** | **32** | **38** |
| Behaviour | 39 | 20 | 19 |
| Internal states: Opinion, belief, thought, attitude, emotion | 11 | 7 | 4 |
| Impression (person, place) | 13 | 4 | 9 |

| | | | |
|---|---|---|---|
| Other abstract thing | 7 | 1 | 6 |
| **Temporal expression** | **41** | **33** | **8** |
| Moment | 21 | 18 | 3 |
| Holiday | 5 | 3 | 2 |
| Weekend | 2 | 1 | 1 |
| Holiday | 5 | 3 | 2 |
| Day | 5 | 5 | 0 |
| Season | 2 | 2 | 0 |

**Table 7. Metatypes and subtypes of the referential objects in the Cx' scope**

Multinomial logistic regression identified several significant relationships between predictors and dependent variables. Men were more likely to use the Cx to characterize abstract things or semiological objects, while women were more likely to characterize animates than either of these types of referential objects. Insiders were more likely to place an event or a temporal expression within the Cx' scope, while outsiders were more likely to do so with animates.

| **Dependent variable** | **Predictors** | **Co-efficient** | **p-value** | **Likelihood ratio p-value** |
|---|---|---|---|---|
| **Abstract thing** | Insider-outsider status | -1.437 | .301 | **.001** |
| | Ethnicity | -.793 | .708 | **.015** |
| | Sex | -.710 | **.002** | **.001** |
| | University education | -.803 | .448 | .584 |
| | Age | -.247 | .330 | .139 |
| **Concrete thing** | Insider-outsider status | .304 | .409 | **.001** |
| | Ethnicity | -.963 | **.028** | **.015** |
| | Sex | -.542 | .144 | **.001** |
| | University education | .012 | .621 | .584 |
| | Age | .071 | .467 | .139 |
| **Event** | Insider-outsider status | .782 | **.020** | **.001** |
| | Ethnicity | -.354 | .318 | **.015** |
| | Sex | -.927 | **.006** | **.001** |
| | University education | .166 | .621 | .584 |
| | Age | .024 | .467 | .139 |

| | | | | |
|---|---|---|---|---|
| | Insider-outsider status | -.231 | .454 | **.001** |
| **Semiological object** | Ethnicity | .167 | .596 | **.015** |
| | Sex | -1.198 | **<.001** | **.001** |
| | University education | .114 | .711 | .584 |
| | Age | .076 | **.018** | .139 |
| | Insider-outsider status | 1.790 | **<.001** | **.001** |
| **Temporal expression** | Ethnicity | .642 | .108 | **.015** |
| | Sex | -.444 | .411 | **.001** |
| | University education | .555 | .200 | .584 |
| | Age | .070 | .097 | .139 |

The reference category is "Animate"

Nagelkerke $R^2$=.111

**Table 8. Multinomial analysis of predictors of referential object type**

***3.3.2. Explicit paradigmatic sets involving the Cx***

The Cx was found to participate in several paradigmatic sets, including:

(1) **Locations**, often to evoke values associated with places (e.g. "*Is beinng called "very la," and upgrade from my usual "very Florida?" When will I be "very New York????"*"), or about different aspects of self, such as a sense of place attachment (e.g. "*Feeling VERY New York right now, but I am a Los Angeleno*") or aspects of personality (e.g. "*my personality is very new york, but my emotional need for open space and fresh air is very colorado. where do i belong?*") or behaviours (e.g. "*Tbh my living habits are very New York but my social habits are very philly.*");

(2) **Positive qualities**, about a person's personality (e.g. "*Very college educated, very left-lea[n]ing, very New York, very media-savvy, very middlebrow, and for lack of a better word, very "cool".*"), a person's appearance (e.g. "*I must look very approachable or very New York because people love asking me for directions now*"), about some broader entity (e.g. "*There's a whole culture. It's very gay. Very fabulous. Very New York.*"), or about a place (e.g. "*Happy to be back in New York. In a very new, very dark, very New York hotel.*");

(3) **Negative qualities**, most often behaviours attributed to New Yorkers (e.g. "*It's very New York. And sleazy.*" or "*I think it's very New York - sarcastic and either passive-aggressive or just aggressive.*"), but also personal characteristics (e.g. "*she doesn't look very New York or like an heiress. maybe this is her incognito style?*") and sensory experiences (e.g. "*Have you seen the state of this city? It's very New York, everything is dirty and smelly here!*");

(4) **Contrasting qualities**, that bear upon, for instance, someone's personality (e.g. "*People at my school think I'm tough and very New York but I don't think so AT*

*ALL I think I'm nice*"), or fashion items (e.g. "*I cropped the pic because Jessica's white tennis shoes were incongruous (though very New York).*").

These paradigmatic sets' raw frequencies are reported in Figure 3.

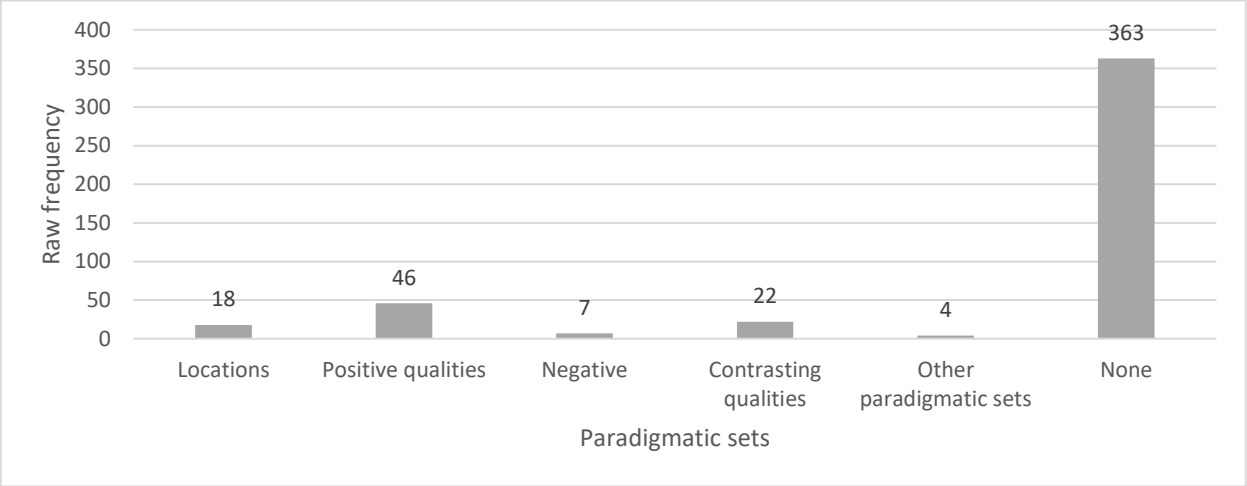


**Figure 3. The types of referential objects used with reported speech (Corpus1)**

The multinomial regression analysis revealed that insiders were more likely to use *very New York* contrastively, while outsiders were more likely to deploy the Cx into a set of positive qualities.

| Dependent variable | | Co-efficient | p-value | Likelihood ratio p-value |
|---|---|---|---|---|
| **Contrasting qualities** | Insider-outsider status | .410 | **.037** | **.044** |
| | Ethnicity | .089 | .886 | .948 |
| | Sex | -.504 | .373 | .809 |
| | University education | .296 | .583 | .188 |
| | Age | .567 | .404 | .720 |
| **Locations** | Insider-outsider status | .022 | .966 | **.044** |
| | Ethnicity | -.194 | .716 | .948 |
| | Sex | .192 | .704 | .809 |
| | University education | -1.096 | **.041** | .188 |
| | Age | -1.054 | .109 | .720 |
| | Insider-outsider status | -.823 | **.017** | **.044** |

| | | | | |
|---|---|---|---|---|
| **Positive qualities** | Ethnicity | .095 | .782 | .948 |
| | Sex | .120 | .712 | .809 |
| | University education | -.073 | .826 | .188 |
| | Age | -.025 | .485 | .720 |

The reference category is "None"

Nagelkerke $R^2$=.044

**Table 9. Multinomial analysis of predictors of paradigmatic sets**

### 3.4. Metalinguistic variables of interest

Of the three purely metalinguistic variables, only reported speech was found to be significantly associated with insider-outsider status (see Figure 2).

According to the multinomial logistic regression test, users who employed the Cx in reported speech (n=79) were more likely to be white, female, in their 20s, and outsiders (see Table 10).

The reason these particular predictors emerge as significant is likely correlated with the function that reported speech serves in this corpus, namely to compliment a person's physical appearance, as in "*Was just told I look "very New York" today...best compliment ever!*" or their fashion choice "*Robert just told me my outfit was "very New York*"".

| **Dependent variable** | **Predictor** | **Co-efficient** | **p-value** | **likelihood ratio p-value** |
|---|---|---|---|---|
| | Insider-outsider status | -.717 | **.032** | **.037** |
| **Yes** | Ethnicity | -3.754 | **<.001** | **<.001** |
| | Sex | 4.611 | **<.001** | **<.001** |
| | University education | .486 | .135 | .163 |
| | Age | 2.187 | **<.001** | **<.001** |

The reference category is "No"

Nagelkerke $R^2$ = .514

**Table 10. Multinomial analysis of predictors of reported speech**

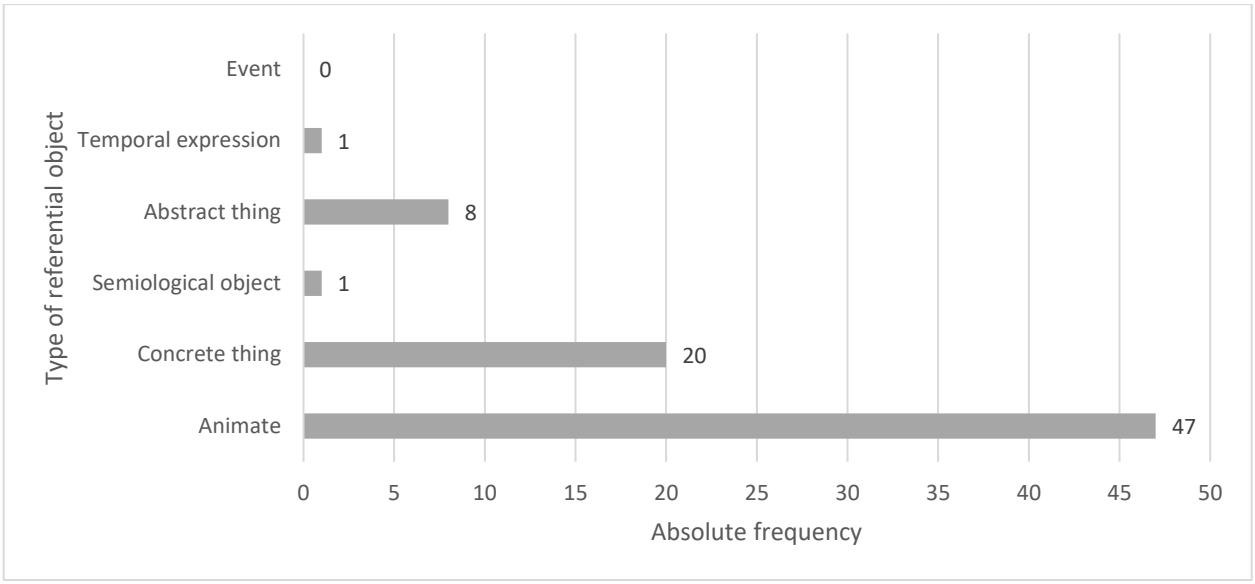


**Figure 4. The types of referential objects characterized in instances of reported speech**

While most instances of reported speech had positive valence, 14% (n=12) of users overtly expressed uncertainty regarding the expressed meaning of the Cx, as in "*My coworker just told me I look "very New York" today. Ok. Well I know I look very cute today. Is NY=cute?*" or "*Just been told I look very "New York" today. I have no idea what that means but I'm hoping it's a good thing*".

## 4. Discussion

Historically, the study of representations associated with a specific city –New York City, for instance— have usually been limited to the representations found in literary works (e.g. Rosenthal, 2016), media sources (e.g. Greenberg, 2009), and songs (e.g. Keeling, 2011), amongst other mediums. It is only with the advent of high literacy rates, widespread internet access, and social media platforms that it has become possible to cheaply investigate the linguistically-expressed representations of a city held by everyday language users. In this study, we investigated the representations that outsiders and insiders have of arguably "the best-known city on the planet" (Mayor Bloomberg, cited in Greenberg, 2003 : 410): New York City. This was done by investigating insider and outsider uses of the construction *very New York* in several hundred datapoints gathered from Twitter, a popular social media platform. The "*very* + city name" construction, an emerging construction in Present-day English, naturally lends itself to a sociosemantic analysis, as it allows speakers to exploit the proper name's semantic potential via a variety of syntactic schemas, and to convey a valence-loaded message. What is more, the construction is amenable to several interesting material manipulations (e.g. use of caps lock and typographic symbols for emphasis on one or even both of the construction's constitutive elements; use of hyphens between the components of the construction), but also different kinds

of interrogations the Twitter users have about the particular meaning or the valence of the construction.

Our study's research questions were the following:

> **Research question 1**: How can individuals' experiences with a semioreferential object be operationalized at scale?
>
> **Research question 2**: When controlling for relevant sociodemographic variables (e.g. sex, age, educational attainment), is the operationalization of individuals' experiences with a semioreferential object a significant predictor of the interactional, semantico-pragmatic or other semiotic characteristics of a particular linguistic construction (here, *very New York*)?

In this study, we operationalized individuals' relationships with a particular type of semioreferential object (a city) by using geographic location information found in Twitter users' profiles as well as their work experience history (when available). It should be noted that insider-outsider status was not conceptualized as a sociodemographic variable that divides speakers into two pre-established sociodemographic groups, but is rather a proxy for individuals' experiences, feelings, and knowledge in relation to New York City. Thus, users who had previously worked in NYC (per, for instance, their LinkedIn profile) were classified as insiders, regardless of their current geographic location.

Using multivariate statistics that assessed multiple factors simultaneously, we showed that individuals' relationship with NYC is not only a significant predictor of many linguistic outcomes, but is more frequently significant or near significant than any other predictor considered in this study (i.e. age, sex, ethnicity, education; see Figure 2). Insider-outsider status was a significant predictor of 48% of the types of dependent variables, not least syntactic schemas (e.g. NN compounds), types of referential objects (e.g. animates, concrete things), the use of paradigmatic sets (e.g. positive qualities, negative qualities), and reported speech.

Reported speech is especially noteworthy with regards to the Cx' semiotization, as it was found to be associated with explicit metacommentary regarding users' interpretation of the Cx' expressed meaning (e.g. "*a girl once told me I didn't strike her as very "new york" and I realize now she meant to say "rich"*") or its valence (e.g. "*My boss told me I was 'VERY New York.' I'm assuming this is a good thing.*"). Reported speech instances of the Cx often function as a semioreferential shifter (Courbon 2020 : 50), a linguistic element whose meaning is partially determined by situational parameters that the individual can use to guide their interpretation of its expressed meaning. Semi-overlapping individual knowledge can thus lead to interpretive difficulties or uncertainties, as in: "*My cousin said that it was very "New York" of me to want good posture??*" or *Walked in to work this morning and friend said "you look very New York today" and I had to ask if that was a compliment*" or "*The dean of the art dept. told me I had a very "New York way of doing things" and I still don't know if that's a compliment or an insult*".

Using aggregate data, we showed that New York is associated with a rich array of knowledge and that the representations selected by Twitter users depend partly on their insider-outsider status. A particularly noteworthy contrast is the fact that insiders, for instance, tend to characterize their lived experiences as *very New York* (i.e. Events; NN compounds containing *moment*, *day*, *evening* or *experience*), whereas outsiders frequently comment semiological

objects (e.g. films, television shows, music) or the way New Yorkers speak (NN compounds containing *accent* or *voice*). These combinatorial tendencies hint at the qualitatively different experiences that insiders and outsiders presumably have with the "same" semioreferential object (NYC): insiders who live or have lived in NYC can, unlike most outsiders, draw on specific, temporalized personal experiences. By contrast, typical outsiders –who have no direct relationship with NYC— have access to "second-hand" representations (e.g. films, music, news, narratives) or seek to understand NYC using such representations even when briefly in contact with the semioreferential object (e.g. "*So, you know that thing in New York movies when they cross a street and a man is singing and talking loudly to everyone? It's true! They do! I truly thought that was only in movies. New York is VERY New York.*"). But being an outsider or an insider does not guarantee that one does or does not possess certain types of knowledge associated with *New York*, as knowledge of New York City is experience-derived and therefore only partly shared. Consider, for instance, the case of two young women, who both independently report having been told that they are "*very New York*". The first –an outsider— reports that "*Someone described me as "very New York" *squeals like the Miami girl I actually am**". The second –an insider— expresses confusion as to the Cx' expressed meaning: "*Someone told me recently that I come off " Very New York". Still figuring out what that means exactly...*". It is no surprise, then, that individuals' knowledge can be laid out on a continuum that reflects their very different relationships with the semioreferential object and very different ways of semantizing the Cx, with many outsiders perplexed as to the Cx' expressed meaning and many insiders imbuing the Cx with personal, highly specific and explicit discussions regarding their intended meaning. For instance, in the following, the Twitter user draws on personal experience to define what they mean by *very New York*: "*A very New York date with @LL tonight. And by "very New York" I mean we fitnessed and had a 45 minute subway situation.*". Individuals with deep relationships with the semioreferential object can even specify which version of New York they have in mind, as in the following: "*It's language from Godfather, Sopranos, Goodfellas. It's very New York, ya, but Gambino New York, not Central Park.*". Sharing similar, well-developed mental representations of New York City is, however, no guarantee that individuals will agree that something is "very New York", as is the case of High Line Park, a greenway opened in 2009. A first user comments "*Finally made it to High Line Park after living in Manhattan my entire life. It's nice, but I must say this isn't very "New York."*". Independently, a second user reflects that they believe High Line Park to be "*very New York, at least what new New York can be*".

Limitations of this study include (1) dependence on self-declared geographic information to operationalize the study's key variable (i.e. insider/outsider status), and thus some uncertainty regarding the actual insider-outsider status of the Twitter users; (2) the division of insiders and outsiders into binary categories that abstract away from their life experiences (e.g. outsiders who have relatives in New York or some other non-professional connection with the city that we failed to assess); (3) the subjectivity involved in the data annotation process, and (4) the relatively small size of the dataset (only a few hundred datapoints). Safeguards against these limitations include (1) use of synchronous location information for Corpus1 (i.e. location data gathered during the year of interest); (2) the meticulous parsing of users' profiles and extended social media accounts (e.g. LinkedIn) to determine whether certain outsiders should be reclassified as insiders on the basis of work experiences; (3) independent verification of all

manual coding; and (4) the generation of highly detailed descriptions of each datapoint, thereby maximizing our exploitation of the available data.

## 5. Conclusion

Investigating a construction that is potentially imbued with such variable meaning requires not only considering multiple parameters found in discourse, but also considering corpus contributors not merely as speakers but as individuals with different kinds of relationships with the semioreferential object. Using multivariate statistics to explore fine-grained annotations of a corpus of 1000+ instances of *very New York*, we showed in this exploratory study that users' relationship with the semioreferential object "New York" is associated with and thus may influence their ways of imbuing the construction with meaning. The study's methodology being laid out transparently, it is our hope that it will be taken up and improved upon, so that language users' complex and varied phenomenal experiences might be better captured when investigating semantic phenomena.

Future research avenues include (1) investigating uses of the Cx when it is applied to new insiders' evolving relationship with NYC; (2) characterizing the evaluative language that tends to accompany the Cx (e.g. recognizing the Cx as a compliment); (3) determining to what extent knowledge regarding NYC is shared; and (4) assessing the stereotypicality of the knowledge drawn on by insiders and outsiders when using the Cx. Looking beyond the Cx studied here, it would also be interesting to investigate the representations that insiders living abroad have of their hometown or those developed by outsiders that move to a new town.

**Conflict of interest declaration**

The authors have no real or perceived conflicts of interest to report.

**Funding**

This study was not funded.

**Author contributions according to the CReDIT taxonomy** (Brand *et al*., 2015)

GFM: conceptualization, methodology, software, data collection, data annotation, statistical analysis, validation, writing original draft (all sections), writing: review & editing.

BC: conceptualization, methodology, data annotation, validation, writing original draft (Abstract, Introduction, Discussion, Conclusion), writing: review & editing, project supervision.

**Data availability**

The study's data is available upon reasonable request.

**Acknowledgements**

The authors would like to thank Noémie Germain and Skye Kiernan for their assistance with manual data cleansing and collection.

# Annex

**Table 1. The characteristics of the Twitter users who produced Corpus1's tweets**

| | | Insiders (n=354) | Outsiders (n=304) |
|---|---|---|---|
| Types of Twitter accounts | Personal | 79.20% | 83.88% |
| | Collectively-managed | 5.41% | 3.29% |
| | Private or deleted | 15.39% | 12.83% |
| Current publicly-stated location [1 B] | United-States | **98.80%** | **71.77%** |
| | Canada | **0.30%** | **6.80%** |
| | Australia & New-Zealand | **0.00%** | **3.74%** |
| | United-Kingdom & Ireland | **0.00%** | **12.25%** |
| | Other location | **0.90%** | **9.18%** |
| Sex[1] | Male | **29.81%** | **46.26%** |
| | Female | **49.69%** | **33.67%** |
| | Other or unknown | 20.48% | 20.07% |
| Estimated age[1 B] | 20s | **30.72%** | **18.98%** |
| | 30s | **24.70%** | **33.74%** |
| | 40s | 9.04% | 12.65% |
| | 50+ | 8.43% | 8.13% |
| | No age-related data | **24.40%** | **14.76%** |
| Ethnicity[1] | White | 48.18% | 48.98% |
| | Black or African-American | 13.55% | 14.97% |
| | Hispanic or Latino | 6.33% | 3.75% |
| | East Asian | 3.92% | 2.04% |

| | | | |
|---|---|---|---|
| | South-East Asian | 1.21% | 1.02% |
| | No profile picture or other | 26.81% | 29.20% |
| Education[1] [A] | University | **46.99%** | **32.99%** |
| | Other or unknown | **53.01%** | **67.01%** |
| Employment history[1] | Documented (e.g. LinkedIn) | **60.40%** | **41.17%** |
| | Undocumented | **39.60%** | **58.83%** |

[A] Values in bold differ significantly from each other according to a 95% binomial proportion confidence interval

[B] Raw values (percentages bolded) differ significantly from each other according to a 2-by-k Fisher exact test (α=5%)

[1] Excluding collectively-managed accounts (e.g. businesses)

**Table 2. The characteristics of the Twitter users who produced Corpus2's tweets (2010-2019)**

| | | Insiders (n=297) | Outsiders (n=289) |
|---|---|---|---|
| Types of Twitter accounts | Personal | 86.87% | 91.00% |
| | Collectively-managed | 10.10% | 6.57% |
| | Private or deleted | 3.03% | 2.42% |
| Current publicly-stated location[1] [B] | **United-States** | **91.39%** | **67.04%** |
| | **Canada** | **1.50%** | **5.19%** |
| | **Australia & New-Zealand** | **1.50%** | **5.19%** |
| | **United-Kingdom & Ireland** | **2.25%** | **11.85%** |
| | **Other location** | **3.34%** | **7.61%** |
| Sex[1] | Male | 47.57% | 45.56% |
| | Female | 47.94% | 50.37% |

| | | | |
|---|---|---|---|
| | Unknown | 3.38% | 2.59% |
| Ethnicity[1 B] | White | 63.30% | 62.22% |
| | **Black or African-American** | **15.73%** | **9.26%** |
| | Hispanic or Latino | 5.62% | 7.03% |
| | **East Asian** | **1.87%** | **5.56%** |
| | South-East Asian | 1.12% | 2.59% |
| | Unknown | 11.24% | 11.85% |
| Estimated age[1] | 20s | 32.95% | 31.66% |
| | 30s | 33.33% | 28.57% |
| | 40s | 18.39% | 19.31% |
| | 50+ | 7.28% | 10.81% |
| | No age-related data | 8.05% | 9.65% |
| Education[1] | University | 54.31% | 49.25% |
| | Other or unknown | 45.69% | 50.75% |
| Employment history[1 A] | Documented (e.g. LinkedIn) | **62.55%** | **52.59%** |
| | Undocumented | **37.45%** | **47.41%** |

[A] Values in bold differ significantly from each other according to a 95% binomial proportion confidence interval

[B] Raw values (percentages bolded) differ significantly from each other according to a 2-by-k Fisher exact test ($\alpha$=5%)

[1] Excluding collectively-managed accounts (e.g. businesses)

## Tools and data-related references

Demographia. (2021). *Demographia World Urban Areas (Built up urban areas or world agglomerations)*. http://www.demographia.com/db-worldua.pdf

IBM Corp. (2020). *IBM SPSS Statistics for Windows, Version 27.0*.

Microsoft Corporation. (2018). *Microsoft Excel for Microsoft 365* (Version 2110) [Computer software].

Twitter. (2022). *Developer policy: Off-Twitter matching*. https://developer.twitter.com/en/developer-terms/policy

## Academic references

AARTS, Bas. (2018). *English syntax and argumentation* (5th ed.). Red Globe Press.

ABU-LUGHOD, Janet L. (2000). *New York, Chicago, Los Angeles: America's global cities*. University of Minnesota Press.

AUDRING, Jenny & BOOIJ, Geert. (2016). Cooperation and coercion. *Linguistics* 54 (4): 617–637. https://doi.org/DOI: 10.1515/ling-2016-0012

AUTHIER-REVUZ, Jacqueline. (1984). Hétérogénéité(s) énonciative(s). *Langages* 73: 98–111.

BAUER, Laurie & HUDDLESTON, Rodney. (2002). Lexical word-formation. In R. Huddleston & G. K. Pullum (Eds.). *The Cambridge grammar of the English language*. 1621–1742. Cambridge University Press.

BELTRAMA, Andrea. (2016). *Bridging the gap: Intensifiers between semantic and social meaning* [Doctoral dissertation]. University of Chicago.

BIBER, Douglas & GRAY, Bethany. (2010). Challenging stereotypes about academic writing: Complexity, elaboration, explicitness. *Journal of English for Academic Purposes* 9 (1): 2–20. https://doi.org/10.1016/j.jeap.2010.01.001

BOLINGER, Dwight W. (1972). *Degree words*. Mouton de Gruyter.

BORDET, Lucile. (2017). From vogue words to lexicalized intensifying words: The renewal and recycling of intensifiers in English. A case-study of very, really, so and totally. *Lexis* 10. https://doi.org/DOI: 10.4000/lexis.1125

BRAND, Amy, ALLEN, Lzi, ALTMAN, Micah, HLAVA, Marjorie & SCOTT, Jo. (2015). Beyond authorship: Attribution, contribution, collaboration, and credit. *Learned Publishing* 28 (2): 151–155.

BREBAN, Tine. (2018). Proper names used as modifiers: A comprehensive functional analysis. *English Language and Linguistics* 22 (3): 381–401. https://doi.org/10.1017/S1360674316000514

BREBAN, Tine & DAVIDSE, Kristin. (2016). The history of very: The directionality of functional shift and (inter)subjectification. *English Language and Linguistics* 20 (2): 221–249. https://doi.org/10.1017/s1360674315000428

BUCHSTALLER, Isa & TRAUGOTT, Elizabeth C. (2006). The lady was al demonyak: Historical aspects of adverb all. *English Language and Linguistics* 10 (2): 345–370.

BYLININA, Elizaveta. (2011). This is so NP! *The Baltic International Yearbook of Cognition, Logic and Communication* 6: 1–29. https://doi.org/10.4148/biyclc.v6i0.1571

CLIFFORD, Colin W. G., WATSON, T. L. & WHITE, D. (2018). Two sources of bias explain errors in facial age estimation. *Royal Society Open Science* 5 (10): 180841. https://doi.org/10.1098/rsos.

COURBON, Bruno. (2015). Pratiques sémantiques et différences interindividuelles à l'ère des corpus informatisés. *Cahiers de Lexicologie* 105: 91–126.

COURBON, Bruno. (2018). Du figement attentionnel au figement dans la langue. *Langues et Linguistique* 37: 4 – 47.

COURBON, Bruno. (2020). Impression de langue française en Amérique septentrionale: Point de vue étholinguistique. In R. Mudrochová & B. Courbon (Eds.). *Diversité et variations de la langue française au XXIe siècle*. 12–72. Éditions NAVA.

CURRID-HALKETT, Elizabeth. (2009). *The Warhol Economy: How fashion, art, and music drive New York City*. Princeton University Press.

DUFFLEY, Patrick J. (2020). *Linguistic meaning meets linguistic form*. Oxford University Press.

ELLIS, Edward Robb. (2004). *The epic of New York City: A narrative history*. Basic Books.

ERENBERG, Lewis A. (1984). *Steppin'Out: New York nightlife and the transformation of American culture, 1860-1930*. University of Chicago Press.

FILLMORE, Charles J. (2006). Frame semantics. In D. Geeraerts (Ed.). *Cognitive Linguistics: Basic readings*. 373–400. De Gruyter Mouton.

FRAZER-MCKEE, Gabriel. (2020). *The semantics and pragmatics of proper names in adverbial degree constructions: A corpus-driven contribution* [Master's thesis]. University Laval.

FRAZER-MCKEE, Gabriel & COURBON, Bruno. (2022). Very New York Behavioural Analysis Protocol v2.3. *www.Protocols.Io*: 1–16. https://doi.org/dx.doi.org/10.17504/protocols.io.b5jaq4ie

FRAZER-MCKEE, Gabriel & DUFFLEY, Patrick J. (preprint). The cognitive mechanisms involved in the "DEGREE ADVERB + PROPER NAME" construction: Evaluating proposals from Construction Grammar and Formal Semantics. *SocArcXiv*. https://doi.org/10.31235/osf.io/r9udp

GILBERT, David. & HANCOCK, Claire. (2006). New York City and the transatlantic imagination: French and English tourism and the spectacle of the modern metropolis, 1893-1939. *Journal of Urban History* 33 (1): 77–107. https://doi.org/10.1177/0096144206290385

GONZÁLVEZ-GARCÍA, Francisco. (2014). "That's so a construction!" Some reflections on innovative uses of "so" in Present-day English. In M. de los Ángeles Gómez González, F. J. Ruiz de Mendoza Ibáñez & F. Gonzálvez García (Eds.). *Theory and practice in functional-cognitive space*. 271–293. John Benjamins.

GREENBERG, Miriam. (2003). The limits of branding: The World Trade Center, fiscal crisis and the marketing of recovery. *International Journal of Urban and Regional Research 27* (2): 286–416. https://doi.org/10.1111/1468-2427.t01-1-00454

GREENBERG, Miriam. (2009). *Branding New York: How a city in crisis was sold to the world*. Routledge.

GRIES, Stefan. Th. (2021). *Statistics for Linguistics with R* (3rd ed.). De Gruyter Mouton.

GRIES, Stefan. Th. & DIVJAK, Dagmar. (2009). Behavioral profiles: A corpus-based approach to cognitive semantic analysis. In V. Evans & S. Pourcel (Eds.). *New directions in Cognitive Linguistics*. 57–75. John Benjamins.

GUILLORY, Jamie & HANCOCK, Jeffrey T. (2012). The effect of LinkedIn on deception in resumes. *Cyberpsychology, Behavior, and Social Networking* 15 (3): 135–140. https://doi.org/10.1089/cyber.2011.0389

HOSKINS, Bryony, D'HOMBRES, Béatrice & CAMPBELL, Joann (2006). Does Formal Education Have an Impact on Active Citizenship Behaviour? *European Educational Research Journal* 7 (3): 386–402. https://doi.org/10.2304/eerj.2008.7.3.386

ISRAEL, Michael. (2002). Literally speaking. *Journal of Pragmatics* 34: 423–432. https://doi.org/10.1016/S0378-2166(01)00048-0

KACHRU, Braj B. (1985). Standards, codification, and sociolinguistic realism: The English language in the outer circle. In R. Quirk & H. Widowson (Eds.). *English in the World: Teaching and Learning the language and the literature*. 11–36. Cambridge University Press.

KAPIDZIC, Sanja & HERRING, Susan C. (2014). Race, gender, and self-presentation in teen profile photographs. *New Media & Society* 17 (6): 958–976. https://doi.org/10.1177/1461444813520301

KEELING, David J. (2011). Iconic landscapes: The lyrical links of songs and cities. *Focus on Geography* 54 (4): 113–125. https://doi.org/10.1111/j.1949-8535.2011.00033.x

KING, Kevin. (2016). Intensifiers and image schemas: Schema type determines intensifier type. *Linguistic Society of America* 1: 1–9. https://doi.org/10.3765/plsa.v1i0.3701

LANDAU, Sarah B. & CONDIT, C. W. (1999). *Rise of the New York skyscraper, 1865-1913*. Yale University Press.

LANGACKER, Ronald W. (2008). *Cognitive grammar: A basic introduction*. Oxford University Press.

LORENZ, Gunter. (2002). Really worthwhile or not really significant? A corpus-based approach to the lexicalization and grammaticalisation of intensifiers in Modern English. In I. Wischer & G. Diewald (Eds.). *New reflections on grammaticalization*. 143–161. John Benjamins.

MARTIN, Jeannett R. & WHITE, Peter R. R. (2005). *The language of evaluation*. Palgrave-Macmillan.

MEISSNER, Christian A. & BRIGHAM, John C. (2001). Thirty years of investigating the own-race bias in memory for faces: A meta-analytic review. *Psychology Public Policy and Law* 7 (1): 3–35. https://doi.org/10.1037/1076-8971.7.1.3

MISLOVE, Alan, LEHMANN, Sune, ONNELA, Jukka-Pekka & ROSENQUIST, J. (2011). Understanding the demographics of Twitter users. *Proceedings of the International AAAI Conference on Web and Social Media*: 554–557.

NEWMAN, Michael. (2014). *New York City English*. De Gruyter Mouton.

NICOLAÏ, Robert. (2022). Dynamique sémiotique, ethnométhodologie… et après? Confrontations, synthèses et ouvertures. *Signifying* 5 (1).

OGDEN, David C. & HILT, Michael L. (2003). Collective identity and basketball: An explanation for the decreasing number of African-Americans on America's baseball diamonds. *Journal of Leisure Research 35* (2). https://doi.org/10.1080/00222216.2003.11949991

PARADIS, Carita. (2001). Adjectives and boundedness. *Cognitive Linguistics 12* (1): 47–65. https://doi.org/DOI: 10.1515/cogl.12.1.47

PARADIS, Carita. (2008). Configurations, construals and change: Expressions of DEGREE. *English Language and Linguistics* 12 (2): 317–343. https://doi.org/DOI: 10.1017/S1360674308002645

PHILLIPS, Woomi J. & JANG, Soocheong. (2010). Destination image differences between visitors and non-visitors: A case of New York City. *International Journal of Tourism Research* 12: 642–645.

PLAG, Ingo. (2003). *Word-formation in English*. Cambridge University Press.

RANTISI, Norma. M. (2004). The ascendance of New York fashion. *International Journal of Urban and Regional Research* 28 (1): 86–106.

RHODES, Matthew G. & ANASTASI, Jeffrey S. (2012). The own-age bias in face recognition: A meta-analytic and theoretical review. *Psychological Bulletin* 138 (1): 146–174. https://doi.org/10.1037/a0025750

ROSENTHAL, Caroline. (2016). "A carnival in hell": Representations of New York City in Siri Hustvedt's novels. In J. Hartmann, C. Marks & H. Zapf (Eds.). *Zones of focused ambiguity in Siri Hustvedt's works*. 51–66. De Gruyter.

SADLER, William J. & HASKINS, Ekaterina V. (2005). Metonymy and the metropolis: Television show settings and the image of New York City. *Journal of Communication Inquiry* 29 (3): 195–216. https://doi.org/10.1177/0196859905275971

SANT, Charlotte. (2018). *The degree modification of multidimensional predicates* [Masters thesis]. University of York.

SCHLOSS, Joseph G. (2009). *Foundation: B-boys, B-girls and Hip-Hop culture in New York*. Oxford University Press.

STANGE, Ulrike. (2020). "Holding grudges is so last century": The use of GenX so as a modifier of noun phrases. *Journal of English Linguistics* 48 (2): 107–136. https://doi.org/DOI: 10.1177/0075424220911070

STOFFEL, Cornelis. (1901). *Intensives and downtoners: A study in English adverbs*. Carl Winter.

STRATTON, James. (2018). The use of the adjective intensifier well in British English: A case study of The Inbetweeners. *English Studies* 99 (8): 793–816. https://doi.org/10.1080/0013838X.2018.1519150

SU, Rong, ROUNDS, James & ARMSTRONG, Patrick. (2009). Men and things, women and people: A meta-analysis of sex differences in interests. *Psychological Bulletin* 135 (6): 859–884. https://doi.org/10.1037/a0017364

TABACHNIK, Barbara G. & FIDELL, Linda S. (2019). *Using multivariate statistics*. Pearson.

TROUSDALE, Graeme. (2018). Change in category membership from the perspective of construction grammar: A commentary. In K. Van Goethem, M. Norde, E. Coussé & G.

Vanderbauwhede (Eds.). *Category change from a constructional perspective*. 291–308. John Benjamins.

TUFEKCI, Zeynep. (2014). Big questions for social media big data: Representativeness, validity and other methodological pitfalls. *Proceedings of the Eighth International AAAI Conference on Weblogs and Social Media* 505: 505–514.

VERBRUGGE, Lois. M., GRUBER-BALDINI, Ann L. & FOZARD, James L. (1996). Age Differences and Age Changes in Activities: Baltimore Longitudinal Study of Aging. *The Journals of Gerontology: Series B*, 51B(1), S30–S41. https://doi.org/10.1093/geronb/51B.1.S30

WEE, Lionel & TAN, Ying Y. (2008). That's so last year! Constructions in a socio-cultural context. *Journal of Pragmatics* 40: 2100–2113. https://doi.org/Wee, L. & Tan, Y. Y. (2008). That's so last year! Constructions in a socio-cultural context. Journal of Pragmatics, 40 (12): 2100–2113. doi:10.1016/j.pragma.2008.04.005

ZWICKY, Arnold. (2006). So in style at the NYT. *Language Log*. http://itre.cis.upenn.edu/~myl/languagelog/archives/002994.html