Names | A Journal of Onomastics

A Corpus Linguistic Approach for Onomastics in Drama: Proper Names and Naming in Edward Childs Carpenter's *The Cinderella-Man*

Maya Sfeir
American University of Beirut, LEBANON

Abstract

Proper names and naming are an integral component of dramatic texts. However, to date, scant and fragmented scholarly attention has been given to the connection between proper names and key dramatic elements—namely, stage directions, characterization, and plot progression. This paper adopts corpus linguistics as a method to show how a corpus linguistic approach to onomastics in drama can, on the one hand, provide analysts with a scoping overview of names and naming in plays, and, on the other hand, offer a richer insight into core dramatic elements. Using as a case study Edward Childs Carpenter's play *The Cinderella-Man* (1915), it argues that corpus analysis and visualization techniques such as keyword analysis, frequency analysis, concordance analysis, collocation analysis, and dispersion plot analysis can enrich the study of names and naming in drama. The paper contributes to and extends recent research that has called for the incorporation of corpus linguistic approaches in onomastics research.

Keywords: corpus linguistics, literary onomastics, proper names, family names, anthroponymy, drama

1. Introduction

Proper names are a fundamental feature of drama. Found in both the stage directions and dialogue, they are rife with symbolic, stylistic, and thematic meanings. As Anne Ubersfeld (1999) remarks, in the stage directions, the playwright names the characters and indicates who is to speak; the character's dialogue or discourse is, thus, "marked by its name" (9). In a statement that encapsulates the significance of names in drama, Nina Ekstein (2019) deems that the different categories of names "all work to develop and to populate the non-visible, non-physical, and yet compelling dramatic universe in which each play's action transpires" (159). Commenting on Shakespeare's usage of names in his plays, Tanner (1987) considers that he "uses names to characterize, to reveal cultural attitudes, prejudices, and superstitions, to show conflict or concord, to enhance themes, and to add humorous and serious dimensions to his dramatic narrative" (164), a statement that may be extended to plays in general. Roberts (1999) recognizes the importance of proper names as they direct both audiences and theater practitioners "to the social, emotional and political dynamics of a given scene, as well as to its narrative significance" (1). Proper names are thus the foundational components of dramatic worlds, both on and off stage, and they play a vital role in bringing plays to life.

A number of studies have delved into the significance of proper names in drama, examining a range of plays from different genres and periods—from Renaissance English and French plays, to more modern and translated plays. Analyzing Shakespeare's *Antony and Cleopatra,* Tanner (1987) reveals how proper names accentuate the play's themes, characterization, and style. Drawing on English Renaissance and Shakespearean plays, Roberts (1999) highlights the importance of switching the attention from the name itself to the words that surround and substitute it (collocation or substitution). Examining both collocations (epithets) and substitutes of names, he states that Shakespeare draws on the ambiguity of the meanings of names to add depth and complexity to the thematics of his plays (Roberts 1999). In a study that focuses on seventeenth-century French tragedies, comedies, and tragicomedies, Ekstein (2019) examines the names of characters who never appear on stage, or what she refers to as unembodied names. As she categorizes unembodied names based on time, space, and importance to the play's actions, she reveals how they "create referential texture and density, thereby expanding the time and place of the play's action" (Ekstein 2019, 141). For Ekstein (2019), unembodied names have a key role in plays; they are "a powerful and compact source of reference, bridging the visible on stage with the invisible realms of the elsewhere and the past" (142). Exploring French seventeenth-century drama, McElveen (1994) considers that in Jean Racine's *Andromache,* the naming of absent characters allows Racine to expand the theatrical unities of time and place. Focusing on translations of names in Ibsen's plays, Maylath (1996) argues that unsuccessful translations of names can lead English audiences to miss some aspects of characterization. Other studies investigated proper names in American plays, probing the playfulness of proper names in Elmer Rice's *The Adding Machine* (Brown 1996), and categorizing proper names and exploring their function in Sam Shepard's drama (Rubenstein 1989).

Names in plays have thus been the subject of analysis from different perspectives. However, despite the centrality of proper names in plays, their exploration in dramatic texts has remained fragmented, and a systematic approach is yet to be developed for their analysis. The connection between proper names and key dramatic elements, namely characterization, plot progression, and stage directions, remains underexplored. This paper thus suggests that computational text analysis, in particular corpus linguistic approaches, can offer a systematic means for exploring and analyzing names and naming in drama. Writing in 1999, Roberts discerns how computational text analysis could enhance the study of names in drama, contending that "Computer-based

studies of play-texts, from spelling patterns to stage directions, have sharpened our understanding of the material distinctiveness of authorial voices", thus enhancing our understanding of how names and naming function in drama (6). Drawing inspiration from Robert's (1999) statement, this paper argues that a corpus linguistic approach to proper names in drama—in particular, an approach that builds on frequency, keyword, dispersion plot, concordance, and collocation analyses—can play a pivotal role in advancing the study of the onomastics of drama. As a case study, it examines Edward Childs Carpenter's play *The Cinderella-Man: A Comedy in Four Acts* (1915).

In this paper, I adopt a specific understanding of the term "proper name" that differs from the definition of "proper noun", though the two terms are often used interchangeably (Schlücker & Ackermann 2017; Van Langendonck 2007; Coates 2006; Allerton 1987). Proper nouns, a distinct lexical category, are used as names for objects, persons, or places; they are capitalized, non-complex, and non-compound words (Coates 2006) (e.g., *Tony*, *Marjorie*).[1] Unlike proper nouns that are non-compound words, proper names vary in terms of their structural complexity, and might consist of proper nouns, a combination of proper nouns, a combination of common and proper nouns, among other configurations (Motschenbacher 2020) (e.g., *Tony*, *Anthony Quintard*, *Cinderella Man*). The next section of this paper discusses the relevance of corpus linguistic analysis in onomastics research, and it introduces key techniques in corpus linguistics and comments on how they can contribute to the analysis of names and naming in plays. Section 3 presents the methodology, along with a short synopsis of the play. Section 4 presents the results and their discussion; the last section is the conclusion of the paper.

2. Towards a Corpus Linguistic Approach to Onomastics in Drama

The seeds for using computational text analysis in onomastics research have long been planted, as suggested by Roberts's (1999) research. However, it is only in recent years that the proponents for integrating corpus linguistic approaches in onomastics research have become more vocal (see Motschenbacher 2020). Motschenbacher (2020) highlights the advantages of applying corpus linguistic approaches to onomastics, considering that corpus linguistic techniques—mainly frequency analysis, concordance analysis, collocation analysis, and keyword analysis—both complement and substantiate onomastics research. Motschenbacher (2020) further argues that using corpus linguistics can provide the "usage context" in which a name occurs (100). In this paper, I engage with Motschenbacher's (2020) claim, and in the process, expand its scope, as I show how the application of corpus linguistic approaches to onomastics in drama can provide analysts with a scoping overview of names and naming in plays, on the one hand, and is integral to the analysis of key dramatic elements: characterization, plot, and stage directions, on the other.

Corpus linguistic approaches can serve as a valuable complement to human analysis. Outlining the capabilities of computers—and by extension corpus tools—Susan Hockey (2000) points out that "Computers are good at mechanical processes, such as searching, counting, and sorting into alphabetical or numerical order [. . .] perform[ing] these tasks not only much faster than a human being, but also very much more accurately" (3–4). The most commonly used corpus techniques—frequency, concordance, collocations, and keyword analyses—are performed using corpus tools specifically designed to count, sort, and group words, including proper names. When applied to play texts, these techniques can greatly enhance the understanding of names and naming in drama.

Frequency analysis, the most basic corpus analysis technique, relies on finding and sorting the frequency of a certain search word or group of words in a text or a number of texts (Anthony 2022; Jones 2022). Given that a substantial number of onomastics studies have focused on finding and categorizing names in drama (see Ekstein 2019; Roberts 1999; Rubenstein 1989), frequency analysis can refine the analysis of names in plays in two main ways: it can help distinguish hapaxes from more prevalent names and their substitutes, and it can enable researchers to compare the occurrences of names in play texts and their translations, thus potentially allowing critics to decide which names should be prioritized for analysis. It is important to mention that most corpus tools, such as AntConc (Anthony 2023) for instance, support case-sensitive searches. The feature is particularly relevant for the analysis of proper names in plays written in languages that support capitalization, as character names are frequently capitalized in the speech prefixes to distinguish them from the rest of the text.

Another form of frequency analysis is keyword analysis. Keywords can be broadly defined as words whose occurrence in a particular text or corpus is significantly higher when compared to another reference corpus. Scott (2010), and Scott and Tribble (2006) consider that a keyword would fall into one of three categories: words relating to the text's "aboutness" and its major themes, markers of style that include high-frequency or grammatical words, and proper nouns and/or titles. While, in general, the analysis of keywords serves an exploratory or confirmatory purpose, a number of studies have suggested a possible connection between proper name keywords and characterization in fiction and drama (Murphy 2015; Mahlberg & McIntyre 2011; McIntyre & Walker 2010). Murphy (2015) examines keywords in Shakespeare's soliloquies and argues that certain proper name keywords underline the "identity, office and relevance of certain male characters" (350). Culpeper (2009) contends that the keyword *Romeo* in Shakespeare's *Romeo and Juliet* suggests that Romeo is the "fulcrum of the play" (38). Other corpus-based studies have also suggested a connection between proper noun keywords and plot progression in short stories (Toolan 2009).

Apart from calculating frequencies of words and names, corpus tools can also show where specific words or names occur in a text—through the use of a dispersion plot viewer—or in their immediate context or co-text—through the use of a concordance (Anthony 2022; Mahlberg & Wiegand 2018; Katz 1996). Dispersion plots, as Mike Scott (2001) describes them, are "maps showing where in the texts the search words were found" (48). By integrating dispersion plots into his analysis of the translation of Clarice Lispector's novel *A hora da estrela* (The Hour of the Star), Scott (2001) shows that characters are referred to using pronouns long before they are named, halfway through the novel.

Another technique for viewing words in a text is a concordance. Described by Mahlberg and Wiegand (2018) as "fundamentally a form of linguistic visualisation" (127), a concordance consists of a listing of words in their context of occurrence (Weisser 2015). As Mahlberg and Wiegand (2018) point out, concordances enable researchers "to see patterns and identify meanings associated with these patterns" (127). Concordances thus aid researchers in analyzing the surrounding context in which a certain name occurs in the dialogue or stage directions. Closely associated with concordance is collocation, or the examination of co-occurring lexical items within a limited distance from each other (four to five words to the left or right) (Weisser 2015). While some collocates can be identified manually, through a concordance analysis, others especially in the analysis of groups of texts, can be extracted statistically. Collocation, like concordance, can reveal which lexical items are associated with names. Thus, the aforementioned corpus techniques can help researchers examine the salience of proper names in drama and visualize these names in their relevant texts(s) or co-texts. I build on these corpus techniques to explore how proper names occupy the world depicted by the play or performance, and how they contribute to characterization and plot progression through a targeted analysis of proper names in Edward Childs Carpenter's play *The Cinderella-Man* (1915).

At this point, I want to draw attention to the multi-layered and intricate nature of dramatic discourse. Apart from having two main constituents, dialogue and stage directions, play texts—which are meant to be read, produced, and performed—also have multiple addressees that include lay readers and theatergoers, as well as theater practitioners, such as stage directors, stage producers, and actors (Feng & Shen 2001; Short 1998). As a result, in a play, names and naming will resonate differently depending on the addressees, and depending on whether a name occurs in the stage directions or dialogue, and whether it occupies the dramatic world during the play's performance or whether it is simply meant to be read. Theatrical audiences, for instance, will only have access to names occurring in the dialogue during a play's performance, while actors will be able to access names in both the dialogue and stage directions; hence, theatrical audiences and actors will thus interpret names and naming differently. Where possible, this paper attempts to take into consideration the different diegetic and extradiegetic functions of names, although it acknowledges its limitations in this regard.

3. Methods

This paper adopts a corpus linguistic approach to explore names and naming in drama, and to examine how personal names are linked to core dramatic components, using as a case study Edward Childs Carpenter's play *The Cinderella-Man* (1915). *The Cinderella-Man* is a comedy that captures the budding romance between Marjorie Caner, an heiress, and Tony Quintard, a penniless artist whom Marjorie nicknames the Cinderella Man. Marjorie first hides her identity from Tony when she visits him in his attic to discuss his art. However, when her identity and immense wealth are finally revealed, Tony is offended. The play ends with their reconciliation, and her father's blessings. *The Cinderella-Man* (35,686 tokens and 3,796 types) is in the public domain and is available electronically through Project Gutenberg and HathiTrust. As the play's own title suggests, proper names occupy a central role in the play.

The analysis began with the extraction of proper names in the play. For this purpose, a systematic approach was adopted, combining keyword analysis and close reading. The dramatis personae of *The Cinderella-Man* are as follows: MARJORIE CANER; MORRIS T. CANER, her father; D. ROMNEY EVANS, a lawyer;

ALBERT SEWALL, a composer; DR. JOSEPH THAYER; ANTHONY QUINTARD; WALTER NICOLLS; BLODGETT, Butler at Caner's; CELESTE, Marjorie's French Maid; JERRY PRIMROSE; and THE GREAT SHE-BEAR!. For this analysis, the proper names were first identified in the dramatis personae, and then extracted through keyword analysis by comparing the play against a literary reference corpus from the period.[2] See Table 1 below:

Table 1: Texts in the Reference Corpus by Author, Title, and Publication Year

Dramas and Plays
Joaquin Miller, *49: Forty-Nine* (1882)
H. B. Eneleh, *Tempest Tossed* (1885)
Augustus Thomas, *Alabama* (1898)
Steele MacKaye, *Hazel Kirke* (1899)
Bret Harte and Thomas Edgar Pemberton, *Sue* (1902)
William Churchill De Mille and Cecil Blount DeMille, *The Genius* (1904)
Jack London, *Scorn of Women* (1906)
Joaquin Miller, *An Oregon Idyl* (1910)
Charles Rann Kennedy, *The Idol-Breaker* (1914)
Clyde Fitch, *The Climbers* (1915)
Roi Cooper Megrue, Unfer Cover (1918)
Comedies and Farces
Augustin Daly, *Our English Friend* (1884)
Madeleine Lucette Ryley, *Christopher Junior* (1889)
Rachel E. Baker, *The Chaperon* (1891)
Rachel E. Baker, *A King's Daughter* (1893)
Brander Matthews, *The Decision of the Court* (1893)
William Gillette, *All the Comforts of Home* (1897)
Verna M. Raynor, *The Bird Family and Their Friends* (1898)
Elizabeth Adshead Hyde, *An Engaged Girl* (1899)
Sarah Folsom Enebuske, *A Detective in Petticoats* (1900)
Clyde Fitch, *The Frisky Mrs. Johnson* (1908)
George Howells Broadhurst, *What Happened to Jones* (1910)
David Belasco, *May Blossom* (1911)
Comedy Dramas and Melodramas
Effie Woodward Merriman, *Diamonds and Hearts* (1897)
Edward Sheldon, *The High Road* (1912)

The extraction of proper names was complemented by a close analysis of the play itself. After establishing a list of character names, their frequencies were generated by using each character's name as a search word. The results are shown in Table 2:

Table 2: Frequency Distribution of the Range of the Proper Names for the Three Main Characters: Tony, Marjorie, and Caner, Per Act

Character	Names	Act I	Act II	Act III	Act IV
Tony	Tony	5 Stage Directions (0)	220 Stage Directions (216), of which Speech Prefixes (174)	225 Stage Directions (211), of which Speech Prefixes (160)	69 Stage Directions (65), of which Speech Prefixes (43)
	Mr. Quintard	0	6 Stage Directions (0)	8 Stage Directions (0)	8 Stage Directions (0)
	Quintard	3 Stage Directions (0)	1 Stage Directions (0)	0	8 Stage Directions (0)
	Cinderella Man	3 Stage Directions (0)	3 Stage Directions (0)	3 Stage Directions (0)	1 Stage Directions (0)
	Anthony Quintard	0	2 Stage Directions (1)	0	1 Stage Directions (0)
	Mr. Anthony Quintard (Non-Keyword)	0	0	0	1 Stage Directions (0)
	Mr. Cinderella Man	0	0	0	1 Stage Directions (0)
Marjorie	Marjorie	206 Stage Directions (183), of which Speech Prefixes (150)	81 Stage Directions (81), of which Speech Prefixes (72)	131 Stage Directions (131), of which Speech Prefixes (96)	154 Stage Directions (141), of which Speech Prefixes (105)
	Veiled Princess	0	11 Stage Directions (0)	0	0 Stage Directions (0)
	Princess	0	5 Stage Directions (0)	6 All in Dialogue *References Are to the Princess in Tony's Story* (4)	8 (Excluding One General Reference to a Princess) *References Are to Marjorie* Stage Directions (1)
	Miss Mudge	0	0	3 Stage Directions (0)	3 *Of which One Placed in Quotes* Stage Directions (0)
	Miss Marjorie	0	0	0	3 Stage Directions (0)
	Marjorie Caner (Non-Keyword)	1 Stage Directions (1)		0	0
	Miss Marjorie Caner (Non-Keyword)	1 Stage Directions (0)	0	0	0
Caner	Caner	84 Stage Directions (84), of which Speech Prefixes (63)	1 Stage Directions (1)	0	107 Stage Directions (104), of which Speech Prefixes (74)
	Morris	2 Stage Directions (0)	0	0	3 Stage Directions (0)
	Morris T. Caner (Non-Keyword)	2 Stage Directions (2)	0	0	0
	Morris Caner (Non-Keyword)	0	0	0	2 Stage Directions (0)

The analysis also took into consideration whether character names are mentioned in the dialogue or stage directions. A dispersion plot was generated for main characters' names occurring as speech prefixes. The analysis was complemented by a collocation analysis and a concordance analysis of selected collocates.

Three different corpus tools were used for the analysis: AntConc version 3.5.9 (Anthony 2020), WordSmith version 8 (Scott 2021), and LancsBox version 6.0 (Brezina et al. 2020). Each of these tools has its distinct strengths, and they complement each other in various aspects. AntConc was used for keyword, frequency, and concordance analysis. The log-likelihood test was selected as an indicator of keyness. Keyness, as defined by McIntyre and Walker (2019), is "a measure of the extent to which a linguistic item is over- or under-represented in comparison with its distribution in a comparator corpus" (23). It is often calculated using a statistical test

that determines the significance of the variation in in the frequency of the linguistic item between the target corpus that the researcher is analyzing and another reference corpus (McIntyre & Walker 2019). In this study, all keywords, including proper name keywords, have a keyness above 15.13, meaning we are 99.99% confident they are statistically significant. The effect size was measured using Hardie's 2014 Log Ratio (LR) statistic; the minimum log ratio was set to 2, indicating that the keyword is likely to occur at least four times in *The Cinder-ella-Man* compared to the reference corpus. The frequencies of selected names were also calculated per act to examine their possible correlation with plot progression.

Dispersion plots were created using WordSmith. The tool has the option of dividing the text into eight segments of equal lengths. Given that the play text is segmented based on its length, the different segments do not align with the acts of the play. One of the advantages of WordSmith is that it allows analysts to compare different search words throughout the text, at the same time, as these can be visualized on designated lines. LancsBox was selected for collocation analysis because it integrates the GraphColl function, which displays collocations in the form of graphs and networks. The Mutual Information (MI) statistic was used and set to a minimum value of 5. The minimum frequency per collocate was also set to 5. Collocates were calculated using a span of 5L/R, that is five words to the left and right.

4. Results and Discussion

This section discusses the results of the keyword, frequency, dispersion plot, collocation, and concordance analysis of selected proper names in *The Cinderella-Man*.

4.1. *Keyword Analysis*

My analysis begins with a comparison and categorization of proper name keywords and character names as listed in the dramatis personae. Keyword analysis, often used as an exploratory or confirmatory technique, has a practical relevance for the mechanical extraction of proper names and their subsequent categorization. Proper names and titles are, after all, one category of keywords described by Scott and Tribble (2006). Because a keyword list is to some extent a frequency list, it can show how frequently a proper name occurs in the play. The more frequent a proper name is, the more prominent a character is likely to be. The frequency of names becomes most insightful when it is compared against the order in which characters' names are listed in the dramatis personae. As Gary Taylor (2007) points out, a character list "touches on issues—of representation, gender, class, national and racial identity" (31).

Analyzing and categorizing the names in the dramatis personae and keyword list led to a number of insights. Proper names in the dramatis personae can be categorized based on a number of variables: their formality and informality, the mention of occupations or not, as well as gender. Following Leuwen's (1996, 2008) model for the categorization of social actors, I divided names in the dramatic personae into formal (family name only), informal (first name), and semi-formal (first and last name) nomination. Results showed that one of the characters, the maid, was referred to informally (*Celeste*) and the six main characters—including Tony, the eponymous *Cinderella Man*—semi-formally (*Marjorie Caner, Morris T. Caner, Albert Sewall, Anthony Quintard, Walter Nicolls, Jerry Primrose*). In terms of titulation, two names had a title (*Dr. Joseph Thayer* and *D. Romney Evans*). One name was metaphorical (*The Great She-Bear*, which refers to Anthony Quintard's landlady). One name, *Blodgett*, was difficult to categorize, as it could be both a first name or a family name. In Smith's (1988) *New Dictionary of American Family Names*, the name is listed as a family name of Welsh origin. Interestingly, a number of the characters' first and family names are of Welsh (*Evans, Nicolls*) or French (*Quintard, Marjorie, Celeste*) origins (Smith 1988; Dunkling & Gosling 1985). While gender as a category does not seem to influence the order of characters in the dramatis personae, there seems to be a connection between the naming of characters and their social class or occupation, as Celeste, Blodgett, and Jerry Primrose, members of the domestic staff, are mentioned towards the end of the list of characters. Including The Great She-Bear at the end of the list is possibly not arbitrary and might be connected to her antagonistic role in the play, as well as her negative perception by readers and audiences.

While each character in the dramatis personae is identified by only one name, the keyword list offers a scoping overview of the different ways in which characters are named and the frequency of each name. To explore each character's range of names, I decided to compare names in the character list to those in the keyword list. A number of observations can be made concerning the top 20 proper name keywords in the play (see Table 3). A quick glance at the top 20 proper names reveals the range of terms used to name characters: Tony is referred to, not only as *Anthony Quintard,* his name in the character list, but also as *Tony, Quintard, Mr. Quintard,* and *Cinderella Man*. Marjorie Caner's names in the top 20 proper name keywords are *Marjorie,*

Princess, Veiled Princess, and *Miss Mudge.* Characters' key proper names also seem to vary in formality. Four names are informal proper names (*Tony/Tony's, Marjorie, Romney/Romney's, Walter*), and seven are formal family names (*Blodgett, Caner, Primrose, Sewall, Thayer, Quintard, Nicolls*). Two key proper names are titulated (*Mr. Quintard, Miss Mudge*), and five are metaphorical names (*She Bear, Princess, Veiled Princess, Cinderella Man, Spongey*). All formal key proper names seem to refer to male characters. In terms of frequency, *Tony* and *Marjorie*, the play's protagonists are mentioned most frequently, followed by *Romney, Caner,* and *Primrose. Romney* and *Caner* are also the second and third names in the dramatis personae, right after Marjorie's. The fact that they occur before Tony's name on the character list seems to emphasize their centrality to the play's actions as Caner is Marjorie's biological father from whom she has been estranged, and Romney is her surrogate father.

Table 3: Top 20 Proper Name Keywords in *The Cinderella-Man*

Rank	Keywords	Frequency	Keyness
1	tony	506	2,865.05
2	marjorie	561	2,732.49
3	romney	282	1,596.73
4	caner	185	1,047.50
5	primrose	181	994.56
6	walter	108	591.76
7	blodgett	103	583.20
8	sewall	88	498.27
9	thayer	53	300.09
10	she_bear	49	277.45
11	princess	24	117.42
12	mr_quintard	17	96.26
13	tony's	13	73.61
14	quintard	12	67.95
15	veiled_princess	10	56.62
16	cinderella_man	10	56.62
17	romney's	9	50.96
18	spongey	9	50.96
19	miss_mudge	6	33.97
20	nicolls	6	33.97

4.2. Frequency Analysis

Following the keyword analysis, which provides analysts with a bird's eye view of the range of names in the play, I performed a frequency analysis of the range of names of the three main characters in the play—Marjorie, Tony, and Caner—per act. My analysis took into consideration whether these names occurred in the dialogue or stage directions to explore the breadth of names permeating both the dramatic world and the world of the readers. Table 2 presents a breakdown of the frequency of mentions per act of the names of the different selected characters; it includes both proper name keywords and non-keyword names. A close examination of the frequency breakdown of these names shows that they lend themselves to a deeper analysis.

While some names only occur in the dialogue, permeating the dramatic world, others only occur in the stage directions, permeating the world of the readers. Some names, for instance, only appear in the stage directions of some acts: *Morris T. Caner* (freq=2) (Act I) and *Marjorie Caner* (freq=1) (Act I). These names do not make their way to the dramatic world of the play; they are only accessible to the readers, including both lay readers and theater practitioners such as actors and stage directors. Other names only occupy the dramatic world of the play during performance—and some only for a few moments in each act. In this respect, Ekstein (2019) highlights the importance of the hapax which she describes as "the most problematic name in that it is articulated once and disappears" (Ekstein 2019, 143). She adds "It may be unreasonable, unless the name is exceedingly well known, to expect that the theatrical audience will even process it" (Ekstein 2019, 143). The two names *Mr. Anthony Quintard* and *Mr. Cinderella Man* occur once in Act IV of the play. Their occurrence is significant: *Mr. Anthony Quintard* is used to announce Anthony's entrance in the play's final act and, in that same act, *Mr. Cinderella Man* is used by Marjorie to address Anthony after her real identity is revealed. While these names are not likely to be remembered by the audiences, they do highlight a shift in plot, mostly manifested in the shift of relationship between the characters.

Names only mentioned in the dialogue could be divided into two categories: terms of address or terms of reference; in turn, terms of reference could be divided into self-references or terms referring to other characters. For instance, *Tony*, which mostly occurs as a speech prefix, is only mentioned in the dialogue 27 times, 14 times in references or self-references, and 13 times as Tony is being addressed. *Tony* is thus used as a term of reference and as a term of address more or less proportionally. A particularly interesting name is the eponymous *Cinderella Man*. *Cinderella Man* (freq=10) is mentioned three times per act in the dialogue of Acts I through III, and only once in Act IV. The name is indicative of the increasing affection between Tony and Marjorie. While in Act I, Marjorie uses the term to refer to Tony when talking to other characters, in Act II, she reveals to Tony that she calls him the Cinderella Man, a term that he begins to use to refer to himself. In the last act, *Cinderella Man* is preceded by the possessive first-person pronoun *my,* as Marjorie uses it to addresses Tony, thus highlighting the bond that has developed between them. As such, frequency analysis can be used to explore ways in which characters populate both the dramatic world and the world of the readers, and how frequently names are uttered in a performed play. These frequencies are not without a connection to the plot progression of the play.

4.3 Dispersion Plot Analysis

The frequency analysis was followed by a dispersion plot analysis of the proper names used as speech prefixes for Marjorie, Tony, and Caner. Figure 1 shows the occurrences of the speech prefixes of the three main characters throughout the play. As previously mentioned, speech prefixes consist of a proper name set apart from the rest of the stage directions and dialogue through capitalization or a punctuation mark. They are also indicative of the turns the characters have, defined in this context as their opportunity to contribute to dialogue. A close analysis of the dispersion plot in figure 1 reveals the sections in which characters' turns intersect. It is important to emphasize that characters' turns are not directly indicative of their stage presence, as some of the characters present on stage might remain silent. A number of interesting insights emerge from examining the dispersion plot in figure 1.
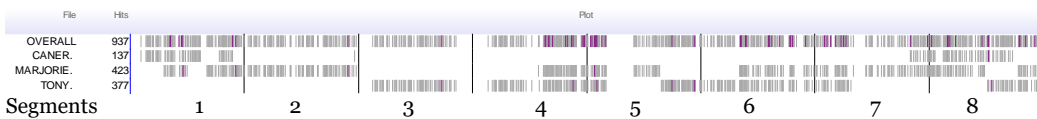
Figure 1: Dispersion Plot of The Proper Names Used as Speech Prefixes for Tony, Marjorie, and Caner

The first is that Marjorie and her father only interact together at the beginning and end of the play (see in figure 1 Segments 1 and 7–8, equivalent to Acts 1 and 4). Their interaction reveals a shift in their relationship throughout the play. The example below shows how, at the onset of the play, Caner who has invited Marjorie to live with him after a long period of estrangement, questions her with irritation, after she had moved some furniture around the house:

> CANER. (Tartly to MARJORIE) Who do you think you are, anyway?
> MARJORIE. I naturally supposed, Pa-pa, when you invited me to live with you, that I would be mistress of the house!

Caner's attitude towards Marjorie takes on a more loving quality in the play's final act, as he becomes more concerned with her happiness, as the example below shows:

> CANER. I'm going to make sure of your happiness. (MARJORIE brightens up. He goes on with a sudden turn to strength) But you can make up your mind to one thing — I shan't disinherit you.

The dispersion plot in figure 1 also shows that Marjorie and Tony do not interact with each other prior to the play's midpoint (Segments 4–5, equivalent to the end of Act II). Their encounter thus seems to be framed as the play's climax. A third insight is that Tony and Caner only interact together at the end of the play (see Segment 8, equivalent to Act 4). The meeting between the two results in a shared admiration and mutual respect between them, thus paving the way for the play's resolution, a predictable outcome considering its genre as a comedy. Caner's role as a father and his reconciliation with Marjorie seems to be pivotal in advancing the play's plot. As such, analyzing the dispersion plot of the three main characters' proper names in the speech prefixes can provide subtle insights into the play's plot progression, which seems to be linked to the play's genre as a comedy.

4.4 Collocation and Concordance Analyses

To further my analysis, I used collocation analysis and examined the collocates of the proper names used as speech prefixes of the three main characters using the GraphColl tool in LancsBox. Figure 2 shows the collocates for the selected proper names. Due to limitations of space, I decided to focus on the most salient collocations, the adverbs *impatiently, excitedly,* and *angrily* that reveal how the actors are expected to depict the characters on stage

Figure 2: Collocations of the Proper Names Used as Speech Prefixes for Tony, Marjorie, and Caner. Corpus: Cinderella Man| Search Term: CANER| Statistic: 03 - MI| Span: 5-5| Collocation Freq. Threshold: 5.0| Statistic Value Threshold: 5.0| CPN: 03 - MI (5.0)/ L5-R5/ C: 5.0-NC: 5.0|

A concordance analysis of these collocates shows that most of them occur in the stage directions, directly following the character's name. Of the eight occurrences of *excitedly*, five describe Marjorie (see figure 3). Similarly, of the 14 occurrences of *impatiently*, eight are associated with Tony, and of the 15 occurrences of *angrily*, 5 are associated with Caner (see figures 4 and 5). As such collocates of proper names, when combined with concordance analysis, can provide a scoping overview of characterization, with possible indications of how the play could be performed.

```
                     to a sandwich. Knock at door below.) TONY. (   Excitedly) Come up! Come up! (He goes to trap,
                   t funny little head of yours? MARJORIE. (Happily,   excitedly, coming back to piano) I've thought of
                 a glimpse of MARJORIE at the window.) WALTER. (   Excitedly, gaily) Ahhhhh! There's a skirt at your
              son him. (Bangs manuscript with fist.) MARJORIE. (   Excitedly) He will — he must! ROMNEY. You'll give
            own off R. MARJORIE rushes to window.) MARJORIE. (   Excitedly) It's Romney's car! Yes, yes — Tony'
                   back to China — he knows, the old dog! (   Excitedly, prancing about.) And what did he tell me? —
              inherit you. (Bell buzzes in hallway.) MARJORIE. (   Excitedly) There he is! (Rushes to window R. ) CANER. (
              've got to do something about this. MARJORIE. (   Excitedly) Yes — yes. SEWALL. Quintard must restore the la
```

Figure 3: Concordance of *Excitedly* in *The Cinderella-Man*

```
              tly) Nothing's important — except my work! (More   impatiently) Confound you — I had just got a wireless
                   hort. WALTER. Are you still writing music? TONY. (   Impatiently) I never wrote music. WALTER. Oh, yes you
                      below L. ) It's most important, sir. TONY. (   Impatiently) Nothing's important — except my work! (More imp
                       of a girl that — that I am. TONY. (   Impatiently) Oh — you're worth a million princesses. MARJORI
                          — is it? (He drops down on trunk.) TONY. (   Impatiently, pacing stage up and down R.) No — no!
              sneaks across the roof to your window? TONY. (   Impatiently, pulling him away from window) Yes! Yes! Come
             TER. Is — is that necessary? MARJORIE. (A little   impatiently) Surely it is! WALTER. (Surprised) You want me
                . Ah, then I don't blame you. TONY. (   Impatiently) There's nothing to blame anybody about. I'
                      a spoon onto the floor.)  TONY. (Looking up   impatiently) Upon my word! How can I work when
                 down to chair at table, back of davenport,   impatiently) Well, well  —  what did you say to her?
         ARJORIE. One moment, Papa. (To SEWALL, anxiously,   impatiently) Well, well! What do you think of it?
             uty! (THAYER comes down to fireplace L. ) CANER. (   Impatiently) What did she say? ROMNEY. Nothing much! But
              man of real talents! MARJORIE. A genius! CANER. (   Impatiently) What's his name? ROMNEY. Quintard. MARJORIE. Ant
                 'll be — be very gentle with him —   CANER. (   Impatiently) Yes, yes, yes! (BLODGETT crosses hallway from L.
```

Figure 4: Concordance of *Impatiently* in *The Cinderella-Man*

```
                      everything. Tony won't have me — rich. CANER. (   Angrily) But what right has this young snip to
                        work. MARJORIE. I'm glad of it. TONY. (   Angrily) Don't you realize that I've got
            o his window. We — we became acquainted. CANER. (   Angrily) Good Lord! ROMNEY. Take it calmly, Morris! CANER
                   ! (To CANER) I made him do it! CANER. (   Angrily) He had no business to mind a chit
                     — don't go all over that again. WALTER. (   Angrily) I can't for the life of me
                     stuff yourself, it isn't my fault. TONY. (   Angrily) I don't stuff myself. It's you —
             giant Nubian slaves. She says to them — very   angrily — "Kill that man!" — and runs up to the
                        be done —  I'll do it. (To PRIMROSE   angrily) Leave me catch you up here again, you
          WALTER. (Twirling box in his hand, ignoring hers   angrily) Marjorie, you'll be sorry for treating me
            vement — don't you, Pa-pa? CANER. (Emphatically,   angrily) No! I do not! (To BLODGETT) Blodgett, what
              y brains have utterly refused to work. MARJORIE. (   Angrily, stamping her foot) It isn't true! They'
                 the tiniest bit of love to me! CANER. (   Angrily) Why didn't he? Doesn't he like
                 've got to get this thing done? MARJORIE. (   Angrily) You'd finish it much better if you
           h her father. TONY. (Starting, growling — almost   angrily) You're not the rich girl next door?
                   's waitin' to see Mr_Quintard. She_Bear. (   Angrily) You're not waiting to see Mr_Quintard.
```

Figure 5: Concordance of *Angrily* in *The Cinderella-Man*

Conclusion

This study set out to examine how a corpus linguistic approach to onomastics in drama can provide a systematic means for exploring and analyzing names and naming in drama. By using as a case study Edward Childs Carpenter's play *The Cinderella-Man*, the study integrates keyword, frequency, concordance, collocation, and dispersion plot analyses to explore proper names in the play and their connection to dramatic elements. The analysis has shown how keyword and frequency analyses can provide a structured approach to extracting and categorizing names, and can reveal the range of names used to refer to characters in the play. A frequency analysis, as the study has also shown, can reveal the number of occurrences of names in dialogues and stage directions across each of the play's acts, thus providing insights into names that are accessible to both readers and audiences, and names that are exclusively accessible to readers. The study also used a dispersion plot to visualize proper names used as speech prefixes for the three main characters in the play, Marjorie, Tony, and Caner. The dispersion plot has revealed moments during which the three characters interact, and has highlighted how their relationship throughout the play evolves, thus providing subtle insights into the play's plot progression, which seems to be linked to the play's genre as a comedy. Overall, this article has found that corpus linguistics as a method can provide a replicable and novel approach to reading and analyzing names in plays, and it can substantiate the analysis of proper names by showing how they are connected to key dramatic elements.

The present study has its limitations. First, the study only focused on selected proper names in one play. The examination of a corpus that includes more plays from different genres will further enrich the analysis of proper names in drama and reveal how corpus linguistics can contribute to onomastics in drama. Second, due to limitations of space and time, the study could not explore in depth some details pertaining to the corpus analyses. In addition to expanding the corpus of analyzed plays, subsequent research could delve more extensively into corpus analyses of proper names, further describing how names are connected to dramatic elements. Further studies could also provide annotation systems for categorizing names in plays, taking into consideration whether these names occur in the dialogue or stage directions, as well as their diegetic and extradiegetic functions. In the future, it will be important to explore how the integration of corpus linguistic methods might result in a meaningful and explanatory typology of dramatic literature that extends beyond mere generic classification. Despite its exploratory nature, this study offers some insight into how a corpus linguistic approach to onomastics in drama can enrich the reading of names in plays and provide an increased understanding of how proper names contribute to the construction of the dramatic and theatrical universe.

Notes

1. It is important to bear in mind that capitalizing proper nouns and names is supported by certain languages that use the Latin script, such as English and French; however, other languages such as Chinese and Arabic do not have capital letters.
2. For a complete keyword list, kindly contact the author.

Acknowledgments

Disclosure Statement

No potential conflict of interest was disclosed by the author.

References

Allerton, David John. 1987. "The Linguistic and Sociolinguistic Status of Proper Names: What Are They, and Who Do They Belong To?" *Journal of Pragmatics* 11, no. 1: 61–92.

Anthony, Laurence. 2020. *AntConc* Version 3.5.9. Mac. Tokyo, Japan: Waseda University. Accessed December 3, 2023. https://www.laurenceanthony.net/software

Anthony, Laurence. 2022. "What can Corpus Software Do?" *The Routledge Handbook of Corpus Linguistics*. Edited by Anne O'Keeffe and Michael J. McCarthy. Oxon: Routledge, 103–125.

Anthony, Laurence. 2023. *AntConc* Version 4.2.4. Mac. Tokyo, Japan: Waseda University. Accessed October 1, 2023. https://www.laurenceanthony.net/software

Brown, Russel E. 1986. "Names and Numbers in *The Adding Machine*". *Names* 34, no. 3: 266–274.

Brezina, Vaclav, Pierre Weill-Tessier, and Anthony McEnery. 2020. *#LancsBox* Version 6.0. Mac. Lancaster University, England. Accessed September 25, 2023. http://corpora.lancs.ac.uk/lancsbox/download.php

Carpenter, Edward Childs. 1915. *The Cinderella-Man: A Comedy in Four Acts*. New York: Tower Brothers Stationery Company.

Coates, Richard. 2006. "Properhood". *Language* 82, no. 2: 356–382.

Culpeper, Jonathan. 2009. "Keyness: Words, Parts-of-Speech and Semantic Categories in the Character-Talk of Shakespeare's *Romeo and Juliet*. *International Journal of Corpus Linguistics* 14, no. 1: 29–59.

Dunkling, Leslie, and William Gosling. 1985. *The New American Dictionary of First Names*. New York: Signet.

Ekstein, Nina. 2019. "Populating the Dramatic Universe through Names". *Papers on French Seventeenth Century Literature* 46, no. 90: 133–164.

Feng, Zongxin, and Dan Shen. 2001. "The Play off the Stage: The Writer-Reader Relationship in Drama. *Language and Literature* 10, no.: 79-93.

Hardie, Andrew. 2014. "Log Ratio: An Informal Introduction". CASS: ESRC Centre for Corpus Approaches to Social Science. Accessed April 23, 2023. http://cass.lancs.ac.uk/?p=1133

Hockey, Susan M. 2000. *Electronic Texts in the Humanities*. Oxford: Oxford University Press.

Jones, Christian. 2022. "What Are the Basics of Analysing a Corpus?" *The Routledge Handbook of Corpus Linguistics*. Edited by Anne O'Keeffe and Michael J. McCarthy. Oxon: Routledge, 126–139.

Katz, Slava. 1996. "Distribution of Common Words and Phrases in Text and Language Modelling". *Natural Language Engineering* 2, no.1: 15–59.

Leeuwen, Theo van. 1996. "The Representation of Social Actors". *Texts and Practices: Readings in Critical Discourse Analysis*. Edited by Carmen Rosa Caldas-Coulthard and Malcolm Coulthard. London: Routledge, 32–70.

Leeuwen, Theo van. 2008. *Discourse and Practice: New Tools for Critical Discourse Analysis*. New York: Oxford University Press.

Mahlberg, Michaela, and Dan McIntyre. 2011. "A Case for Corpus Stylistics: Ian Fleming's *Casino Royale*". *English Text Construction* 4, no. 2: 204–227.

Mahlberg, Michaela, and Viola Wiegand. 2018. "Corpus Stylistics, Norms and Comparisons: Studying Speech in *Great Expectations*". *Rethinking Language, Text and Context: Interdisciplinary Research in Stylistics in Honour of Michael Toolan*. Edited by Ruth Page, Beatrix Busse, and Nina Nørgaard. New York: Routledge, 123–143.

Maylath, Bruce. 1996. "The Trouble with Ibsen's Names". *Names* 44, no. 1: 41–58.

McElveen, Nancy M. 1994. "Name, Memory, and Time in Racine's Trojan War Plays". *Names* 42, no. 1: 27–37.

McIntyre, Dan, and Brian Walker. 2010. How Can Corpora be Used to Explore the Language of Poetry and Drama?" *The Routledge Handbook of Corpus Linguistics*. Edited by Anne O'Keeffe and Michael McCarthy. London: Routledge, 516–530.

McIntyre, Dan, and Brian Walker. 2019. *Corpus Stylistics: Theory and Practice*. Edinburgh: Edinburgh University Press.

Motschenbacher, Heiko. 2020. "Corpus Linguistic Onomastics: A Plea for a Corpus-Based Investigation of Names". *Names* 68, no. 2: 88–103.

Murphy, Sean. 2015. "I Will Proclaim Myself What I Am: Corpus Stylistics and the Language of Shakespeare's Soliloquies". *Language and Literature* 24, no. 4: 338–354.

Roberts, David. 1999. "Shakespeare's Fellows: Names and Collocations". *Cahiers Élisabéthains* 56, no. 1: 1–12.

Rubenstein, Gilbert M. 1989. "The Onomastic Sam Shepard". *Names* 37, no. 3: 231–244.

Schlücker, Barbara, and Tanja Ackermann. 2017. "The Morphosyntax of Proper Names: An Overview. *Folia Linguistica* 51, no. 2: 309–339.

Scott, Mike. 2001. "Comparing Corpora and Identifying Key Words, Collocations, Frequency Distributions through the *WordSmith Tools* Suite of Computer Programs". *Small Corpus Studies and ELT: Theory and Practice*. Edited by Mohsen Ghadessy, Alex Henry, and Robert L. Roseberry. Amsterdam: John Benjamins, 47–77.

Scott, Mike. 2010. "Definition of Key-ness". Lexically. Accessed November 11, 2023. https://lexically.net/downloads/version5/HTML/index.html?keyness_definition.htm

Scott, Mike. 2021. *WordSmith Tools* Version 8. Windows. Stroud, England: Lexical Analysis Software. Accessed September 13, 2023. https://lexically.net

Scott, Mike, and Christopher Tribble. 2006. *Textual Patterns: Key Words and Corpus Analysis in Language Education*. Amsterdam: John Benjamins.

Short, Mick. 1998. "From Dramatic Text to Dramatic Performance." *Exploring the Language of Drama: From Text to Context*. Edited by Jonathan Culpeper, Mick Short, and Peter Verdonk. London: Routledge, 6–18.

Smith, Elsdon C. 1988. *New Dictionary Of American Family Names*. New York: Gramercy Publishing.

Tanner, Jeri. 1987. "The Power of Names in Shakespeare's *Antony and Cleopatra*". *Names* 35, no. 34: 164–174.

Taylor, Gary. 2007. "The Order of Persons". *Thomas Middleton and Early Modern Textual Culture: A Companion to the Collected Works*. Edited by Gary Taylor and John Lavagnino. Oxford: Clarendon Press, 31–79.

Toolan, Michael. 2009. *Narrative Progression in the Short Story: A Corpus Stylistic Approach*. Amsterdam: John Benjamins.

Ubersfeld, Anne. 1999. *Reading Theatre*. Translated by Frank Collins. Toronto: University of Toronto Press.

Van Langendonck, Willy. 2007. *Theory and Typology of Proper Names*. Berlin: Mouton de Gruyter.

Weisser, Martin. 2015. *Practical Corpus Linguistics: An Introduction to Corpus-Based Language Analysis*. Oxford: Wiley.

Notes on the Contributor

Maya Sfeir is a lecturer in the Department of English at the American University of Beirut, Lebanon. Her research interests revolve around using corpus and digital methods to examine patterns and meanings in texts. She has taught courses in the areas of Corpus Linguistics, Digital Humanities, and Discourse Analysis.

Correspondence to: Maya Sfeir, Department of English, American University of Beirut, Lebanon
Email: ms213@aub.edu.lb