

# Adjective-Noun Placement Variation in French Twitter Corpora

Joan Palmiter Bajorek

University of California, Davis  
One Shields Ave, Davis, CA 95616  
jpbajorek@ucdavis.edu

## Abstract

French written and spoken corpora reveal that variation of adjective-noun placement occurs more often within written corpora than in spoken ones (Lightfoot, 1979; Prévost, 2009; Thuilier, 2013). I examine adjective lemma variation of placement, adjective categories, frequency, and syllable length as a means of assessing current Twitter trends of adjective placement. In this preliminary study, a cross-section of 30 adjective lemmas within a corpus of approximately 6,000 French tweets demonstrates that parallels exist between the usage of adjectives on Twitter and norms of spoken French.

## 1 Introduction

This article examines French adjective-noun placement in the determiner phrase within written, spoken, and Twitter corpora. Placement of French adjectives in the determiner phrase has long been controversial and misconstrued within the literature (Benzitoun, 2014; Grevisse & Goosse, 2011). It is generally accepted that “[i]n Romance languages such as French and Spanish, postnominal adjectives are the rule rather than the exception” (Alexiadou, Haegeman, & Stavrou, 2007).

However, contemporary written and spoken corpora indicate that adjective placement has greater flexibility and ambiguity than has previously been understood. Statistical analysis of corpora indicate that adjective placement is more restrictive in spoken usage than written usage by a large margin (Thuilier, 2013). This research investigates adjective placement in

French tweets through the examination of two French Twitter timelines of users *Le Monde* and *Cyprien*. It is hypothesized that French tweets will display higher levels of adjective placement restriction than Thuilier’s written and spoken corpora findings. This may be attributed to Twitter’s character limit and the curt, spoken-like treatment of language on the social network and overall online usage of language.

## 2 French Adjectives

In French, attributive adjectives “adjectifs épithètes” can be placed before or after the noun as seen in (1) (Grevisse & Goosse, 2011; Laenzlinger, 2005), an example modified from (Thuilier, 2013).

- (1) a. un beau requin                      b. un requin sympa  
          *a beautiful shark*                      *a shark nice*

In example (1a), the adjective “*beau*,” found in the prenominal position and modifies the noun “*requin*.” Conversely, in (1b), the adjective “*sympa*” is postnominal (Alexiadou et al., 2007; Delbecque, 1990; Thuilier, 2013). French adjectives are widely believed to exist in one of three categories: fixed prenominal, fixed postnominal, or accepted alternator, with recognized semantic shift attributed to either position.

Although these categories are frequently used, “[i]t has proven notoriously difficult to define the functions of the two positions” (Sleeman, Van de Velde, & Perridon, 2014) and variation is oft attributed to semantic shifts (Alexiadou et al., 2007; Laenzlinger, 2005). For example, the adjective “*brave*” means good or decent in the preposition, but courageous or brave in the postposition. In this way, positions are correlated with meaning. Yet, Thuilier, a native speaker of

French, provides several examples of exceptions of words that are considered as regular alternators with specified meanings correlated to positions (2013). Furthermore, many adjectives retain the same meaning regardless of placement, such as the adjective “*charmant*,” meaning charming (Thuilier, 2013). In many ways, this variation in placement thus is a choice of the user and varies widely on the usage context. It should be noted that the primary concern of this research is variation in placement and patterns of adjective corpora rather than semantic interpretations, which are outside the scope of this study.

## 2.1 Syntactic Theory

In considering the syntactic basis of adjective placement, modern frameworks explain the variation through Branching Direction Theory (Song, 2012). It is posited that nouns move leftward and up the syntactic tree with “pied-piping” and “snowballing” movement and interact with linear and mirror image matching among languages (Cinque, 1994, 2010; Laenzlinger, 2005; Stavrou, 2012). However, much is left to be desired in these theoretical understandings of syntactic underpinnings. If adjectives that alternate in position and have semantic shifts relegated to those positions, why does the noun moving up and down the tree account for the adjective’s semantic change (Alexiadou et al., 2007)? How does this theory account for languages that display varied adjective-noun ordering systems? In the words of Alexiadou et al., “[a] question that remains open is why various languages have to resort to different means (order, excessive articles, morphology) in order to encode different interpretations of adjective-noun combinations” (parentheses original (2007)). Do to the recent growth in Optimality Theory (Song, 2012), there is hope that elegant, more fully explanatory theories can be developed.

## 2.2 Historical Background

French evolved from Latin, a language with no fixed adjective placement (Sleeman et al., 2014). Between the 13th and 19th Century, adjective-noun placement shifted from prenominal preference in Old French to

postnominal dominance in Modern French (Boucher, 2006; Sleeman et al., 2014).

Several theories postulate reasons for this phenomenon. According to Glatigny, “preposed adjectives belong in very great majority to the ancient foundations of the language” (translated by the author (Thuilier, 2013)). The shift may resulted from the influence and exposure of French and Romance languages to Germanic languages (Grevisse & Goosse, 2011). Yet some reject these “influenced by another language” interpretations due to the lack of documented evidence (Lightfoot, 1979). Another theory indicates that prenominal adjectives have marked positions that licenses specialized reading and greater frequency (Ledgeway, 2012). Evidence for this theory comes from the stronger syntactic bond of prenominal adjectives to nouns as compared to the relationship of their postnominal counterparts and nouns. The phonological connections between prenominal adjectives and nouns are demonstrated through the use of liaisons and irregular adjective agreements (Ledgeway, 2012).

## 3 Usage Norms in Corpora

For decades, newspapers, literary texts, university essays, and spoken corpora have been analyzed for their adjective placement norms. While patterns arise, gaps in information and uniformity of analysis abound in the research. Most studies cite raw counts of adjectives found to be prenominal or postnominal. They do not take into account adjective lemma repetition, aspects of frequency, and the possibility of alternation of position. Recognizing the importance of the source of corpora, Delbecq writes that prenominal and postnominal proportions of placement are “mainly to be attributed to the text genre” (1990). Echoing these sentiments, Thuilier writes that intense difference exist according to the type of production (2013).

### 3.1 Literary Texts

In Wydler’s study of the famous text “*La Chanson de Roland*,” 70% of overall adjectives were prenominal (Gerard J. Brault, 1978; Lightfoot,

1979). “The Song of Roland” is the written version of the epic tale that was once performed orally, though its precise origins are clouded in “considerable speculation” (G.J. Brault, 2010). Delbecque notes that modern French novels demonstrate the same proportional split in placement, but provides no data to support the claim (1990).

In another analysis of French adjectives, Wagner separated adjectives into three basic categories. “*Cardinals*” are basic, essential, and high frequency adjectives throughout the French language. “*Populaires*” are mainstream, colloquial, and wide spoken adjectives. Finally, “*savants*” are “learned” adjectives of scholarly registers (translations by the author (Lightfoot, 1979)). For clarity in the following section, these three categories will be referred to as Frequent, Colloquial, and Scholarly. In an examination of 13<sup>th</sup> and 14<sup>th</sup> century prose, Wagner found that Frequent adjectives strongly favored prenominal positions: 2,393 prenominal and 11 postnominal occurrences, a 99.5% to 0.5% split (Lightfoot, 1979). 75% of Colloquial adjectives were prenominal whereas roughly 70% of Scholarly adjectives were postnominal.

When the Frequent and Colloquial categories were combined, the average prenominal to postnominal division was 95% to 5%. This combined category was of extremely high frequency, almost 16 times as common as the Scholarly category (Lightfoot, 1979). It is important to note that all of the data literature cited in this section explores the raw counts of the data and categorization of lemmas, but does not specify the frequency or other attributes of the adjectives. Wagner’s study underlines the importance of adjective categorization when considering adjective placement. In addition, Lightfoot notes that the difference between source adjectives and those that are “fairly recent” borrowings reinforcing Glatigny’s theories of ancient French versus modern linguistic acquisitions (1979).

### 3.2 University Essays

In 1980, Wilmet investigated 90 philology essays of university students dating from 1977 to 1978 (Delbecque, 1990). Of the 3,835 adjectives found, Wilmet reported that 2.3% were exclusively prenominal, 16.9% alternated in position, and

80.8% were exclusively postnominal (1980). Wilmet did not account for categories of adjectives, but what can be interpreted from this study is the propensity for the work of university students, who focus their studies on literary criticism, history, and linguistics, to use literary, worked, polished language that displays postnominal preferences, an idea supported by Thuilier’s work (2013).

### 3.3 Newspapers

Postnominal preference is also found in most of the studies in newspaper corpora. Five studies of French adjective placement in newspapers published between 1911 and 1978 by Damourette and Pichon, Hug, and Forsgren returned homogenous results of 65% postnominal placement. With an average of 3,000 adjectives per study, all of this research compared raw counts of adjective-noun placement (Delbecque, 1990).

### 3.4 Spoken and Modern Corpora

Contemporary analyses provide more thorough and nuanced approaches to corpora studies as seen through the work of Thuilier (2013). For reduced redundancies, section 2.8, refers to Thuilier’s 2013 work. In 2013, Thuilier compiled a corpus of over 1.3 million French words from written and spoken French. The written corpus was composed of *Le Monde* newspaper articles dating from 1989-1993 sourced from the French Treebank (FTB). For the spoken corpus, Thuilier used the 2005 edition of the CORAL Romance corpus of transcribed speech. Thuilier’s data reveals placement norms between the French written and spoken usages.

Spoken data displayed placement that was more rigid and more likely to follow prescriptivist norms than written language. In explanation, Thuilier hypothesizes that written language is “*travaillée*,” worked and polished, and therefore may have greater opportunity for flexibility, nuance, and variations of style (translation by the author). Conversely, it was hypothesized that spoken language is spontaneous and instinctive tendency follows mainstream norms.

Thuilier’s corpus comprised of 1,750 adjective lemmas of which 170 were found to alternate in position. Oral corpora lemmas demonstrated a prenominal preference of 74% while their written

corpora counterparts demonstrated a prenominal preference of 67%.

Additionally, Thuilier coined a term about the syllable length patterning demonstrated by the data, “*court avant long*,” short before long (translated by the author). Prenominal adjectives were frequently monosyllabic and “adjectives of one syllable are more than 80% in anteposition” (translated by the author).

Summary of Previous Studies			
By Raw Count of Adjectives			
Study Authors	Corpus Origin	Prenominal	Postnominal
Wydler (1956)	Song of Roland	70%	
Damourette and Pichon (1911-1968)	Newspaper		61%
Hug (1971)	Newspaper		63%
Hug (1971)	Newspaper		69%
Forsgren (1978)	Newspaper		67%
Wilmet (1981)	Modern Novels		67%
Corpus Defined by Categorization of Adjective Lemmas			
Study Authors	Corpus Origin	Prenominal	Postnominal
Wagner (Reiner, 1968)	13th & 14th Century Prose, Frequent	99.5%	
Wagner (Reiner, 1968)	13th & 14th Century Prose, Scholarly		70%
Wagner (Reiner, 1968)	13th & 14th Century Prose, Frequent & Colloquial	95%	
Corpus Comprised of Only Alternating Lemmas			
Study Authors	Corpus Origin	Prenominal	Postnominal
Thuilier (2013)	Newspaper, French Treebank (FTB)		67%
Thuilier (2013)	Oral, CORAL Romance corpus		74%

Table 1: Summary of Literature Review Percentages

From these studies and especially that of Thuilier’s work, further experiments can be designed to investigate several factors in adjective-noun placement in French corpora. To summarize the above sections succinctly the percentages quoted in this section are summarized in Table 1. The following factors have been previously demonstrated to be significant in classifying French adjective placement: lemmas as a better gauge of corpora rather than raw counts of adjectives, variation of placement, categorization of adjectives, frequency, origin of corpora, and syllable length.

### 3.5 Twitter

Twitter is a real-time social networking and microblogging service (Lomicka & Lord, 2012). Users can read, post, and repost messages called “tweets.” Tweets are limited to 140 characters that

are retained on a user’s profile page, called a “timeline” (Lomicka & Lord, 2012).

A global phenomenon, Twitter has 288 million users as of 2015 (Popper, 2014; Statistica, 2015). While the company is American and based in San Francisco, 77% of Twitter accounts are outside of the United States and over 35 Twitter company offices are outside of the United States, including one in Paris (Twitter, 2015a). Despite its American origins, which might imply high volumes of English language tweets, only 34% of tweets are in English (Smith, 2015). Twitter users from France tripled from 2009 to 2013 (Statistica, 2013).

## 4 Defining the Corpus

Approximately 6,000 tweets were analyzed from the Twitter feeds of the French newspaper *Le Monde* (@lemondefr) and the French comedian Cyprien (@MonsieurDream). These users were chosen because of 1) their comparable four million Twitter followers (“Twitter statistics for France,” 2015), 2) *Le Monde*’s connection to Thuilier’s (2013) and Wilmet’s (1980) corpora studies, and 3) the users have the 2nd and 3rd largest audiences of Twitter in France. The top 24 French Twitter accounts are attributed to men or groups (Socialbakers, 2015) thus the Twitter timelines analyzed in this study are a sample of the most popular, mainstream content on French Twitter.

### 4.1 Le Monde

Launched in March 2009, *Le Monde*’s Twitter timeline feed of tweets reports events related to France, the world, and politics. Written by journalists Pauline Croquet and Clément Martel, the timeline has over 132,000 tweets and over 6,000 photos and videos (Twitter, 2015c). As the online presence of one of the largest French newspapers, the tweets from this source are objective, political, newsworthy, and diplomatic in nature.

### 4.2 Cyprien

While Cyprien has significantly fewer tweets than the *Le Monde* timeline, Cyprien’s popularity was determined to be a more important factor than number of tweets. Launched in June 2007,

MonsieurDream has over 10,000 tweets and is written by Cyprien Iov who goes by Cyprien (Twitter, 2015d). Cyprien is a prolific 24-year-old French blogger, illustrator, comedian, and poster of videos on YouTube.com (Iov, 2015a, 2015b). His work ranges from the banal to the political, blogging about the experience of moving apartments, gay marriage, and general self-promotion. Cyprien has more Twitter followers than the Twitter accounts of movie celebrities, for example Gad Elmaleh with 300,000 fewer followers (Socialbakers, 2015). As a young comedian, Cyprien's humor is colloquial, sensationalized, dramatic, and irreverent in nature. His most popular YouTube video is a rap song where Cyprien responds a man who criticized his jokes (Cyprien, 2011). As of May 2016, the video has almost 37 million views and includes jokes about sex, online reputation, clothing style, and poor education, concluding with a joke about killing handicapped people (Cyprien, 2011).

By juxtaposing the linguistic trends from these two users, adjective placement norms on Twitter can be examined. It was hypothesized that French tweets would display higher levels of adjective placement restriction than Thuilier's corpora findings (2013). It was expected that Cyprien's usage of language would demonstrate even greater restriction in adjective placement than *Le Monde's* tweets do to the higher register and scholastic nature of the latter's content.

## 5 Methodology

The research looked into the placement norms of *Le Monde* and Cyprien on Twitter to determine whether corpora more closely resembled written or spoken corpora norms. In addition, questions of distribution, alternation, syllable length and frequency were considered in the data collection and analysis.

Tweets were downloaded from twitter using Python scripts and the Python module "Tweepy" (Roesslein, 2009). To access the tweets, a Twitter account was created with a corresponding Twitter website giving the user access to personalized Application Program Interface (API) credentials. Modified from scripts posted on GitHub (Yanofsky, 2013), roughly 6,000 tweets were downloaded from Twitter users *Le Monde* and *MonsieurDream*. Due to restrictions on tweet

access to the general public, roughly 3,200 tweets were downloadable per Twitter timeline. Each user has a homepage, a unique page of their tweets and retweets called a "timeline," synonymous with the term "Twitter feed" (Twitter, 2015b). The scripts created csv files of the tweet's date published and text content. Due to the prolific tweeting of *Le Monde*, all tweets analyzed were published between April 1<sup>st</sup>, 2015 and May 4<sup>th</sup>, 2015. For the Cyprien corpus, the tweets from his timeline were published between December 9<sup>th</sup>, 2009 and May 2<sup>nd</sup>, 2015. Hereafter for reasons of clarity, the corpora will be termed "*Le Monde*" and "Cyprien" and will no longer be italicized.

For the scope of this study, in the Twitter corpus 30 adjective lemmas were investigated, see Appendix. These lemmas were chosen based on their prevalence in the linguistic literature and were placed into categories: High Frequency, Colors, Contemporary/Neologisms, Famous Alternators, and Others, see Appendix A.

Tweets containing the specific adjective lemmas were appended to separate csv files to isolate all instances of the adjective. Due to the added complexity of multiple adjectives, see mirror imaging (Cinque, 1994, 2010), only tweets with a single adjective were analyzed. Tallies were created of how many instances were found for each adjective lemma within the two Twitter feeds.

These tallies were then sorted to categorize adjectives found in a determiner phrase. From those that fit the correct syntactic environment, determiner phrase, they were annotated as prenominal and postnominal. Tallies were noted for all the above values. In Python, a Chi square test was conducted on the raw counts of the 30 adjectives found in the determiner phrase of either timeline were statistically significant.

## 6 Findings

6,406 tweets were data-mined from the Cyprien and *Le Monde* Twitter feeds. After isolating the 30 adjective lemmas, 1,181 adjective occurrences were analyzed. Roughly 82% of these single adjectives were found in a determiner phrase or noun phrase, a total of 949 adjective occurrences. Some adjective lemmas were not attested within the corpora. In the analysis of the data, variation in placement, categorization of adjectives, adjective frequency, syllable length and origin of the corpora

were analyzed. The following analysis considers the Le Monde and Cyprien datasets separately as well as combined, to consider the implications of Twitter as a usage context.

It is crucial to note that despite the 6,406 tweets data-mined, the resulting adjectives found in determiner phrases was lower than initially anticipated. It is therefore recognized the findings from this corpus must be considered a preliminary study that has recommendations for further investigations.

### 6.1 Placement Variation

Of adjectives found within the corpora, 8% of adjective lemmas in Cyprien tweets and 30% adjective lemmas in Le Monde’s tweets varied in placement, Figure 1.

Of the lemmas which alternated in this corpus, they demonstrated prenominal preference at an average of 80%/20% prenominal/postnominal split. The distribution of this variation ranged from 60% to 99.67% prenominal placement. These finding are in line with Thuilier’s finding that alternators are more inclined to be prenominal placed (2013).

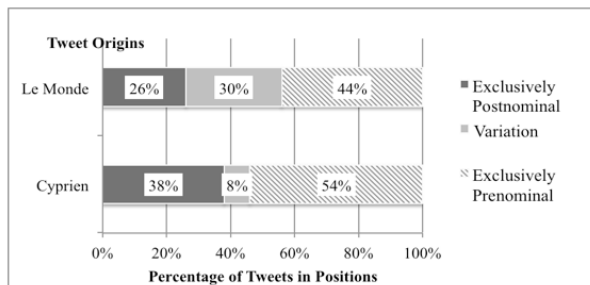


Figure 1: Raw Counts and Percentages of Tweets

Trends of the placement of alternators differed between the Cyprien and Le Monde corpora. Le Monde’s adjectives displayed more variation in placement whereas very few of Cyprien’s adjectives varied in placement. Supporting the initial hypothesis, adjective placement is more rigid in Cyprien’s tweets than in Le Monde’s tweets. The margin of difference in usage norms is a significant finding of this study and would be a strong starting point, to analyze Twitter feeds of news sources as compared to the tweets of individuals, for future investigations.

### 6.2 Adjective Categories

In consideration of the initial categories of the 30 adjective lemmas, adjectives were analyzed within their categorizations as a comparison point to Wagner’s categories of Frequency, Colloquial, and Scholarly, see Table 2 (Lightfoot, 1979). Of the Famous Alternators, most of these were prenominal placed, though many of this category were adjective lemmas that were unattested in the corpus. Contemporary/Neologisms and Others displayed the greatest amount of variation in placement within the corpus. Mirroring Wagner’s results, High Frequency adjectives were placed prenominal for the most part, on average 92% of the adjective lemmas. Colors were exclusive found in postnominal positions. This complements Wilmet’s findings which demonstrated that “*blanc, bleu, noir et rouge*” were found postnominal 97.4% of the time (1980; Thuilier, 2013). Thus, category of adjective was helpful for grouping some adjectives, but not for all the categories of the study.

Combined Cyprien and Le Monde Data Sets		
Categorization	Average Total Prenominal	Average Total Postnominal
High Frequency	92%	8%
Colors	0%	100%
Contemporary/Neologisms	33%	67%
Famous Alternators	95%	5%
Others	46%	54%

Table 2: Combined Cyprien & Le Monde Data Sets

### 6.3 Frequency

Adjectives used with highest frequency are correlated with those that are prenominal placed. Of the top 10 adjective lemmas in the Cyprien tweets, all were prenominal dominant at an average of 97.79% overall. Of the top 10 adjective lemmas in the Le Monde tweets, 70% were prenominal dominant, see Table 3. These high frequency adjectives in the Le Monde tweets were exclusively postnominal, “politique,” “blanc,” and “rouge.” As seen also with the above section,

adjectives describing colors are famously postnominally placed (Grevisse & Goosse, 2011). Of the highest frequency adjectives from both timelines, 6 out of 10 overlap between the Cyprien and Le Monde tweets.

Frequency: Top 10 Lemmas	Cyprien	Prenominal	Le Monde	Prenominal
	Lemmas	Percentage	Lemmas	Percentage
1	nouveau	99.7%	nouveau	98.7%
2	petit	100.0%	grand	100.0%
3	premier	100.0%	premier	95.5%
4	grand	100.0%	politique	0.0%
5	bon	100.0%	ancien	100.0%
6	dernier	100.0%	petit	100.0%
7	gros	100.0%	dernier	100.0%
8	prochain	78.3%	blanc	0.0%
9	beau	100.0%	rouge	0.0%
10	mauvais	100.0%	bon	100.0%

Table 3: Combined Cyprien & Le Monde Data Sets

In the Cyprien corpus, there was a significant divide between the high frequency adjectives prenominal dominance to low frequency postnominal dominance. The Le Monde corpus is more nuanced in this separation of variables. In the Le Monde corpus, 7 of the 10 high frequency adjectives displayed prenominal dominance. For the low frequency adjectives, 10 of 13 also displayed prenominal preference, with only 3 that were exclusively postnominal in placement. Overall, the Le Monde corpus has a larger number of adjectives that alternate in placement than those that are exclusively postnominally placed, which leads to lower levels of stratification in frequency and placement. When the corpora were combined, the mixed nature of the Le Monde corpus was demonstrated. Frequency was strongly correlated with prenominal adjective placement in the Cyprien corpus where it was highly stratified, but this trend was less significant in the Le Monde corpus.

## 6.4 Syllable Length

The results of this study suggest that syllable count was not a significant factor in adjective placement. Previous literature indicated that “short before long” patterns were observed in larger data sets (Thuilier, 2013), however these trends were not observed in these data, see Appendix B. The average percentage for the average prenominal placement for lemmas with one syllable was 60%, two syllables at 82%, and three syllables at 61%.

While two syllable adjectives display prenominal preference, this is not a major finding of this study.

## 6.5 Twitter Corpora Overall

As seen in the variation of placement, Le Monde’s adjective lemmas demonstrated greater flexibility of placement than Cyprien’s, which supports the hypothesis. Le Monde’s flexibility was more in line with written corpora norms as seen in Thuilier, while Cyprien’s follows more spoken norms of restriction (2013).

As compared to the corpora in Table 1, section 2.8, this Twitter corpora demonstrated significant stratification at a 76%/31% prenominal split. From a raw count of the 30 adjectives found in either position within the corpora, Cyprien’s adjective placement was more stratified than Le Monde’s placement, Table 4. When combining the counts in both Twitter feeds, there was a strong tendency toward prenominal placement: 504 prenominal to 100 postnominal, 5:1 a ratio.

Raw Count of Adjectives in Determiner Phrase				
Source of Tweets	Prenominal	Percentage of Prenominal	Postnominal	Percentage of Postnominal
Cyprien	303	92%	26	8%
Le Monde	201	73%	74	27%
Combined Data Set	504	83%	100	17%

Table 4: Raw Counts and Percentages of Tweets

It should be noted that the lemma “nouveau” was removed from the raw count comparisons due to its nature as an outlier. While the lemma posed no problem as an example of categories, high frequency, and syllable length, its high raw count skews the data set. Within the Cyprien data set, “nouveau” was found 301 instances in the determiner phrase compared to all the other adjectives with values ranging from 0 to 59 occurrences. Due to the limited scope of this article “nouveau” was not analyzed independently, a case for future study.

## 7 Conclusions

Twitter corpora displayed adjective placement norms that were more closely aligned with spoken corpora norms than written corpora norms when compared to the patterns explored in Thuilier (2013). As hypothesized, the Cyprien corpus was

more restricted in adjective placement alternation than the Le Monde corpus. In the Le Monde tweets, 30% of the adjectives were found in both prenominal and postnominal positions, much higher than the 8% alternating in the Cyprien corpus. This suggests the greater flexibility of adjective placement in the Le Monde corpus. Syllable count was not a significant factor in adjective placement, though there was a tendency for adjectives of two syllables to be prenominal place. Frequency was a significant factor in adjective placement and was highly correlated with prenominal placement in the Cyprien corpus.

Limitations of this study include small final numbers of adjectives attested per lemma. Despite the approximately 6,000-tweet sample size, some adjectives occurred less than 5 times. In the future, larger data sets would be preferred. In addition, the 30 adjectives chosen may have influenced the data collected and in the future, different sampling methods could remedy this issue. Overall, these findings provide preliminary insight into the French adjective placement norms on Twitter.

## 7.1 Further Research

**Sampling:** Using a parts of speech tagger, a larger analysis of all the adjectives in the corpora could be conducted. Tweets could be data mined over a specific time interval rather than from user timelines. In addition, this study showed the longitudinal trends of specific users. Further research could also examine cross-sectional data of streams of tweets and isolate specific collocations or adjective usage. Collocations of adjective-noun pairings and a study of ngrams of the corpus could illuminate stylistic preferences of Twitter users. Gender and other demographics data could be compiled to generate studies that compare different types of Twitter users.

**Nouveau:** Further study might investigate specific adjective lemmas, such as nouveau, for a microanalysis interpretation of the data.

**Dialects and Languages:** While it is clear that French is spoken around the world, varieties of the language based on location were not chosen for the this study (Thuillier, 2013), but could be analyzed to compare language norms of this corpus. Semantic shifts in placement were not analyzed within this study, but could be an aspect of future work. French is not the only language that displays

adjective-noun prenominal and postnominal alternation; Spanish does as well (Alexiadou et al., 2007; Delbecque, 1990). An analysis of the adjectives in Spanish tweets, or other languages with comparable alternation trends could be compared to this study. This study provides a jumping-board for further adjective-noun Twitter corpora studies.

## Appendix A

### Full Combined Data Sets for Prenominal and Postnominal Placement of Adjectives found in Determiner Phrases

Combined Cyprien and Le Monde Data Sets						
Categorization	Total Adjectives in Determiner Phrases			Percentage Placement		
	Prenominal Placement	Postnominal Placement	Total Adjectives	Total Prenominal	Total Postnominal	
<b>High Frequency</b>						
1	beau	18	0	18	100%	0%
2	bon	41	0	41	100%	0%
3	dernier	46	0	46	100%	0%
4	grand	85	0	85	100%	0%
5	gros	23	3	26	88%	12%
6	long	3	3	6	50%	50%
7	mauvais	16	0	16	100%	0%
8	nouveau	374	2	376	99%	1%
9	petit	81	0	81	100%	0%
10	premier	87	2	89	98%	2%
11	prochain	24	8	32	75%	25%
<b>Colors</b>						
12	rouge	0	12	12	0%	100%
13	bleu	0	9	9	0%	100%
14	blanc	0	21	21	0%	100%
<b>Contemporary/ Neologisms</b>						
15	cool	0	1	1	0%	100%
16	merdique	0	1	1	0%	100%
17	sexy	0	0	0	-	-
18	super	14	0	14	100%	0%
<b>Famous Alternators</b>						
19	ancien	25	0	25	100%	0%
20	brave	0	0	0	-	-
21	pauvre	0	0	0	-	-
22	propre	9	2	11	82%	18%
23	pur	1	0	1	100%	0%
24	simple	0	0	0	-	-
25	seul	11	0	11	100%	0%
<b>Others</b>						
26	amoureux	0	3	3	0%	100%
27	important	3	3	6	50%	50%
28	joli	3	0	3	100%	0%
29	politique	9	28	37	24%	76%
30	violent	5	4	9	56%	44%
<b>Complete Total</b>						
		878	102	980	90%	10%
<b>Total (nouveau removed)</b>						
		504	100	604	83%	17%



## Appendix B

### Full Combined Data Sets of Syllable Counts

Combined Data Set: Syllable Count								
	Adjective	Syllables	Total Prenominal	Total Postnominal	Total Adjectives	Prenominal Preference	Postnominal Preference	Average Prenominal Preference
1	beau	1	18	0	18	100%	0%	One Syllable 60%
2	blanc	1	0	21	21	0%	100%	
3	bleu	1	0	9	9	0%	100%	
4	bon	1	41	0	41	100%	0%	
5	brave	1	0	0	0	-	-	
6	cool	1	0	1	1	0%	100%	
7	grand	1	85	0	85	100%	0%	
8	gros	1	23	3	26	88%	12%	
9	long	1	3	3	6	50%	50%	
10	pauvre	1	0	0	0	-	-	
11	propre	1	9	2	11	82%	18%	
12	pur	1	1	0	1	100%	0%	
13	rouge	1	0	12	12	0%	100%	
14	seul	1	11	0	11	100%	0%	
15	simple	1	0	0	0	-	-	
16	joli	2	3	0	3	100%	0%	Two Syllables 82%
17	mauvais	2	16	0	16	100%	0%	
18	merdique	2	0	1	1	0%	100%	
19	nouveau	2	374	2	376	99%	1%	
20	petit	2	81	0	81	100%	0%	
21	prochain	2	24	8	32	75%	25%	
22	sexy	2	0	0	0	-	-	
23	super	2	14	0	14	100%	0%	
24	amoureux	3	0	3	3	0%	100%	Three Syllables 61%
25	ancien	3	25	0	25	100%	0%	
26	dernier	3	46	0	46	100%	0%	
27	important	3	3	3	6	50%	50%	
28	politique	3	9	28	37	24%	76%	
29	premier	3	87	2	89	98%	2%	
30	violent	3	5	4	9	56%	44%	

## Acknowledgments

Thank you to the reviewers for their insightful comments and to Dr. Raúl Aranovich, Dr. Susan Palmiter, and Alan Wong for their remarks on earlier versions of this work.

## References

Alexiadou, A., Haegeman, L., & Stavrou, M. (2007). Noun Phrase in the Generative Perspective. *Noun Phrase in the Generative Perspective*, 71, 1-664. doi: 10.1515/9783110207491

Benzitoun, C. (2014). The place of the attributive adjective in French: what we learn speech corpora. 4e Congrès Mondial De Linguistique Française, 8, 16. doi: 10.1051/shsconf/20140801066

Boucher, P. (2006). Mapping Function to Form: Adjective Positions in French. *Linguisticae Investigationes*, 29(1), 43-60.

Brault, G. J. (1978). *The Song of Roland*. University Park: Pennsylvania State University Press.

Brault, G. J. (2010). *Song of Roland: An Analytical Edition: Introduction and Commentary*. Pennsylvania State University Press.

Chomsky, N. (1986). *Knowledge of Language: Its Nature, Origin, and Use*. Praeger.

Cinque, G. (1994). On the evidence for partial N-movement in the Romance DP. *Paths towards universal grammar*, 85-110.

Cinque, G. (2010). *The Syntax of Adjectives: A Comparative Study*. MIT Press.

Cyprien. (2011, May 21, 2016). Cyprien répond à Cortex. Retrieved from <https://www.youtube.com/watch?v=dKwzZZKibUs>

Delbecq, N. (1990). Word Order as a Reflection of Alternate Conceptual Construals in French and Spanish. *Similarities and Divergences in Adjective Position*. *Cognitive Linguistics*, 1(4), 349-416.

Grevisse, M., & Goosse, A. (2011). *Le bon usage : grammaire française : 75 ans*. Bruxelles; [Paris]: De Boeck-Duculot.

Iov, C. (2015a). Presse. Retrieved 30 May 2015, from <http://www.cyprien.fr/index.php/presse/>

Iov, C. (2015b). YouTube Channel: Cyprien.fr. Retrieved 30 May 2015, from <https://www.youtube.com/user/MonsieurDream>

Laenzlinger, C. (2005). French adjective ordering: perspectives on DP-internal movement types. *Lingua*, 115(5), 645-689. doi: 10.1016/j.lingua.2003.11.003

Ledgeway, A. (2012). *From Latin to romance : morphosyntactic typology and change*. Oxford; New York, NY: Oxford University Press.

Lightfoot, D. (1979). *Principles of Diachronic Syntax*. Cambridge University Press.

Lomicka, L., & Lord, G. (2012). A tale of tweets: Analyzing microblogging among language learners. *System*, 40(1), 48-63.

Matthews, P. H. (2007). *Syntactic relations: A critical survey (Vol. 114)*. Cambridge University Press.

Popper, B. (2014). Twitter Now Has 255 Million Users, but Activity Has Dropped Year over Year. Retrieved 20 Mar. 2015, from <http://www.theverge.com/2014/4/29/5665752/twitter-q1-2014-earnings>

Prévost, P. (2009). *The Acquisition of French: The Development of Inflectional Morphology and Syntax in L1 Acquisition, Bilingualism, and L2 Acquisition*. John Benjamins Publishing Company.

Roesslein, J. (2009). Tweepy Documentation. from <http://docs.tweepy.org/en/v3.2.0/>

Sinclair, J., & Carter, R. (2004). *Trust the text : language, corpus and discourse*. London; New York, N.Y.: Routledge.

Sleeman, P., Van de Velde, F., & Perridon, H. (2014). *Adjectives in Germanic and Romance*. John Benjamins Publishing Company.

Smith, C. (2015, 20 May. 2015). *By the Numbers: 150 Amazing Twitter Statistics (May 2015)*. 30 May 2015, from

- <http://expandedramblings.com/index.php/march-2013-by-the-numbers-a-few-amazing-twitter-stats/10/>
- Socialbakers. (2015). Most Popular Twitter Accounts in France. Twitter Statistics for France. Retrieved 30 May 2015, from <http://www.socialbakers.com/statistics/twitter/profiles/france/>
- Song, J. J. (2012). *Word order*: Cambridge University Press.
- Statistica. (2013). Facebook, Twitter and Google Penetration 2009-2013 | France. Retrieved 30 May 2015, from <http://www.statista.com/statistics/417082/social-network-subscribers-among-internet-users-france/>
- Statistica. (2015). Social Networks: Global Sites Ranked by Users 2015. Leading Social Networks Worldwide as of March 2015, Ranked by Number of Active Users (in Millions). Retrieved 30 May 2015, from <http://www.statista.com/statistics/272014/global-social-networks-ranked-by-number-of-users/>
- Stavrou, M. (2012). The syntax of adjectives: A comparative study. *Language*, 88(2), 419-423. doi: 10.1353/lan.2012.0024
- Thuilier, J. (2013). *Syntaxe du français parlé vs. écrit : le cas de la position de l'adjectif épithète par rapport au nom*. TIPA.
- Twitter. (2015a). About Twitter, Inc. Retrieved 30 May 2015, from <https://about.twitter.com/company>
- Twitter. (2015b). Help Center: What's a Twitter timeline? Retrieved 04 Jun 2015, from <https://support.twitter.com/articles/164083-what-s-a-twitter-timeline>
- Twitter. (2015c). Le Monde.fr Twitter. Retrieved 30 May 2015, from <https://twitter.com/lemondefr>
- Twitter. (2015d). Monsieur Dream Twitter. Retrieved 30 May 2015, from <https://twitter.com/MonsieurDream>
- Twitter statistics for France. (2015). Retrieved 20.March.2015, from <http://www.socialbakers.com/statistics/twitter/profiles/france/>
- Wilmet, M. (1980). Anteposition and Postposition of Qualificative Epithets in Contemporary French. *Travaux de Linguistique*, 7, 179-201.
- Yanofsky. (2013, 1 Nov. 2013). A Script to Download All of a User's Tweets into a Csv. Yanofsky/tweet\_dumper.py. Retrieved 30 May 2015, from <https://gist.github.com/yanofsky/5436496>