

Problems and procedures in the construction of diachronic corpora: a case study of the *passé composé* and *passé simple* in Classical French

Magdalene Jacobs

University of Georgia, Athens
magdalene.jacobs@gmail.com

1 Introduction

Diachronically, the Romance languages have been subject to a bevy of developments within their respective aspectual systems. First, a new complex perfect developed in Vulgar Latin and began to share aspectual territory with the existing preterit and imperfect. In the various Romance languages, this compound perfect--formed by combining the auxiliary *habere* with a past participle--began to be employed for many functions that were previously expressed through the simple preterit (Harris 1982: 50). Within the history of French, the morphologically complex *passé composé* developed from the original Latin perfect to a perfective. The perfective *passé simple*, which developed from the Latin preterit, eventually disappeared from use in the spoken language sometime before the 19th century (Dauzat, qtd in Vincent & Harris 1982: 58). The *passé simple* was subsequently relegated to restricted stylistic settings in the written language. Thus, in contemporary spoken French, past events are expressed solely through the use of three tenses: the *passé composé*, the *imparfait*, and the *plus-que-parfait*. Even in contemporary written French, the *passé simple* is arguably used in increasingly smaller frequencies (Hollerbach: 220).

In French, the exact trajectory of the decline of the *passé simple* remains murky, although the change is typically treated as a result of grammaticalization, in which a newer form, in this case the *passé composé*, gradually came to be in semantic variation with the preterit--the *passé simple*--before completely replacing it (c.f. Howe 2009, Bybee 1994, Hopper & Traugott, 1993). According to Hopper and Traugott (1993), across languages, increases in the frequency of new forms in relation to older forms is considered to be an important factor in the replacement of one grammatical structure by another. Along the grammaticalization path, a new category of forms arises--in this case the *passé composé*--and becomes gradually less concrete, i.e. less lexical--and more grammatical. At a given moment, the new forms begin to share semantic territory with the older forms. In the majority of cases, if the new paradigm increases in its frequency, one will observe that it encroaches on the semantic territory of the older forms (Howe: 151). In the case of the *passé composé* and the *passé simple* in French, the *passé composé* began as a 'perfect of result' (See Section 2 of this paper) and became a perfective over time, before completely usurping the *passé simple* in the spoken language (Fleischmann: 83).

The current project does not endeavor to offer an in-depth analysis of the evolution of the relationship between the *passé simple* and the *passé composé*; rather, it is a methodological study designed to observe whether diachronic corpora can provide clarification as to the character of this shift in terms of the frequency of occurrence of both the *passé composé* and the *passé simple*; more specifically, the current project aims to analyze how ongoing linguistic change manifests in different textual genres. To address these questions, this paper first outlines the construction and analysis of a specialized diachronic corpus of Classical French. The corpus is comprised of three different text types made available through the University of Chicago ARTFL archive: Poetry, Treaties and Essays, and Personal Writings, which was comprised of both correspondences and journals.

To carry out this study, it was necessary to compare the frequency of forms in each chronological section of the corpus, as well as to compare relative frequencies across text types in order to observe the extent to which grammar would vary according to genre and whether or not grammatical change could be observed to follow the same trends across textual categories. Such observations are particularly crucial given the fact that the *passé simple* continues to be used in the written language, despite having fallen out of spoken use. Because of the continued literary use of the *passé simple*, a methodological analysis of how to collect optimally representative historical data is a vital component for further study on the decline of *the passé simple* and the rise of the *passé composé* in the spoken language.

The difficulties in tracking historical semantic change are manifold. First, the nature of the change itself is as difficult to define as the character of the variation that inspires it. Additionally, frequent deficits in historical data can render comprehensive quantifications of contexts problematic, and thus diachronic corpus linguistics must often make do with small sampling frames. The lack of data also leads to the creation of linguistic abstractions. In historical studies such as the current one, we often unavoidably obscure the fact that there was never one French Language, but rather a complex and dynamic network of varieties of French. Similarly, the lack of both data and metadata renders it difficult to strictly control for all sociolinguistic variables: such as age, gender, ethnicity, region, and socioeconomic status. Furthermore, since our only data is written, each individual author's stylistic choices are likely to affect the nature of the results.

The impossibility of absolute certainty in historical analyses does not, however, render a corpus-based study of grammatical change fruitless; rather, we must approach all historic data with caution, as providing us with a method to identify patterns and to analyze whether or not these patterns fluctuate or stabilize in systematic ways. Thus, this paper addresses the methodology of building a diachronic corpus while accounting for the problems inherent in historical analyses. In this aim, a comparison of corpora of different sizes facilitates observations on the effect of sampling size on the frequency of distributions of grammatical forms. Furthermore, an examination of the internal composition of each corpus sample, paired with an analysis of the frequency of forms in individual authors' texts, addresses the relationship between varying samples, corpus size, and representativeness.

The following conclusions can be drawn from this study: First, that increased rigor in text type delineation is necessary for accurate results, i.e. the Personal Writings Category, which originally contained both correspondences and memoirs, is erroneous and should be more finely subdivided. Second, that, given a strict enough sampling frame, the results given by a 108,000 Word Corpus can approximate those given by a smaller, less diverse 67,500 Word Corpus in certain textual categories, such as in Treaties and Essays. However, in a text type that is subject to massive stylistic variation between authors, such as Poetry, the results from a 108,000 Word Corpus are not analogous to those from a 67,500 Word Corpus. Finally, in terms of diachronic change, there is an observable increase in the *passé composé* over the *passé simple* over the time periods sampled. In the Treaties and Essays Section of the corpus, there is a particular increase in the use of the *passé composé* with third person pronouns.

2

2.1 Proposed Developments of the *passé simple/passé composé* opposition from Vulgar Latin to Modern French

In contemporary French, the *passé composé* seems to have two broadly delineated domains: 1) it shares perfect territory with the present in prose and 2) it serves as the only past perfective in the spoken language, indicating that an action was completed in the past, whereas the *imparfait* indicates that an action was not completed in the past (Martin, 10). The *passé simple* is restricted to written language, where it continues to be

used in opposition with the *passé composé* for disputed degrees of aspectual opposition and/or stylistic effect. Various studies on Modern written French (Engel 1990, Stavinochova 1974) indicate the *passé composé* is primarily used to relate “anteriority to the present/future, result, successive actions, [and] accomplishment” whereas the *passé simple* is employed for narratives and sequences of discrete events (Engel 1990: 7).

In *Tense and Aspect: A study of French Past Tenses*, Engel (1990) writes that “the *passé simple* evolved from the Classical Latin perfect *feci*, which had the values of a preterite and a present perfect” (4). The *passé composé*, on the other hand, evolved from the Vulgar Latin compound construction which combined the auxiliary *habere* with a past participle and had a primarily aspectual value (Harris 1982: 46-9). According to Harris, the *habeo factum* paradigm originally arose as a resultative construction and gradually developed an aspectual opposition to the *feci* paradigm in Vulgar Latin (147). In Table I, below, the *feci* category represents the category of morphologically simplex forms that would develop into the *passé simple* in French. The *habeo factum* category represents the new aspectual category of morphologically complex forms that are considered to have been grammaticalized, those that would eventually develop into the *passé composé* in French. According to this categorization, the French of the 17th and 18th centuries appears to have been at Stage III in the schema proposed by Harris, whereas contemporary French is at Stage IV.

STAGE	FECI	HABEO FECTUM	LANGUAGE
Stage I.	i. All perfect and perfective functions.	i. Only present states that are results of actions in the past.	Sicilian, Calabrian.
Stage II.	ii. Most past tense functions.	ii. Begins to have “the function of a perfective, but is limited to situations of a particular aspectual profile” (Fleischman: 195).	Galician, Portuguese, Most Varieties of Latin American Spanish
Stage III.	iii. Preterit.	iii. Perfect.	Peninsular Spanish
Stage IV.	iv. Restricted to formal registers, eventually eliminated.	iv. All perfect (completed) past functions	French, Northern Italian, Romanian

Table I: Proposed path of grammaticalization of the *habeo factum* perfect > perfective across the Romance Languages (Table adapted from Fleischman 1983: 195).

One can compare the development of the aspectual system in French with that of the other Romance languages. Spoken varieties of Spanish, for example, productively retain both the complex and simplex forms. However, how these forms are employed for indicating aspect varies across geographic varieties of the language. In the majority of varieties of Spanish in Latin America, the aspectual systems are considered to be at Stage III of Harris’s schema. Contrarily, it is possible that Standard Mexican Spanish is at Stage II, whereas certain varieties of Peninsular Spanish are arguably at or moving towards Stage IV (Howe: 34).

2.2 Old French

The nature of the opposition between the *passé simple* and the *passé composé* in Old French is far from clear. According to Engel (1990), in Old French, use of the *imparfait*, the *passé composé*, and the *passé simple* was

not systematic (4). In *La Disparation du prétérit*, Foulet (1920) argues that the *passé composé*, the *passé simple*, and the *imparfait* were employed “concurrently” for stylistic variation in Old French (see also the discussion from Detges 2000, below). Engel (1990) claims the same, adding that the *passé composé* “had a preterite sense in verse, and a perfect sense in prose” (4). According to Engel, the *passé simple* was still capable of receiving durative readings during the Old French period (4). Furthermore, Engel argues that the *passé simple* was the only category used for sequenced narratives, an aspectual territory which continued to be considered off-limits to the *passé composé* (4).

According to Caudal & Veters (2007: 124), in Old French, the *passé composé*, as opposed to the *passé simple*, was apparently systematically incompatible with “*des modifieurs de localisation temporelle*” that did not include the moment of speech (Caudal & Veters, 124). Caudal & Veters argue that this might indicate that the *passé composé* was not, in fact, semantically compatible with perfectivity in Old French. However, Caudal & Veters also maintain that the *passé composé* could be used to denote a series of discrete past events as early as the 11th century and that, in such instances, the *passé composé* could be analyzed as pragmatically, if not semantically, “close to an aorist” (124). This leads Caudal & Veters to conclude that the *passé composé* was already beginning to take on the function of encoding perfectivity on the “semantic-pragmatic interface” during this period (124-5). Caudal & Veters argue that literary evidence, such as the following examples from *La Chanson de Roland*, indicates that the *passé composé* was already compatible with aoristic contexts. To illustrate this point, in the following extract, the *passé composé* alternates with the *passé simple* in denoting a series of discrete past events.

(1) Sun destre guant a Deu en PUROFFRIT;
Seint Gabriel de sa main l'AD PRIS.
Desur sun braz teneit le chef enclin;
Junte ses main EST ALET a sa fin.
Deus TRANMIST sun angle Cherubin,
E seint Michel del Peril;
Ensembl'od els sent Gabriel VINT. . .”
(La Chanson De Roland CLXXVI, v. 2375-2396, qtd. in Caudal & Veters et al., 125)

In his reference grammar on Old French, diverging from the findings of Caudal & Veters, Buridant (2000) describes the *passé composé* as already fully functionally equivalent with the *passé simple* during this period. If this were the case, the *passé composé* would have already been a perfect IV in Old French. Detges (2006), however, argues that this conflicts with the “widely accepted view than in later epochs of French (which lasted well until the 18th century), the *passé composé*, very much like the present perfect in Modern English or the *perfecto compuesto* in Modern Spanish, was a perfect III” (48). According to Detges, in Old French, the *passé composé* was employed “exclusively” to mark past events with current relevance (c.f. also Fleischman 1982). If this was indeed the case, Detges argues, it would follow that the *passé composé* would not have been used for narratives during this period. Detges maintains that, in Old French narrative texts, the *passé composé* only *appears* to occur in variation with both the *passé simple* and with the historical present for “the marking of foregrounded events” (48). According to Detges, arguments for neutralization between the two categories variation arise from the fact that different translations of the same text often suggest possible functional overlap. Detges cites the following example from the *charroi*:

(2) Si vit ester Guilin et Betran. . . (vit: *passé simple*).
Si a veuz Guyelin et Betreant. . . (a veuz: *passé composé*).
Et voit ester Guielin et Betrant. . . (voit: *présent*).

Detges argues that such seeming neutralization of the opposition between the *passé simple* and the *passé composé* in Old French does not indicate that the *passé composé* was already a variant of the *passé simple*; rather, this neutralization is nothing but an “optical illusion” (49). Detges maintains that such variation is the result of the historical present being employed as a “stylistically marked. . . variant of the *passé simple*” (49).

According to Detges, if the *passé composé* did not encode temporal information but was exclusively aspectual during this period, then it would be default-marked as present tense. The *passé composé* was then actually a variant of the historical present, encoding for resultative aspect, which was in turn a stylistic variant of the *passé simple* (49).

Detges goes on to argue that the primary function of the *passé composé* in Old French was purely aspectual and not temporal, making it a perfect I, or, in Detges's terms a "Resultative B" (49-51). For Detges, Resultative B constructions are those in which the agent of the past event receives the focus, rather than the current result of the past event (51):

- (3) Resultative B
Li quens Rollant, il l' *ad e prise e fraite*
'Count Roland, he is its conqueror and destroyer'
(*ChRol* 663, qtd. in Detges 2000: 50. Emphasis original in Detges).

These types of resultative constructions contrast with those which Detges terms "Resultative A" constructions in which the "current results of past events" are highlighted (50).

- (4) Resultative A
Escababi i *ad le chef trenchet*
'Escababi has had his head cut off there'
(*ChRol* 1555, qtd. in Detges 2000: 50. Emphasis original in Detges).

Detges argues that Resultative A constructions are "extremely rare" in both Old French and Old Spanish texts and that Resultative B constructions are much more frequent (50-1). Furthermore, Detges maintains that Resultative B constructions are in fact the "starting points of the evolution of perfects;" i.e. they are stage I perfects within Harris's schema (51).

2.3 Classical French

How to characterize the *passé simple/passé composé* opposition in the period following Old French is not entirely clear. According to Martin (1971: 383 ff.), in Middle French, the *passé composé* continued to be a pure resultative, which would follow Detges' observations on Old French. The *passé simple*, on the other hand, was a "past punctual" for all events, even those with current relevance. According to Caudal & Veters (2007), in the period of transition from Old French to Classical French, the *passé composé* retained its distribution while the *passé simple* gradually lost its potential preterit readings as an imperfective and a resultative (125). Caudal & Veters argue that the diachronic evolution of the *passé simple/passé composé* opposition then reached a stage of stability from the post-Classical period that corresponds with contemporary readings of the two forms (125).

Some observations regarding the nature of the *passé simple/passé composé* opposition in Classical French can be made from attempted codifications of the aspectual system by prescriptivists (Engel, 4). This codification applied to both the theater by the establishment of *l'unité de temps* (which stated that the action of a play should take place during no more than 24 hours) and also to the language as a whole. In the 17th century, it was declared that the proper use of the *passé composé* was exclusively for denoting that an event occurred during the same day as the moment of speech (the '24 hour rule') (Engel, 5). The exception to this "rule" was the usage of the *passé composé* for denoting events in the distant past if and only if these events seemed "close" to the speaker, i.e. were pragmatically relevant (Engel, 5). Engel maintains that "the question of how rigid this distinction [between the *passé composé* and the *passé simple*] was is disputed, but it is clear that the situation in C.F. was far less confusing and haphazard than in O.F." (Engel: 6).

Caudal & Veters maintain that, in Classical French, the *passé composé* remained in a transitional period between being a semantic resultative given to pragmatically perfective readings and as a "mixed" resultative

and perfective. However, during this period, the *passé composé* does not appear to have gained “*terrain perfectif*” (132). In fact, the *passé composé* appears to have been completely marginal, and perhaps even inexistent, in certain constructions, such as in combination with past adverbial complements. According to Caudal & Veters, in Moliere, Montesquieu, and La Bruyere, *hier + passé composé* is absent as a construction, which contrasts with the abundance of *hier + passé simple* constructions (132). The co-occurrence of *hier + passé simple* is most likely a result of the of the so-called 24 hour rule (Fournier 1998: 398-99).

Dahl (1984: 105) suggests that it was in fact the case that the *passé composé* was characteristically restricted to hodiernal contexts in the 17th century. However, according to Caudal & Veters, the evolution of the *passé composé* towards an aorist appeared earlier, in the 16th century. Caudal & Veters (2007) support this argument by citing cases in which the *passé composé* can be used *without* a temporal complement to refer to the distant past, as seen in (5), below.

(5). . .encore qu’il en eust acqize autant que Cezar memes en a acquis (Monluc qtd. in Caudal & Veters, 131)

In the same vein as Caudal & Veters (2007), other scholars have also focused on characterizing the evolution of the opposition between the *passé simple* and *passé composé* according to the co-occurrence of each with specific temporal adverbial phrases. Following results from his corpus of Middle French, Wilmet (1970) found that “*le passé composé semble mieux toléré lorsque sa détermination temporelle ne précise pas l’intervalle séparant l’événement de l’actualisée*” (283). Wilmet recorded the rate of occurrence of both the *passé simple* and *passé composé* with temporal adverbial phrases in plays from the 15 and 16th centuries, with the results shown in Table II. Wilmet observed that both *hier* and *avant-hier* co-occurred with only the *passé simple* during these periods.

Temporal Reference	<i>passé simple</i>	<i>passé composé</i>
<i>trestout venant</i>	0	1
<i>tout maintenant</i>	0	2
<i>Tantost</i>	0	3
<i>Orains</i>	2	0
<i>ce matin</i>	0	6
<i>Ennuyt</i>	4	4
<i>aujourd’huy</i>	1	9
<i>Huy</i>	6	2
<i>Hier</i>	25	0
<i>avant-hier</i>	6	0
Total	44	27

Table II: The *passé simple*, *passé composé*, and temporal reference from the 15th and 16th centuries (Wilmet 1970: 278)

In the 17th century, according to Caudal & Veters, the *passé composé* was used over the *passé simple* when a temporal complement referred to an interval that included the moment of speech (Caudal & Veters, 127). If

there was no such temporal complement, the choice between the *passé composé* and the *passé simple* appeared to be free. Furthermore, in the 17th century, there are still observable occurrences of the *passé simple* with the adverbial *depuis*, which, due to the connection it creates between a past moment and the moment of speech, is typically considered to have favored almost exclusively the *passé composé*. According to Caudal & Vetters, in such occurrences, the interval described by *depuis* encodes a resulting state, as in (6), below.

(6) Car DEPUIS que le temple de Salomon fut bâti, il n'était plus permis de sacrifier ailleurs, et tous ces autres autels qu'on élevait à dieu sur des montagnes, appelés par cette raison dans l'écriture les hauts lieux, ne lui étaient point agréables" (Racine, *Athalie*, Acte II, Scène 7, qtd. in Caudal & Vetters & Vetters)

Caudal & Vetters maintain that *depuis* could be used with both the *passé composé* and the *passé simple* within the same texts during this time period and that such residual co-occurrence between the two categories of forms indicates that the *passé simple* still allowed resultative readings during this time period. Such readings would be "vestiges" of a much earlier stage of the *passé simple* (130).

Fournier (1998: 399) localizes the disappearance of the *passé simple* from spoken French as occurring in the latter half of the 17th century. However, observations from grammarians of this time period indicate that the *passé simple* was still in use at least in until the early 18th century. In his *Essay d'une parfaite grammaire de la langue française*, Chiflet (1659) described the *passé composé* as having a substantial degree of functional overlap with the *passé simple*, writing "[le] *prétérit indéfini [passé composé]* se peut dire de toute sorte de temps simple passé. Hier j'ay bien soupé : & aujourd'hui j'ay mal dîné," whereas in *Grammaire française sur un plan nouveau*, Buffier (1709) observed that, although speakers were more likely to employ the *passé simple* with bounded events that were anchored by temporal adverbs, it was also possible to use the *passé composé* under such circumstances: "avec un mot qui marque un temps entièrement écoulé, on mettra plutôt le *prétérit simple*, je fis cela hier, j'ai voyagé l'année passée : bien qu'on put dire, j'ay fait cela hier, j'ay voyagé l'année passée." Such descriptions of the language suggest that both forms were in use in the 18th century and that there was an observable degree of functional overlap between them.

According to Caudal & Vetters, it is only in the 18th century that the *passé composé* began to acquire full systematic compatibility with past temporal modifiers. Similarly, in his corpus of epistolary literature, Liu (1999) found that the combination of *passé composé* + *hier* was marginal in the 17th century (1.6 %), then gained the majority in the 18th century (51.7%) and had almost eliminated the *passé simple* + past temporal modifier constructions in the 19th century (97.8 %). The data for *passé composé* + "days of week" evolve in the same vein: 18.3% in the 17th century, 59.7% in the 18th century, and 93.1% in the 19th century. Based on these observations, Caudal & Vetters argue that the transition of the *passé composé* from a value of resultative to one of "mixed resultative and perfective" was spread over a much longer period than is generally acknowledged in the literature (134). According to Caudal & Vetters, this transition can be delineated in the following stages:

Stage I: A pragmatic stage, attained after the Old French period, in which the *passé composé* acquired the possibility of perfective interpretations in "*successions temporelles*," as seen in (7), below.

(7) Vers le palés est alés;
Il en monta les degrés.
En une canbre est entrés,
Si comença a plorer
Et grant dol a demener (Auccassin, VII, 6-10, qtd. in Buridant 2000: 381, qtd. in Caudal & Vetters: 134)

Stage II: A semantic stage, fully terminated in the Post-Classical period, in which the *passé composé* became compatible with "*compléments de temps passés*" (134), as seen in (8).

(8) Hier au soir, avec le secrétaire, tu ne t'es pas ménagé davantage; quand tu remontas tu chancelais, tu ne savais pas ce que tu disais; et aujourd'hui tu as fait dix haltes, et je gage qu'il ne reste pas une goutte de vin dans ta gourde? (Diderot, Jacques le Fataliste, qtd. in Caudal & Vetter).

Thus it would appear that the *passé composé* and the *passé simple* were in variation in classical French but that the character of this variation is not entirely clear. The current study is thus concerned with an initial quantification of the textual contexts in which the *passé composé* and the *passé simple* occurred in order to establish maximally representative sampling frames. The establishment of such sampling frames will then facilitate a more in-depth analysis of the distribution of the two categories of forms in historical text-based data. Thus, the current study does not attempt to mount an analysis of the variable contexts (such as temporal reference) of the *passé simple* and the *passé composé* during the three time periods under consideration; rather this study aims to first address the issue of how to construct representative sampling frames to maximally eliminate bias in the data.

3 The corpora

For the construction of the corpus, texts from three different periods were analyzed, covering a span of 100 years. Samples from texts were selected from the following three chronological frames: 1619-1630, 1660-1670, and 1720-1735. Within each of these frames, samples from three different text types were analyzed: Poetry, Treaties and Essays, and Personal Writing, e.g. journals and correspondences (a complete list of the sampled texts is available in Appendix I). To initially represent these text types, 2,500 word samples from three different authors were analyzed within each category. In sum, this amounted to 7,500 words from each text type in each period and 22,500 total words from each chronological frame. Following this sampling frame, the original corpus consists of 67,000 words.

In order to augment and diversify the original corpus, an additional randomly selected 500 word passage was selected from each author already represented in the corpus, resulting in a 3,000 word sample from each individual author. In the aim of diversification, 3,000 words were added from an additional author for each text type within each chronological frame. The results drawn from this larger, more diverse corpus were then compared with the results from the smaller, less diverse corpus. The compositions of the two corpora are outlined below.

- 67,500 Word Corpus of 17th-18th Century French

I. 1620-1630: 22,000 Words Total

Poetry: 7,500 words

Correspondences and Personal Journals: 7,500 words

Treaties and Essays: 7,500 words

II. 1670-1680: 22,000 Words Total

Poetry: 7,500 words

Correspondences and Personal Journals: 7,500 words

Treaties and Essays: 7,500 words

III. 1720-1735: 22,000 words total

Poetry: 7,500 words

Correspondences and Personal Journals: 7,500 words

Treaties and Essays: 7,500 words

- 108,000 Word Corpus of 17th-18th Century French

I. 1620-1630: 36,000 Words Total

Poetry: 12,000 words

Correspondences and Personal Journals: 12,000 words

Treaties and Essays: 12,000 words

II. 1670-1680: 36,000 Words Total

Poetry: 12,000 words

Correspondences and Personal Journals: 12,000 words

Treaties and Essays: 12,000

III. 1720-1730: 36,000 Words Total

Poetry: 12,000 words

Correspondences and Personal Journals: 12,000 words

Treaties and Essays: 12,000 words

Each 2,500 (and, subsequently, each 500) word sample were taken from the middle of each document, rather than from the beginning or end of a given text. Each verb was manually tagged according to the following categories: *passé composé* (PC), *passé simple* (PS), *imparfait* (VBI), present tense/inflected (VB), conditional (VBC), *plus-que-parfait* (PQP), infinitive (VBT), and future (VBF). Forms that were inflected for both tense and mode were coded accordingly; e.g. a verb that was inflected both as subjunctive and *imparfait* was tagged as <VBI SUBJ>. However, only the tense of these verbs will be included in the final analysis.

Additionally, each verb was tagged for the grammatical person in which it was conjugated. In modern written French, when the *passé simple* does occur, it is overwhelmingly in the context of third person pronouns (Hollerbach, 220). Tagging for grammatical person was undertaken in the aim of observing whether the frequencies with which each tense occurred with each grammatical person were comparable across text types or were subject to variation. Furthermore, such tagging facilitates the comparison of the rate of change of each category within the context of grammatical person across the three time periods.

In addition to being tagged for grammatical person, each verb was tagged according to two other variable contexts: the presence or absence of an object pronoun preceding the verb and the type of verb used in the construction. Six broad categories of verb type were delineated: transitive, intransitive, reflexive, *être* 'to be', *avoir* 'to have,' and *faire* 'to make/do.' A verb was determined to be transitive if it was followed by a direct object. Finally, each verb was tagged for whether or not it was preceded by an object pronoun. The tagging of variables was undertaken in the goal of providing a method of searching and isolating the syntactic environments and semantic contexts of the tenses. However, only the results for grammatical person are presented in this paper.

4 The effect of corpus size on observed frequencies

Frequencies were calculated using the following criteria: 1) the frequencies of the *passé composé* and the *passé simple* were calculated over the total number of both forms due to the fact that these are the forms considered to be in variation, 2) the frequency of the *imparfait*, which is not considered to be in variation with either the *passé composé* or the *passé simple* during these periods, but rather to be semantically distinct, was calculated over the total of number of past tense verbs, i.e. over the total number of occurrences of the *passé composé*, the *passé simple*, and the *imparfait*.

Under these criteria, there is not a great difference in the frequency of forms between the 36,000 Word Treaties and Essays Corpus and the 22,000 Word Corpus. Both show a significant increase in the *passé composé* between 1620 and 1670, from 29% to 52% in the smaller corpus and from 32% to 51% in the larger corpus. Additionally both indicate a leveling in the frequencies of distribution between 1670 and 1720: in the smaller corpus, the frequencies of the *passé composé* (52% in 1670 and 53% in 1720) and the *passé simple* (48% in 1670 and 47% in 1720) are relatively stable between these two periods. In the larger corpus, the *passé composé* shows an increase from 51% to 60% and the *passé simple* shows a decrease from 49% to 40%. The percentages of the *imparfait* between the two corpora are roughly analogous. In the smaller corpus, the *imparfait* maintains a similar frequency across the three periods (1620: 42% > 1670: 43% > 1720: 41%). In the larger corpus, the *imparfait* shows a slight increase between 1620 and 1670 (1620: 43% > 1670: 48%), before stabilizing between the 1670 and 1720 periods. Given the relatively small size of both corpora, a change as slight as 5% is not considered to be significant.

	1620-1630	1670-1680	1720-1730
Passé Simple	123 (71%)	42 (48%)	55 (47%)
Passé Composé	51 (29%)	45 (52%)	63 (53%)
TOTAL	173	87	118
VBI	124/298 (42%)	66/153 (43%)	81/199 (41%)
TOTAL	298	153	199

Table III: Treaties and Essays: 67,500 Word Corpus

	1620-1630	1670-1680	1720-1730
Passé Simple	134/196 (68%)	77/156 (49%)	61/154 (40%)
Passé Composé	62/196 (32%)	79/156 (51%)	93/154 (60%)
TOTAL	196	156	154
VBI	149/345 (43%)	143/299 (48%)	110/264 (42%)
TOTAL	345	299	264

Table IV: Treaties and Essays: 108,000 Word Corpus

For Poetry, there are greater differences between the smaller and larger corpora, a fact which highlights the degree to which authorial style impacts the frequency of forms in this textual category. Whereas the results in the distribution of forms from both corpora in the 1670 period are similar, the results from the other two periods are completely altered when the sample is both augmented and diversified. In the smaller, 7,500 word 1670-1680 Poetry Sample, the *passé simple* occurs 73% of the time, whereas the *passé composé* occurs 27%

of the time. In the larger 12,000 1670-1680 Poetry Sample of the 108,000 Word Corpus, the *passé simple* occurs 72% of the time, whereas the *passé composé* occurs 28% of the time. This is in stark contrast with the other two time periods. In the smaller 7,500 word 1620 Poetry Sample, the *passé composé* is more frequent than the *passé simple*, at 59% compared to 41%. In the sample from the larger corpus, the situation is reversed. The *passé composé* occurs 42.5% of the time, at a lower frequency than the *passé simple* at 57.5%.

In the 1720 period for poetry, both corpora show a greater frequency of the *passé simple*, at 70.5% and 62% in the smaller and larger corpora, respectively. Notably, the frequency of the *imparfait* seems to stabilize in the larger corpus, while being subject to quite massive variation in the smaller sample. In the 65,700 Word Corpus, there is a large degree of difference in the frequency of the *imparfait* between the three time periods: 1620: 27% > 1670: 18% > 1720: 40%. In the 108,000 Word Corpus, the *imparfait* is basically stable between the three periods: 1620: 30% > 1670: 33% > 1720: 28.5 %.

	1620-1630	1670-1680	1720-1730
Passé Simple	48/116 (41%)	82/112 (73%)	86/122 (70.5%)
Passé Composé	68/116 (59%)	30/112 (27%)	36/122 (29.5%)
PS/PC TOTAL	116	112	122
Imparfait	43/159 (27%)	25/137 (18%)	82/204 (40%)
All past tense verbs	159	137	204

Table V: Poetry: 67,500 Corpus

	1620-1630	1670-1680	1720-1730
Passé Simple	111/193 (57.5%)	85/120 (72%)	151/243 (62%)
Passé Composé	82/193 (42.5%)	34/120 (28%)	92/242 (38%)
PS/PC TOTAL	193	120	243
Imparfait	83/276 (30%)	33/153 (33%)	97/340 (28.5%)
TOTAL	276	153	340

Table VI: Poetry: 108,000 Corpus

In addition to creating a larger corpus for all textual categories, it was necessary to construct an even smaller corpus for Personal Writings. The creation of this smaller corpus was crucial due to aberrant results in the Personal Writings category, which initially showed a dramatic increase of the *passé simple* over the *passé composé* across the three time periods. These results are in stark contrast with the hypothesis that the *passé composé* would become more frequent across the three time periods. These results can be observed in Table VII, below and will be discussed in Section 4 of this paper.

	1620-1630	1670-1680	1720-1730
Passé Simple	25/130 (19%)	115/181 (63%)	179/260 (69%)
Passé Composé	105/130 (81%)	66/181 (37%)	81/260 (31%)
PS/PC TOTAL	130	181	260
Imparfait	102/232 (44%)	135/316 (43%)	144/404 (36%)
TOTAL	232	316	404

Table VII: Personal Writings: 67,500 Word Corpus

The aberrance of these results was not greatly improved after augmenting and diversifying the data in this category, as can be seen in Table VIII.

	1620-1630	1670-1680	1720-1730
Passé Simple	39/207 (19%)	134/266 (51%)	202/330 (61%)
Passé Composé	168/207 (81%)	132/266 (49%)	128/330 (39%)
PS/PC TOTAL	207	266	330
Imparfait	151/358 (42%)	201/467 (43%)	199/592 (34%)
TOTAL	358	467	592

Table VIII: Personal Writings: 108,000 Word Corpus

However, a closer investigation into the individual authors in this text type revealed that all Personal Writings samples taken from journals and memoirs, as opposed to samples taken from letters, greatly influenced the frequency of forms in these samples (refer to Section 4).

Under ideal circumstances, one would remove the journals and memoirs from the Personal Writings category and replace them with correspondences. Accordingly, one would then set up a separate corpus--equal in size and diversity--of memoirs in order to measure them against the correspondences. However, due to the lack of sufficiently large samples of correspondences from different authors from the given time frames, this was not possible. Rather, first 1720-1730 was isolated as the time frame in which memoirs were most skewing the data. In 1720-1730, two of the four texts were memoirs, amounting to 50% of the sampled data. These texts were removed from the corpus. Two texts were then removed from the 1620-1630 and the 1670-1680 frames to ensure that each period was equally represented. Thus, once the sample was altered, the Personal Writings category was rendered both smaller and less diverse, being comprised of only 6,000 words from two different authors, the results of which can be observed in Table IX, below.

	1620-1630	1670-1680	1720-1730
Passé Simple	18/90 (20%)	13/97 (13%)	19/119 (16%)
Passé Composé	72/90 (80%)	84/97 (87%)	100/119 (84%)
PS/PC TOTAL	90	97	119
Imparfait	72/162 (44%)	61/158 (39%)	62/181 (34%)
TOTAL	162	158	181

Table IX: Personal Writings 18,000 Word Corpus

When the Personal Writings Corpus was augmented and diversified, there was little change in the frequency of forms between the smaller and the larger samples in the 1620 period. Similarly, when the Personal Writings Corpus was made drastically smaller (6,000 words per time period), the frequency of forms was stable between the different sample sizes within the 1620 period. In the largest and most diverse 1620 sample from the 108,000 Word Corpus, the *passé simple* occurred 19% of the time, whereas the *passé composé* occurred 81% of the time. These numbers are exactly the same for the 1620 sample from the mid-range 67,500 Word Corpus. In the 1620 sample from the smallest and least diverse corpus, containing only 6,000 words from two different authors, the frequencies are almost exactly the same as in the larger corpora: the *passé simple* occurred 20% of the time, whereas the *passé composé* occurred 80% of the time. Importantly, the 1620 period in the Personal Writings Corpus is the only period in which every sample was taken from letters. Since the frequencies between the three corpora in 1620 period are so similar, it is thus possible that, given a rigorous enough sampling frame, a larger corpus is not absolutely necessary to track change in this textual category; rather, it is more pressing to account for the sensitivity of the sampled text types to aspectual variation.

The effect of text type sensitivity becomes even more salient when the 22,000 word sample from 1670 Personal Writings Corpus is compared to both the 36,000 word and the 6,000 word samples. In the original 22,000 word sample, the *passé simple* showed a massive increase from the previous period (1620: 19% > 1670 : 63%). In the textual category that is assumed to most closely approximate spoken language, such an increase in the category of forms that would eventually disappear completely from spoken and/or “informal” language is not only unexpected, but jarring. As will be addressed in Section IV: Text Type Stability, this increase is due to text-type sensitivity and individual authorial style. Within the current discussion, when this increase is compared to the increase of the *passé simple* between the two periods in the larger corpus, we can observe that although small augmentation and diversification of the corpus did not totally eliminate the effects of individual authors and the sensitivity of text types on the frequency results, these changes did diminish the aberrance of the data.

In the 12,000 Personal Writings word sample from the 1670-1680 section of the 108,000 Word Corpus, the *passé simple* occurred 51% of the time and the *passé composé* occurred 49% of the time. The 12,000 1720-1730 word sample from the 108,000 Word Corpus shows similar results: the *passé simple* occurred 61% of the time, whereas the *passé composé* occurred 39% of the time. In terms of the relationship between representativeness and corpus size, one might make the following observation: a larger and more diverse corpus can reduce the effects that individual texts have on the overall frequencies of forms; however, if the construction of such a corpus is not feasible, the determination of extremely precise text type categories might compensate for the paucity of the sample size.

5 Text type stability

Two criteria were used for determining the relative stability of a given text type sample. First, for this preliminary study, since the *imparfait* is not considered to have undergone distributional change between the three time periods, it was analyzed as a control category. The rate of occurrence of the *imparfait* was thus measured against the total occurrence of past tense verbs conjugated in the *imparfait*, the *passé simple*, and the *passé composé*. If the distribution of the *imparfait* varied significantly for the majority of different authors within the same text type, the sample was analyzed as being subject to large degrees of internal variation.

Second, if individual authors remained within ten percentage points of each other in terms of their usage of the *passé composé* and the *passé simple*, respectively, the category to which they were each a member was determined to be relatively “stable” in terms of representativeness. In a much larger corpus, such inter-authorial variation could be relatively stabilized by augmentation and diversification of the data. However, as we will continue to see, in a smaller, less diverse sample, such variation can lead to an individual author’s narrative content or stylistic choices completely altering the frequency results.

In the 1620-1630 Treaties and Essays Category, Gournay and Naude showed similar distributions of all past tense forms, as seen in Table X. Coeffeteau, on the other hand, used almost exclusively the *imparfait*, employing both the *passé simple* and *passé composé* relatively infrequently. Chapelaine displays a different kind of variation, using the *imparfait* at a comparatively stable rate (33%) and mostly employing the *passé composé*, using the *passé simple* only twice.

	GOURNAY	CHAPELAIN	COEFFETEAU	NAUDE
Passé Simple	15/44 (34%)	2/31 (6.5%)	3/7 (43%)	21/65 (32%)
Passé Composé	29/44 (66%)	29/31 (93.5%)	4/7 (57%)	44/65 (68%)
TOTAL	44	31	7	65
VBI	29/73 (40%)	15/46 (33%)	105/112 (94%)	65/130 (50%)
TOTAL	73	46	112	130

Table X: Individual authors from the 1620-1630 12,000 Word Treaties and Essays Word Sample

In the 1620-1630 Poetry Category, Viau, Auvray, and Du Lorens all employ the *passé simple* and the *passé composé* at roughly the same rates. Evaluated against these three authors, the sample from Maynard could be viewed as quite distinct, employing the *passé simple* at a much higher rate than the *passé composé*.

	VIAU	AUVRAY	DU LORENS	MAYNARD
Passé Simple	16/55 (29%)	12/38 (32%)	12/40 (30%)	53/57 (93%)
Passé Composé	39/55 (71%)	26/38 (68%)	28/40 (70%)	4/57 (7%)
TOTAL	55	38	40	57
VBI	21/76 (28%)	22/60 (37%)	12/52 (23%)	28/85 (32%)
TOTAL	76	60	52	85

Table XI: Individual authors from the 1620-1630 12,000 Word Poetry Sample

The 1620-1630 Personal Writings category shows significantly less variation between authors than any other text type in any other period. Notably, this is the only time period in which only correspondences were sampled to meet the text type quota. As seen in Table XII, Balzac uses the *imparfait* at a higher rate than the other authors. However, the rates of occurrence of the *passé simple* and the *passé composé* remain relatively stable between all four authors.

	BALZAC	PEIRESC	MERSENNE	COLIGNY
Passé Simple	6/34 (18%)	12/74 (16%)	10/55 (18%)	11/44 (25%)
Passé Composé	28/34 (82%)	62/74 (84%)	45/55 (82%)	33/44 (75%)
TOTAL	34	74	55	44
VBI	38/72 (53%)	47/121 (39%)	34/89 (38%)	27/71 (38%)
TOTAL	72	121	89	71

Table XII: Individual authors from the 12,000 Word 1620-1630 Personal Writings Sample

The 1670-1680 Treaties and Essays category shows a great degree of inter-authorial variation in the distribution of all tensed forms, as seen in Table XI. Interestingly, the author with the lowest past tense token number overall, Nicole, is also the author with the highest rate of occurrence of the *passé composé*. In terms of inter-authorial variation in the distribution of the *passé simple* and *passé composé*, each author shows a strikingly different percentage, which is also the case for the 1670-1680 Poetry category. Thus, while it was expected that poetry would be subject to wide ranges of stylistic variation, this appears to be the case for Treaties and Essays as well. Additionally, there is a possible effect of the pragmatic context of the sample on the calculated frequencies, which will be discussed in Section 5: Clustering and Sampling.

	ARN. D'ANDILLY	BOUHOURS	NICOLE	AUBIGNY
Passé Simple	27/52 (52%)	42/50 (84%)	1/19 (5%)	10/38 (26%)
Passé Composé	24/52 (48%)	8/50 (16%)	18/19 (95%)	28/38 (74%)
TOTAL	52	50	19	38
VBI	48/99 (48%)	38/88 (43%)	8/27 (30%)	49 (56%)
TOTAL	99	88	27	87

Table XIII: Individual authors from the 12,000 Word 1670-1680 Treaties and Essays Sample:

	LA FONTAINE	MERE	BOILEAU	BENSERADE
Passé Simple	29/54 (53%)	8/37 (22%)	51/56 (91%)	78/92 (85%)
Passé Composé	25/54 (46%)	19/27 (78%)	5/56 (9%)	14/92 (15%)
TOTAL	54	27	56	92
VBI	6/60 (10%)	45/72 (62.5%)	19 (25%)	36/128 (28%)
TOTAL	60	72	75	128

Table XIV: Individual authors from the 12,000 Word 1670-1680 Poetry Sample

The 1670-1680 Personal Writings category shows comparable frequencies between Bussy, Sevigne, and Fournay: each have a much lower percentage (between 10% and 25%) of *passé simple* verbs than *passé composé* verbs (between 77% and 91%). The numbers for Retz are markedly different. Out of 124 *passé composé/passé simple* constructions, Retz employed the *passé composé* only once, whereas his use of the *imparfait* was similar to that of the other three authors. Importantly, Retz's is the only text that is an excerpt from a journal in this category, rather than from a group of letters. In the process of the data analysis, this was the first clue as to the extreme sensitivity of the *passé composé/passé simple* variation between text types. It was thus determined to be insufficient to separate texts into broad categories such as "literary" and "non-literary/personal." The observation that memoirs and journals, as a text type, appear to have a radically different distribution of past tense forms will be apparent in the 1720-1730 sample as well.

	BUSSY	SEVIGNE	RETZ	FOURNAY
Passé Simple	4/44 (9%)	9 (17%)	123/124 (>99%)	14/60 (23%)
Passé Composé	40/44 (91%)	44/53 (83%)	1/124 (<1%)	46/60 (77%)
TOTAL	44	53	124	60
VBI	29/73 (40%)	32 (38%)	110 (47%)	30/90 (43%)
TOTAL	73	85	234	90

Table XV: Individual authors from the 12,000 Word 1670-1680 Personal Writings Sample

A breakdown of the 1720-1730 Treaties and Essays Corpus shows similar inter-authorial variation to the 1670-1680 Corpus. Again, each author has a significantly different distribution of the *passé simple* as compared to the *passé composé*, with the numbers being as low as 5% *passé simple* in Lambert and as high as 61% *passé simple* for Montesquieu. The numbers for the *imparfait* are more stable, ranging from 38% of the total past tense verbs for Meslier to 51% for Dursumais.

	LAMBERT	MESLIER	MONT.	DURSUMAIS
Passé Simple	2/38 (5%)	10/55 (18%)	52/85 (61%)	6/16 (37.5%)
Passé Composé	36/38 (95%)	45/55 (82%)	33/85 (39%)	10/16 (62.5%)
TOTAL	38	55	85	16
VBI	28/66 (42%)	34/89 (38%)	54/139 (39%)	17/33 (51%)
TOTAL	66	89	139	33

Table XVI: Individual authors from the 12,000 Word 1720-1730 Treaties and Essays Sample

A closer examination of the different texts sampled for this category leads to some preliminary observations. First, Lambert, one of only three female authors sampled for the entire 108,000 Word Corpus, overwhelmingly uses the *passé composé* in her essay *Réflexions nouvelles sur les femmes*. As a treatise on the conditions of women in France, the character of Lambert's discourse is thus grounded in the recent past and ongoing present. Similarly, Meslier and Dumarsais's essays are both discussions of rhetoric and philosophy and also show greater occurrences of the *passé composé* over the *passé simple*, whereas Montesquieu's essay, *Réflexions sur la monarchie universelle en Europe*, including discussions of both Rome and of the recent past in Europe, has a higher occurrence of the *passé simple*. The high rate of *passé simple* in Montesquieu could

also be taken to be an effect of analyzing homogenous samples from texts with a high degree of clustering of tenses, which will be discussed in Section 5: Clustering and Sampling.

For the 12,000 Word 1720-1730 Poetry Sample from the 108,000 Word Corpus, one can observe the same trend of extreme stylistic variation between authors as was reflected in the Poetry Categories of the other two time periods. Racine uses the *passé simple* and the *passé composé* almost equally (*passé simple*: 48%, *passé composé*: 52%), whereas La Motte employs a greater percentage of verbs in the *passé simple* (*passé simple*: 79%, *passé composé*: 21%) and Rousseau employs a greater percentage in the *passé composé* (*passé simple*: 16%, *passé composé*: 84%). The excerpts from La Chaussée are interesting in that he uses the *imparfait* at a markedly lower rate than any other author sampled in the entire 108,000 word corpus. This serves as a further indication of extreme stylistic variability in the poetic genre.

	RACINE	ROUSSEAU	LA MOTTE	LA CHAUSÉE
Passé Simple	66/137 (48%)	4/25 (16%)	30/38 (79%)	53/75 (71%)
Passé Composé	71/137 (52%)	21/25 (84%)	8 (21%)	22/75 (29%)
TOTAL	137	25	38	75
VBI	55/192 (29%)	11/36 (30.5%)	20 (34%)	11/86 (13%)
TOTAL	192	36	58	86

Table XVII: Individual authors from the 12,000 Word 1720-1730 Treaties and Essays Sample

The distributions of the past tense forms in the 12,000 Word 1720-1730 Personal Writings Sample from the 108,000 Word Corpus were equally as various as those from the 12,000 Word 1670-1680 Personal Writings Sample. A breakdown of the distributions for each author indicates that the sensitivity of the text types is, again, the probable source for the seemingly anomalous numbers. Saint-Simon and Dagneau both show a significantly greater frequency of the *passé simple* (94% in Saint Simon and 75% in Dagneau) over the *passé composé* (6% in Saint Simon and 26% in Dagneau). The samples taken from these two authors were taken from journals and not from letters, which was the case in the 1670-1680 corpus for the author (Retz) who significantly favored the *passé simple* over the *passé composé*. This further indicates that more finely drawn textual boundaries are necessary for tracking grammatical change through historical corpora. The samples taken from the other two authors in the 1720-1730 Period, Rousseau and Voltaire, were excerpted from letters. It is necessary to note that there is a degree of difference between the distributions of forms between these two authors, but nothing as wildly divergent as the numbers reflected the Saint-Simon and Dagneau samples.

	SAINT-SIMON	ROUSSEAU	DAGNEAU	VOLTAIRE
Passé Simple	99/105 (94%)	7/69 (10%)	105/142 (74%)	12/51 (23.5%)
Passé Composé	6/105 (6%)	62/69 (90%)	37/142 (26%)	39/51 (76.5%)
TOTAL	105	69	142	51
VBI	97/202 (48%)	27/96 (28%)	46/188 (24%)	35/86 (41%)
TOTAL	202	96	188	86

Table XVIII: Individual authors from the 12,000 Word 1720-1730 Personal Sample

6 Clustering and sampling

A significant clustering effect was observed in the Treaties and Essays text types, which clearly influences the representativeness of a given sample. For example, when Montesquieu addresses modern European History, he overwhelmingly uses the *passé composé*. In the sections of his essay where he discusses Rome, however, he uses the *passé simple* almost exclusively. In a randomly selected 800 word sample from the 4,500 word Montesquieu sample, the *imparfait* occurred 17 times, the *passé simple* occurred 37 times, and the *passé composé* occurred only once. In a different randomly selected 800 word sample from the same essay, the *imparfait* occurred 17 times, the *passé composé* occurred 8 times, and the *passé simple* did not occur at all.

This reflects an apparent tendency in the Treaties and Essays category: the author writes a series of paragraphs using almost exclusively the *imparfait* and the *passé composé* and then switches to using exclusively the *imparfait* and the *passé simple*, or vice versa. When a single paragraph contains all three tenses, it is overwhelmingly the case that the author has switched from using the *passé simple* with third person pronouns in an account of past events to using the *passé composé* with a first person pronoun in order to comment on the narrative account itself. This type of narrative switch can be observed in the excerpts (9) and (10) below, taken from Montesquieu:

(9)La monarchie d'Espagne dans les guerres de Philippe III contre la France. . .ne perdit qu'une petite portion d'un coin de terre qu'on attaquoit. Le plus petit peuple qu'il y eut pour lors en Europe soutint contre elle une guerre de cinquante ans avec un avantage égal; et nous avons vu de nos jours un monarque. . . .

(10)La raison qui leur fit établir cette sorte de gouvernement. . . .Or, pour les raisons que nous avons dites, un grand empire, où le prince n'avoit pas une autorité absolue, devoit nécessairement se diviser. . . .Voilà l'origine des royaumes de France, d'Italie, de Germanie, d'Aquitaine, et de tous les démembrements que l'on vit. . . .

The fact that such clustering occurs suggests that it is insufficient to take a homogenous sample from such a text, as was done for this study. In the case of Montesquieu, the thematic content of a given sample of text is shown to influence the distribution of past forms. It would therefore be preferable to sample several smaller portions of text from different paragraphs across the entire textual body.

7 Diachronic Change

Having examined the composition and degrees of representativeness of each corpus, we are now in a position to examine the rate of change in the distribution of forms across each category. In the Treaties and Essays section of the 108,000 Word Corpus, the *passé simple* decreases across all three time periods: 1620: 68% > 1670: 49% > 1720: 40%, whereas the *passé composé* increases: 1620: 32% > 1670: 51% > 1720: 60%. The change of 19 percentage points between 1620 and 1670 is greater than the change of 9 percentage points between 1670 and 1720. The *imparfait*, as predicted, rests stable between all three periods.

	1620-1630	1670-1680	1720-1730
Passé Simple	134/196 (68%)	77/156 (49%)	61/154 (40%)
Passé Composé	62/196 (32%)	79/156 (51%)	93/154 (60%)
TOTAL	196	156	154
VBI	149/345 (43%)	143/299 (48%)	110/264 (42%)
TOTAL	345	299	264

Table XIX: Diachronic distributions across the 36,000 Treaties and Essays Word Sample

In the Poetry section of the 108,000 Word Corpus, the distribution of forms is relatively stable between all three periods. This is as predicted by the hypothesis that change would appear to occur at a slower rate in this text type. Given the great degree of literary variation between individual authors and texts in this category, it is arguable that a much larger and more diverse sample would be necessary before drawing any firm conclusions regarding the rate of change in this category. However, given the stability of the control category, the *imparfait*, one could arguably make the preliminary observation that, in these data, the rate of occurrence of the *passé simple* and the *passé composé* did not change across the three time periods in this category.

	1620-1630	1670-1680	1720-1730
Passé Simple	111/193 (57.5%)	85/120 (72%)	151/243 (62%)
Passé Composé	82/193 (42.5%)	34/120 (28%)	92/242 (38%)
TOTAL	193	120	243
VBI	83/276 (30%)	33/153 (33%)	97/340 (28.5%)
TOTAL	276	153	340

Table XX: Diachronic distributions across the 36,000 Treaties and Essays Word Sample

The imbalanced samples in the 1670-1680 and 1720-1730 sections of the Personal Writings in the 108,000 Word Corpus yielded wildly divergent numbers, as illustrated in table XXI below.

	1620-1630	1670-1680	1720-1730
Passé Simple	39/207 (19%)	134/266 (51%)	202/330 (61%)
Passé Composé	168/207 (81%)	132/266 (49%)	128/330 (39%)
TOTAL	207	266	330
VBI	151/358 (42%)	201/467 (43%)	199/592 (34%)
TOTAL	358	467	592

Table XXI: Diachronic distributions across the 36,000 Personal Writings Word Sample

When the journals and memoirs samples from two authors were removed from each period, yielding a 6,000 word sample for each, the smaller, less diverse corpus shows a relatively stable distribution of the forms across the three time periods.

	1620-1630	1670-1680	1720-1730
Passé Simple	18/90 (20%)	13/97 (13%)	19/119 (16%)
Passé Composé	72/90 (80%)	84/97 (87%)	100/119 (84%)
TOTAL	90	97	119
VBI	72/162 (44%)	61/158 (39%)	62/181 (34%)
TOTAL	162	158	181

Table XXII: Diachronic distributions across the 36,000 Personal Writings Word Sample

8 Grammatical person as a variable

Since the representativeness of 1670-1680 and the 1720-1730 Personal Writings categories in the 108,000 Word Corpus has been deemed erroneous, only the variable findings from 1620-1630 (where all the samples are comprised of letters) will be discussed in the current paper. Additionally, the findings from the smaller 6,000 word samples from all three time periods will also be presented.

Across all the time periods, the extremely small number of verbs conjugated for second person singular *tu* renders it impossible to draw any solid conclusions about the effect of this variable on the occurrence of the *passé composé* or the *passé simple*. Similarly, there are very few occurrences of verbs conjugated for first person plural *nous* and second person plural *vous*. However, when the first or second person plurals are used in the corpus, it is overwhelmingly with the *passé composé*. In the 12,000 word 1620-1630 corpus, 78% of the occurrences of *nous* and all occurrences of *vous* are in the *passé composé*. 90% of the occurrences of first person singular *je* are with the *passé composé*. As predicted, the frequency of third person singular and plural pronouns with the *passé composé* is smaller than that of the first person singular, although both occur at greater frequencies in the *passé composé* than in the *passé simple*. Third person singular pronouns occur 73% of the time with verbs in the *passé composé*. Third person plural pronouns occur 78% of the time with verbs in the *passé composé*.

1620-1630							Total
	1S	2S	3S	1P	2P	3P	
Passé Simple	6 (10%)	0 (0%)	23 (27%)	2 (22%)	0 (0%)	8 (22%)	39
Passé Composé	55 (90%)	0 (0%)	61 (73%)	7 (78%)	17 (100%)	28 (78%)	168
TOTAL	61	0	84	9	17	36	207

Table XXIII: Grammatical Person Distributions of *Passé Simple* and *Passé Composé* in the 12,000 Word Personal Writings Sample from the 108,000 Word Corpus

The percentages drawn from the 12,000 word sample from 1620-1630 sample are similar to those drawn from the 6,000 word sample from the same time period, as seen in Table XXIV. 90.5% of the occurrences of *je*, 84% of the occurrences of *il/elle/on*, and 70% of the occurrences of *ils/elles* are in the *passé composé*.

1620-1630							
	1S	2S	3S	1P	2P	3P	TOTAL
Passé Simple	2 (9.5%)	0 (0%)	5 (16%)	2 (33%)	2 (22%)	7 (30%)	18
Passé Composé	19 (90.5%)	0 (0%)	26 (84%)	4 (66%)	7 (78%)	16 (70%)	72
TOTAL	21	0	31	6	9	23	90

Table XXIV: Grammatical Person Distributions of *Passé Simple* and *Passé Composé* in the 1620-1630 6,000 Word Personal Writings Sample

In the 6,000 Word Corpus from 1670-1680, there is an apparent decrease in the frequency of occurrences of *je* in the *passé composé* (76%) when compared to the 6,000 word sample from 1620-1630 (90.5%), whereas there is an increase in the occurrence of *il/elle/on* (93%). Since there are so few occurrences of either the *passé simple* or the *passé composé* conjugated for the third person plural, these results indicate very little.

1670-1680							
	1S	2S	3S	1P	2P	3P	
Passé Simple	9 (24%)	0 (0%)	3 (7%)	0 (0%)	0 (0%)	1 (20%)	13
Passé Composé	28 (76%)	0 (0%)	43 (93%)	1 (100%)	8 (100%)	4 (80%)	84
TOTAL	37	0	46	1	8	5	97

Table XXV: Grammatical Person Distributions of *Passé Simple* and *Passé Composé* in the Revised 1670-1680 6,000 Word Personal Writings Sample

In the 6,000 word sample from 1720-1730, the results are similar to those from the other two time periods: all grammatical persons appear to favor the *passé composé*. Of any grammatical person, *je* occurs with the *passé composé* the most frequently, at 91% of the time, whereas *il/elle/on* and *ils/elles* occurs with the *passé composé* at 81% and 78%, respectively.

1720-1730							
	1S	2S	3S	1P	2P	3P	
Passé Simple	3 (9%)	0 (0%)	10 (18.5%)	0 (0%)	1 (12.5%)	5 (22%)	19
Passé Composé	30 (91%)	1 (100%)	44 (81%)	0 (0%)	7 (87.5%)	18 (78%)	100
TOTAL	33	1	54	0	8	23	119

Table XXVI: Grammatical Person Distributions of *Passé Simple* and *Passé Composé* in the Revised 6,000 1720-1730 Word Personal Writings Sample

In the Poetry section of the 108,000 word corpus, there is a stability between the three time periods in the relative frequencies of the *passé simple* and the *passé composé* when conjugated with the first person plural pronoun *je*. There is, however, a great degree of fluctuation between the distributions of both categories when conjugated with the third person singular or plural pronouns, as seen in tables XXVII, XXVIII, and XXIX. Third person singular pronouns *il/elle/on* show the following fluctuations when conjugated with the *passé composé* across the three time periods: 1620-1630: 36% > 1670-1680: 19% > 1720-1730: 24%. Third person plural pronouns *ils/elles* show the following fluctuations: 1620-1630: 68% > 1670-1680: 66% > 1720-1730: 89%. Thus, there are no observable trends in terms of how often a third person pronoun occurs with the *passé composé* over the *passé simple* in the Poetry text type, whereas we might posit that likelihood of the *passé composé* being conjugated for first person singular *je* remains relatively stable across the three time periods.

1620-1630							
	1S	2S	3S	1P	2P	3P	TOTAL
Passé Simple	6 (35%)	7 (87.5)	88 (64%)	1 (100%)	1 (20%)	8 (32%)	111
Passé Composé	11 (65%)	1 (12.5%)	49 (36%)	0 (0%)	4 (80%)	17 (68%)	82
TOTAL	17	8	137	1	5	25	193

Table XXVII: Grammatical Person Distributions of *Passé Simple* and *Passé Composé* in the 12,000 1620-1630 Word Poetry Sample from the 108,000 Word Corpus

1670-1680							
	1S	2S	3S	1P	2P	3P	TOTAL
Passé Simple	3 (27%)	1 (33%)	75 (81%)	1 (100%)	3 (100%)	3 (33%)	86
Passé Composé	8 (73%)	2 (66%)	18 (19%)	0 (0%)	0 (0%)	6 (66%)	34
TOTAL	11	3	93	1	3	9	120

Table XXVIII: Grammatical Person Distributions of *Passé Simple* and *Passé Composé* in the 12,000 1670-1680 Word Poetry Sample from the 108,000 Word Corpus

1720-1730							
	1S	2S	3S	1P	2P	3P	TOTAL
Passé Simple	2 (33%)	1 (50%)	143 (76%)	1 (33%)	0 (0%)	4 (11%)	151
Passé Composé	4 (66%)	1 (50%)	44 (24%)	2 (66%)	6 (100%)	35 (89%)	92
TOTAL	6	2	187	3	6	39	243

Table XXIX: Grammatical Person Distributions of *Passé Simple* and *Passé Composé* in the 12,000 1720-1730 Word Poetry Sample from the 108,000 Word Corpus

In the Treaties and Essays section of the 108,000 Word Corpus, first person singular verbs are overwhelming conjugated in the *passé composé*. The overall token count of first person singular verbs is very small,

however, so these results are not particularly revelatory. Third person singular verbs conjugated in the *passé composé* increase over the three time periods: 1620-1630: 19% > 1670-1680: 33% > 1720-1730: 43.5%. Third person plural verbs conjugated for the *passé composé* show a degree of fluctuation across the three time periods: 1620-1630: 48% > 1670-1680: 86% > 1720-1730: 62%.

1620-1630							
	1S	2S	3S	1P	2P	3P	TOTAL
Passé Simple	1 (12.5%)	0 (0%)	109 (81%)	1 (20%)	0 (0%)	23 (52%)	134
Passé Composé	7 (87.5%)	0 (0%)	26 (19%)	5 (83%)	3 (100%)	21 (48%)	62
TOTAL	8	0	135	6	3	44	196

Table XXX: Grammatical Person Distributions of *Passé Simple* and *Passé Composé* in the 1620-1630 12,000 Word Treaties and Essays Sample from the 108,000 Word Corpus

1670-1680							
	1S	2S	3S	1P	2P	3P	TOTAL
Passé Simple	5 (31%)	0 (0%)	68 (66%)	0 (0%)	0 (0%)	4 (14%)	77
Passé Composé	16 (76%)	0 (0%)	34 (33%)	4 (100%)	0 (0%)	25 (86%)	79
TOTAL	21	0	102	4	0	29	156

Table XXXI: Grammatical Person Distributions of *Passé Simple* and *Passé Composé* in the 1670-1680 12,000 Treaties Sample from the 108,000 Word Corpus

1720-1730							
	1S	2S	3S	1P	2P	3P	TOTAL
Passé Simple	0 (0%)	0 (0%)	37 (43.5%)	4 (50%)	1 (33%)	19 (38%)	61
Passé Composé	8 (100%)	0 (0%)	48 (56%)	4 (50%)	2 (66%)	31 (62%)	93
TOTAL	8	0	85	8	3	50	154

Table XXXII: Grammatical Person Distributions of *Passé Simple* and *Passé Composé* in the 1720-1730 12,000 Treaties and Essays Sample from the 108,000 Word Corpus

9 Conclusion

Diachronically, the *passé composé* appears to increase in its frequency across the three time periods when only the most stable text type, Treaties and Essays, is taken into account. In both Poetry and Correspondences, the distribution of the *passé composé* remains fairly stable across all three periods. In this study, text type stability was determined by comparing the relative frequencies of forms between the specific sampled texts against the total frequency in each text types. Two varieties of text type sensitivity are indicated as being

significant. First, there is the stylistic nature of the text itself. In the case of Poetry, there are wide ranges of variation between authors. This diversity is taken to be an effect of the highly variable stylistic nature of the poetic genre. Second, it is clear that “Personal Writings” is too broad a category. In fact, personal letters display significantly different frequencies of the targeted forms than memoirs. When letters and memoirs are grouped together within the same text type, the frequency of forms becomes highly aberrant. A small degree of augmentation and diversification significantly decreases the erroneousness of the frequencies in this category, which suggests that a much larger corpus might compensate for such sensitivity. However, given the fact that historical data is difficult to obtain, the creation of a larger and more diverse corpus would be a difficult goal to reach. A more realistic goal would be increased rigor in sampling frames. When the sampling frame is more finely delineated, i.e. when personal correspondences are separated from memoirs, it seems that corpus size has a decreased effect on the results, i.e. the frequency of forms in a 6,000 word sample is analogous to the frequency of forms in both a 7,500 word sample and a 12,000 word sample when the text type is rigorously controlled for.

10 Appendix I: Texts Analyzed

I. 1620-1630

10.1 A. Treaties and Essays

- 1) Gournay, Marie Le Jars de, 1565-1645. [1622], *Égalité des hommes et des femmes* (Éd. Milagros Palma. Paris: Côté-femmes, 1989) [genre: traité ou essai].
- 2) Chapelain, Jean, 1595-1674. [1623], *Lettre ou discours de Monsieur Chapelain à Monsieur Favereau, Conseiller du Roi en sa cour des aides portant son opinion sur le poème d'Adonis du Chevalier Marino* (In *L'Adone, Poema Del Cavalier Marino*, Parigi, O. Di Varano, 1623.) [genre: traité ou essai]
- 3) Coeffeteau, Nicolas, 1574-1623. [1623], *Histoire romaine* (Paris, G. Loyson, 1646.) [genre: traité ou essai]
- 4) Naudé, Gabriel, 1600-1653. [1625], *Apologie pour tous les grands hommes qui ont esté accusez de magie*. (Paris, Eschart, 1669.) [genre: traité ou essai]

10.2 B. Poetry

- 1) Viau, Théophile de, 1590-1626. [1621], *Oeuvres poétiques. 1re partie* (Genève, Droz, 1951.) [genre: poésie]
- 2) Auvray, Jean 1590?-1633. [1623], *Le banquet des muses* (Rouen, D. Ferrand, 1636.) [genre: poésie]
- 3) Viau, Théophile de, 1590-1626. [1623], *Oeuvres poétiques. 2e partie* (Genève, Droz, 1958.) [genre: poésie]
- 4) Maynard, François de, 1582?-1646. [1619], *Le philandre* (Genève, J. Gay et Fils, 1867.) [genre: poésie]

10.3 C. Correspondences and Memoirs

- 1) Balzac, Jean-Louis Guez, seigneur de, 1597-1654. [1624], *Lettres, 1624* (In *Les Premières Lettres*, T.I. Paris, Droz, 1934) [genre: correspondance].
- 2) Peiresc, Nicolas Claude Fabri de, 1580-1637. [1625], *Lettres de Peiresc. Tome sixième, Lettres de Peiresc à sa famille et principalement à son frère, 1602-1637* (Paris, Impr. Nationale, 1896.) [genre: correspondance]
- 3) Mersenne, Marin, 1588-1648. [1627], *Correspondance. T. 1, 1617-1627* (Paris, P.U.F., 1945.) [genre: correspondance].
- 4) Coligny, Louise de, 1515-1620. [genre: correspondance]

II. 1670-1680

10.4 A. Treaties and Essays

- 1) Arnauld d'Andilly, Monsieur (Robert), 1588-1674. tr. [1670], Fondations Monasteres Carmel. (In Les Oeuvres de Sainte- Thérèse. Paris, P. le Petit, 1670.) [genre: traité ou essai]
- 2) Bouhours, Dominique, 1628-1702. [1671], Les entretiens d'Ariste et d'Eugène (Paris, A. Colin, 1962.) [genre: traité ou essai]
- 3) Nicole, Pierre, 1625-1695. [1671], Essais de morale. T. 1 (Paris, G. Desprez, 1701.) [genre: traité ou essai]
- 4) Aubignac, François-Hédelin, abbé d', 1604-1676. [1676], Conjectures académiques (Paris, Hachette, 1925.) [genre: traité ou essai]

10.5 B. Poetry

- 1) La Fontaine, Jean de, 1621-1695. [1671], Fables nouvelles (In Oeuvres Diverses, Paris, Gallimard, 1942.) [genre: poésie]
- 2) Méré, Antoine Gombaud, chevalier de, 1610-1684. [1671], Les conversations, discours de la justesse (Ed. Ch. H. Boudhours, F. Roches, 1930.) [genre: poésie]
- 3) Boileau Despréaux, Nicolas, 1636-1711. [1674], L'art poétique (In Oeuvres Complètes, T.2. Paris, les Belles Lettres, 1952.) [genre: poésie]
- 4) [Benserade, Isaac de, 1613-1691. [1676], Métamorphoses d'Ovide (Paris, Imprimerie Royale, 1676.) [genre: poésie]]

10.6 C. Correspondences and Memoirs

- 1) Bussy, Roger de Rabutin, comte de, 1618-1693. [1672], Les lettres de messire Roger de Rabutin, Comte de Bussy, T. 3, 1666-1672 (Paris, F. Delaulne, 1720.) [genre: correspondance]
 - Sévigné, Marie de Rabutin-Chantal, marquise de, 1626-1696. [1675], Correspondance. T. 1, 1646-1675 (Paris, Gallimard, 1972.) [genre: correspondance]
 - Retz, Jean François Paul de Gondi de, 1613-1679. [1679], Mémoires. T. 5 (In Oeuvres, Paris, Hachette, 1880.) [genre: mémoires]
 - Foigny, Gabriel de, ca. 1630-1692. [1676], La terre australe connue (In F. Lachevre, les Successeurs de Cyrano de Berg., Paris, Champion, 1922.) [genre: récit de voyage]
- 1720-1730

10.7 A. Treaties and Essays

- 1) Lambert, Anne Thérèse de Maguenat de Courcelles, marquise de, 1647-1733. [1727], Réflexions nouvelles sur les femmes (Éd. Milagros Palma. Paris: Côté-femmes, 1989.) [genre: traité ou essai]
- 2) Meslier, Jean, 1664-1729. [1729], Mémoire des pensées et des sentiments. T. 1 (In Oeuvres Completes, Ed. R. Desne, T.1. Paris, Ed. Anthropos, 1970.) [genre: traité ou essai]
- 3) Montesquieu, Charles de Secondat, baron de, 1689-1755. [1727], Réflexions sur la monarchie universelle en Europe (In Oeuvres Compl., Ed. R. Caillois, T.2. Paris, Gallimard, 1951.) [genre: traité ou essai]
- 4) Dumarsais, 1676-1756. [1730], Des tropes ou Des différens sens dans lesquels on peut prendre un même mot dans une même langue: ouvrage utile pour l'intelligence des Auteurs, & qui peut servir d'introduction à la Rhétorique & à la Logique. (Paris, Vve Brocas, 1730.) [genre: traité ou essai]]

10.8 B. Poetry

- 1) Racine, Louis, 1692-1763. [1720], *La grâce* (Paris, J-B Coignard, J. Desaint, 1742.) [genre: poésie]
- 2) Rousseau, Jean-Baptiste, 1670-1741. [1723], *Odes, cantates* (Paris, Didot, An 7-1799.) [genre: poésie]
- 3) La Motte, M. de (Antoine Houdar), 1672-1731. [1719], *Fables* (In *Oeuvres*, T.9. Paris, Prault L'Aine, 1754.) [genre: poésie]
- 4) La Chaussée, Nivelle de, 1692-1754. [1732], *Épître de Cléo* (In *Oeuvres*, T.5. Paris, Prault, 1762.) [genre: poésie]

10.9 C. Correspondences and Memoirs

- 1) Saint-Simon, Louis de Rouvroy, duc de, 1675-1755 [1720], *Mémoires Tome 18* (Chéruef, Paris)
- 2) Rousseau, Jean-Jacques, 1712-1778. [1728], *Lettres* (1728-1778) (La Guilde du Livre, Lausanne), Ed. Marcel Raymond [genre: correspondance]
- 3) Dangeau, Philippe de Courcillon, marquis de, 1638-1720. [1713], *Journal du marquis de Dangeau. T. 14, 1711-1713* (Paris, F. Didot, 1858.) [genre: mémoires]
- 4) Voltaire. *Correspondance Tome 1 : Décembre 1704 - Décembre 1738 de Voltaire.* [genre: correspondance]

Bibliography

- Buffier, C. (1709). *Grammaire française sur un plan nouveau*. Paris : Nicolas Le Clerc.
- Buridant, C. (2000). *Grammaire nouvelle de l'ancien Français*. Paris : Sedes.
- Bybee, J., Revere P. and Pagliuca W. (1994) *The evolution of grammar: Tense, aspect and modality in the languages of the world*. Chicago: University of Chicago Press.
- Caudal, P. & Vettters C. (2007). *Passé composé et passé simple: sémantique diachronique et formelle*. In P. Caudal, E. Labeau, & C. Vettters (eds) *Sémantique et Diachronie du Système Verbal Français*.
- Detges, U. (2006). *Aspect and Pragmatics. The passé composé in Old French and the Old Spanish perfecto compuesto*. In Eksell, Kerstin; Vinther, Thora (Hrsg.) *Change in Verbal Systems. Issues on Explanation*. Frankfurt/M. [u.a.]: Lang, S. 47-72.
- Engel, D. (1990) *Tense and Aspect: A study of French Past Tenses*. London: Routledge.
- Fleischman, S. (1983). From pragmatics to grammar : diachronic reflections on complex pasts and futures in Romance. *Lingua* 60:183-214.
- Foulet, L. (1920). La disparition du prétérit , *Romania*, 46, 271-313.
- Foulet, L. (1923/1930). *Petite syntaxe de l'ancien français*. 3rd rev. ed., Paris: Champion.
- Hollerbach, W.. (1994) *The Syntax of Contemporary French*, Lanham, MD: University Press of America.
- Hopper, P. J., and Traugott E. C. (1993). *Grammaticalization*. Cambridge: Cambridge University Press.
- Howe, C. (2009). Revisiting perfect pathways: Trends in the grammaticalization of periphrastic pasts. *New Challenges in Typology II*, ed. by Pattie Epps and Alexandre Arkhipov, Berlin: Mouton de Gruyter, 151-174.
- Harris, M. (1982). The past simple and the present perfect in Romance. in N. Vincent and M. Harris (eds), *Studies in the Romance Verb--Essays offered to Joe Cremona on the Occasion of his 60th Birthday*. London, Canberra: Croom Helm, 42-70.
- Martin, R. (1971). *Temps et Aspect: Essai sur l'emploi des temps narratifs en moyen français*. Paris : Centre de Philologie et de Littérature romanes.

Wilmet, M. (1970). *Le système de l'indicatif en moyen français. Etude des tiroirs de l'indicatif dans les farces, sotties et moralités françaises des XV et XVI siècles*. Genève: Droz, Publication romanes et françaises CVII.