The Fall of So, Esto, Do, and Vo and Rise of Soy, Estoy, Doy, and Voy

Jana M. Thomas Coffman
University of Alabama

Abstract: Modern students and speakers of the Spanish language often note that the first-person present tense singular indicative forms of the Spanish verbs ser, estar, dar, and ir ("to be," "to be," "to give," and "to go," respectively) are strangely irregular, as each is spelled with a word-final [y] that is absent from other first-person present tense verbs in Spanish. Yet from the emergence of Proto-Iberian, the mother tongue of modern Portuguese, Spanish, Galician, and other dialects and languages found on the Iberian Peninsula, until the Middle Ages, the first-person singular indicative forms of these Spanish verbs were actually regular.

While prior research in Spanish historical linguistics succeeded in finding patterns among the time and rate of these verbal shifts, modern access to vast online corpora has opened the field to new and reinvigorated study. This article outlines prior scholarship related to the gradual shift and replacement of these regular verbs with their modern-day counterparts; it continues by delving anew into the shifts undergone by these verbs in the light of global access to broader corpora of historical Spanish documents. Data tokens of verb pairs were pulled from the Corpus de Español (CdE) and descriptive and inferential statistical analyses were performed. Fisher exact and χ^2 tests revealed that while the timeline of the most-studied verbal shift, ser, remained loyal to the findings of previous research, the order and rate of change of the other three verbs, especially estar, differed from prior literature.

Keywords: Spanish, historical linguistics, Old Spanish, yod, *ser, estar, ir, dar*

Introduction to the Literature

Until the 1200s, the Spanish first-person present indicative forms of *ser* ("to be"), *estar* ("to be"), *dar* ("to give"), *and ir* ("to go) were regular in the *so, esto, do,* and *vo,* respectively. After the 1200s, however, the verb *so* began to exist in variation with *soy* [soi̯], until finally surpassing the older form in the 1400s and eliminating its rival form by the middle of the 1500s. *Esto, do,* and *vo* followed a century later but completed their changes faster, around the same time as *ser.*

Much of the literature on the topic is quite old, including non-scientific speculations for the changes beginning in the 1400s (Nebrija 1492) and continuing to the present day. Given the unresolved question of how and why these verbs changed, several resources (Díaz 2016; Granvik 2009; Pensado Ruiz 2000; Santano Moreno 2009) simply include textbook-like descriptions of the verbs over the years without attempting to isolate a definitive explanation for why the changes occurred. Many of the articles, therefore, are very similar in content and ideas and serve simply to summarize the main theories prior linguists had

hypothesized about the epenthesis of the yod. These theories include the idea that, through analogy with *haber* (*ha*<*hay*), the verbs had fused with the Old Romance particle still seen in Modern French *y* ("there") (Lathrop 2003; Lloyd 1987; Penny 2002; Pharies 2007; Rini 1999) or the agglutination of post-verbal *yo* (Gago-Jover 1997; de Gorog 1980; Lloyd 1987). Other linguists posited the addition of -y served to distinguish the stressed -ó of these monosyllabic verbs from the third-person singular preterite forms of regular Spanish verbs, which also carried a tonic -ó (Lloyd 1987; Penny 2002). Others reported perhaps *so* changed to *soy* through analogy with the final yod present in the first-person singular indicative preterite form *fui* (Gutiérrez-Rexach 2016; Wanner 2006), or even from contact with Leonese or Portuguese verbs that exhibited similar patterns (Gutiérrez-Rexach 2016; de Gorog 1980; Santano Moreno 2009).

It is clear from the history of Spanish verbs that the Classical Latin ESSE ("to be") morphed into the Vulgar Latin ESSERE, but its forms were influenced by those of other verbs during the transition into Old Spanish, as the modern Spanish verb *ser* ("to be") is derived from forms borrowed from both Latin ESSE < ESSERE and from Latin SEDERE ("to sit)" (Díaz 2000; Lathrop 2003; Nadeau & Barlow 2013). The present indicative paradigm of ESSE is seen below:

SUM	SUMUS
ES	ETIS
EST	SUNT

Penny (2002) states the first-person form of the verb suffered apocope of the final *-m*; he assumes this apocope, while unusual for monosyllabic words, came about through analogy with other first-person verb forms in Spanish, none of which maintained a final *-m*, yielding sum < so. Lathrop (2003) offers a somewhat different explanation for the medieval Spanish form *so*: he projects SUM would have developed first into SUN, much as seen with TAM < tan ("as") and QUEM < quien ("who"), but believes the final *-n* suffered deletion to maintain a difference between first-person singular *son* and third-person plural *son* (> SUNT). Regardless, the Latin SUM became the Old Spanish *so*, which as early as the thirteenth century, began to coincide with *soy*.

The earliest speculations are found in Nebrija's 1492 *Gramática de la lengua castellana*, who, in speaking of the formulation of indicative verbs, noted that with monosyllabic verbs, "por ser tan cortos algunas vezes por hermosura añadimos *.i.* sobre la *.o.* como diziendo *.do. doi. vo. voi. so. soi. sto. stoi*" (Esparza & Saramiento 1992: 345). This idea that the verbs were pronounced as a diphthong for aesthetic reasons (*hermosura* is "beauty" in Spanish) was, however, rejected in Valdés' 1535 *Diálogo de la lengua* (Santano Moreno 2009) and is at any rate a subjective opinion rather than a scientifically-based linguistic conclusion.

Addition of the final yod was also proposed via analogy with the well-attested merger of Old Spanish *ha + y / ha + i*. The adverb *y*, and its orthographic variation *i,* existed in Old Spanish and is well-attested to have combined with the third-person singular form of *haber* ("to have"), itself from the Latin HABERE, yielding *hay* ("there is/there are") from *ha + y* (Lathrop 2003; Lloyd 1987; Pharies 2007; Penny 2002; Rini 1999). Santano Moreno (2009) notes this may have been possible due to the shared meaning of existence of both *haber* and *ser.* This same phenomenon also gave rise to the modern French equivalent, *il y a* ("there is/there are") from corresponding French infinitive *avoir* ("to have"), also descended from HABERE (Granvik 2009; Penny 2002). Lathrop (2003) attributes the /j/ ending of these verbs to the permanent merging of the verbs with this Old Spanish adverbial affix *-y,* although the sequence of his verb changes differs somewhat from much of the literature in that he believes the change started with the Leonese *do* and then spread to *so* and *vo* by virtue of their monosyllabic first-person singular forms, and then finally to *esto* by analogy with *soy.*

Researchers have suggested the final yod may also have resulted from the agglutination of post-verbal *yo.* It is possible to imagine *so yo* < *soy,* etc. (Lloyd 1987). In his corpus study of these verbs, Gago-Jover (1997) found that all four of the old forms were found with a post-verbal *yo*, and *so* and *do* exhibited more cases of a final overt subject pronoun than the other two. As in the fourteenth century he found a higher proportion of the modern forms *soy* and *doy,* he surmised the change may have been precipitated by the presence of *yo* after the verb. This theory can be phonologically represented as [so jo] < [so̯ jo], leaving [so̯i] when the optional pronoun was omitted (Pensado Ruíz 2000).

Lloyd (1987) hypothesized the epenthetic yod might have arisen due to the tonic nature of the final /o/ in the three monosyllabic verbs, *so, do,* and *vo.* Very few Spanish verbs are monosyllabic, resulting in an atonic pronunciation of the regular -o ending of first-person indicative verbs. The tonic pronunciation of Old Spanish [só], [dó], and [vó] may have presented difficulties in distinguishing the forms of the -ó of these monosyllabic verbs from the third-person singular preterite forms of regular Spanish verbs, which also carried a tonic -ó (Lloyd 1987; Penny 2002).

Like Lloyd, Wanner (2006) also proposed the changes were affected by the verbs' preterite forms, but through analogy rather than differentiation. The first-person singular indicative of both Spanish *ser* and *ir* in the preterite is *yo fui,* an irregularity inherited from the original Latin verbs; it is possible the final /j/ of *fui* was simply transferred to its corresponding present-tense form (Gutiérrez-Rexach 2016; Santano Moreno 2009; Wanner 2006).

Finally, some linguists suggested the changes came about through contact with other medieval Spanish dialects; there are Leonese and Portuguese verbs that exhibited similar patterns of diphthongization. This idea was first put forth

by Staaff in his 1907 analysis of medieval Leonese texts, where he found tokens of *do* + *y:*

> *"do y la otra heredat a este monasterio"* (Staaff 1907, 39)
> *"do hy cuanto eredamiento a Sancta Maria de Piasca"* (Staaff 1907, 39)

Warren also noted the emergence of synchronous instances of *so/soy* in Leonese in the early thireenth century, followed quickly by cases of do coexisting with Leonese *doi* in the late thirteenth century (Santano Moreno 2009; Warren 2006). Certain Portuguese words likewise allow for a tonic /ou/ diphthong to be realized as an /oi/; Lloyd (1987) and de Gorog (1980) note this is seen both from Latin descendants in Portuguese and in Portuguese dialectal differences, such as *coisa* ("thing") from Latin CAUSA and alternative forms of *doitor – doutor* ("doctor"), *oitro – outro* ("other"), and *oiro – ouro* ("gold"). Therefore, contact with other dialects or languages on the Iberian Peninsula at that time is a potential explanation, given the corresponding verbal Portuguese forms of *sou, dou, vou,* and *estou*, and the known contact between medieval Portuguese to Galician to Leonese to Castilian (de Gorog 1980; Gutiérrez-Rexach 2016; Lloyd 1987; Santo Moreno 2009).

Weaknesses of Existing Theories. Many of these theories have been partially or completely debunked. While the *y* particle might work for *ha* > *hay* ("there is"), *so* > *soy* ("I am there"), *esto* > *estoy* ("I am there"), and even *vo* > *voy* ("I go there"), it cannot explain the epenthesis of yod for *do* > *doy*: (*"I give there") (Pharies 2007). It likewise fails to explain why these first-person forms formed permanent attachments to the adverb when the only other example found is the third-person singular *ha* (Rini 1999). A survey of available texts proves that immediate proximity to an overt post-verbal *yo* did not influence whether those verbs epenthesized the yod (Santano Moreno 2009). Moreover, Granvik (2009) finds the agglutination theory doubtful because it failed to yield the forms *soyo, *doyo, voyo, or *estoyo, as well as an epenthetic yod to other disyllabic verbs such as *traigo* ("I bring"), which, although seen in ancient texts with the post-verbal *yo,* did not yield **traigoy*.

Likewise, an Old Spanish preference toward diphthongizing monosyllabic verbs explains *so, do,* and *vo,* but cannot explain the disyllabic *estoy.* If *so* < *soy* through analogy with its preterite form *yo fui* ("I was") (Gutiérrez-Rexach 2016; Wanner 2006), one would expect *vo* to have changed at the same time, as it shares the first-person preterite form *fui.* However, *vo* changed over a century later and, notably, after the verb *dar,* making that theory unlikely. While contact with Leonese (or Portuguese via Galician via Leonese) is a possibility, Castilian forms dominated their Leonese counterparts in almost all cases, making Leonese contact an unlikely explanation (Lloyd 1987). Therefore, the cause and nature of the change remains unsolved (de Gorog 1980; Gutiérrez-Rexach 2016; Martínez-Gil 2012; Pharies 2007; Santano Moreno 2009; Wanner 2006).

Research Questions

That first-person present indicative verbs *so, do, esto,* and *vo* underwent a word-final epenthesis of the yod /j/ is incontrovertible; this class of verbs draws historical linguists because *why* and *how* they suffered this change remains an unsolved mystery. As it is clearly impossible to interview the long-deceased speakers of medieval languages, their attitudes toward adopting these new changes will never be known. Historical linguists are relegated, therefore, to investigating the written relics left behind by a minority of that epoch's population. It is only through quantifying the *how much* and *when* that linguists can hope to shed light on the *why* and *how.*

Spanish is unique among extant Romance languages in the epenthesis of the /j/ to these first-person verbs. That, and the continuing puzzle as to why they changed, has traditionally made these verbs a hot topic among historical linguists, if the term "hot topic" can truly be applied to a centuries-old linguistic change. It is, at least, a continuing anomaly. Given that these verbs all underwent the same change, the overarching research questions guiding the hypotheses are:

1. When did each verb begin to change?
2. When did the older forms die? (Put another way, how long did the two varieties exist in competition?)
3. In what order did the verbs change, or was it a synchronous change?

Hypotheses

H0: There is no significant difference between tokens of first-person singular present indicative verbs with /j/ and without /j/ throughout the thirteenth through seventeenth centuries.

H1: There is a significant difference between tokens of first-person singular present indicative verbs with /j/ and without /j/ throughout the thirteenth-seventeenth centuries.

$H1_A$: There is a significant difference between tokens of the first-person singular present indicative form of the verb *ser* with /j/ and without /j/ throughout the thirteenth–seventeenth centuries.

$H1_B$: There is a significant difference between tokens of the first-person singular present indicative form of the verb *estar* with /j/ and without /j/ throughout the thirteenth–seventeenth centuries.

$H1_C$: There is a significant difference between tokens of the first-person singular present indicative form of the verb *dar* with /j/ and without /j/ throughout the thirteenth–seventeenth centuries.

$H1_D$: There is a significant difference between tokens of the first-person singular present indicative form of the verb *ir* with /j/ and without /j/ throughout the thirteenth–seventeenth centuries.

$H1_E$: There is a significant difference between tokens of *all four* first-person singular present indicative verbs with /j/ and without /j/ throughout the thirteenth-seventeenth centuries.

Methodology

Data were pulled from the searchable corpus *Corpus de Español (CdE)*. A search was conducted on each verb pair (so/soy, vo/voy, esto/estoy, and do/doy) with a one-word collocate of *yo* in either direction. For the verb *ser,* the orthographic variant *soi* was also searched and grouped with *soy.* The corpus did not yield results for orthographic variations of the other verbs—*voi, estoi,* or *doi*—so these were omitted. Then, data were aggregated in Excel and tokens of homonyms removed. Examples of homonymous tokens include *so* as an abbreviation of *solo* ("only/just"), *do* as a shortened form of *donde* ("where"), and *esto* ("this"). While the demonstrative pronoun *esto* was not truly a homophone with the first-person singular form of *estar*—Wanner (2006) maintains they were pronounced *esto and estó,* respectively—the audial stress was not preserved in orthography, and therefore the search returned examples of each.

Once descriptive scatterplots and/or other charts were completed for each verb, two statistical tests were chosen to determine the statistical relationship among the data. A series of χ^2 tests was performed to determine if a statistically significant difference existed among the *ser* verb pairs and the total verb pairs over the centuries. Due to data samples < 5, making the use of a χ^2 test inappropriate for this data set, the relationship between the competing verb pairs *estar, dar,* and *ir* were compared running Fisher exact analyses at $p \leq 0.05$. Results are tabulated and discussed below.

Results

Once tokens from the *CdE* had been counted and homonyms removed, the resulting data set can be seen below in Table 1.

Table 1: Total Tokens of All Verbs

Years C.E.	Without -/j/	With -/j/
1200	520	59
1300	383	44
1400	321	324
1500	24	2235
1600	32	3007

As seen in Table 1, in the thirteenth century, there were 520 instances of *so, do, esto,* and *vo,* whereas there were only 59 examples of their counterparts with

yod /j/. This indicates the shift had begun in the thirteenth century but not yet taken hold in the language of most speakers. In the fourteenth century, tokens without /j/ still maintained a clear majority of 383 vs. 44 with /j/. However, in the fifteenth century, the use of verbs with and without /j/ was almost equal at 324 and 321, respectively. From the sixteenth century on there were exponentially more data tokens with the increase of literacy and publication, but the yod-less forms had almost disappeared, with only 24 tokens in the sixteenth century and 32 in the seventeenth in comparison with 2,235 with yod in the sixteenth century and 3007 in the seventeenth. By the seventeenth century, the change was complete and the yod-less verb forms ceased to be used. However, the distribution of the four verbs was not equal: for instance, there were far more recorded instances of *ser* verb forms than the other three, at 5,540 tokens, over 3,000 higher than the next largest category, *dar* at 1,136, followed by *estar* at 572 and *ir* at 234.

The relationship between each verb pair (or, in the case of *ser,* a verb triad due to the orthographic variation of the yod as *soi/soy*) was graphed using scatterplots. The relationship among competing *ser* forms, without statistical analysis, is shown in Figure 1.

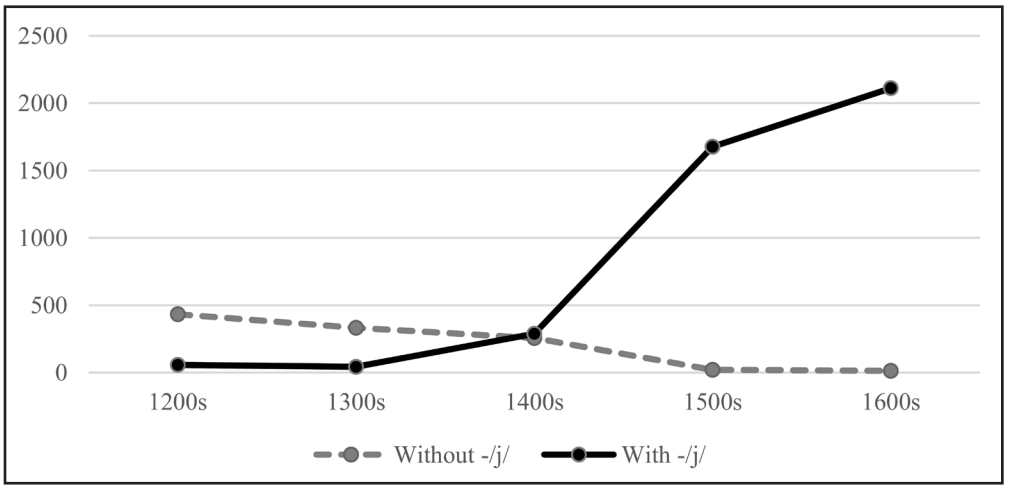

Figure 1: Tokens of Yo + First-Person Singular Present Indicative of *Ser* by Century

In Figure 1, there already were 57 tokens of *soy/soi* during the thirteenth century, but *so* was the dominant form and remained dominant until the 1400s, when it was slightly overtaken by *soy/soi,* and from there the yod form of the verb attained clear dominance. *Estar, dar,* and *ir* underwent a similar pattern, but the change gained momentum at a later time; analysis of the results showed the verb *esto* was still more prominent than *estoy* in the 1400s, but shortly the yod form gained precedence and followed the same pattern as *ser.* The results from *estar* are significant for modern historical linguistics as this is the first study that suggested such an early change from the original *esto* to the modern *estoy;* earlier research consistently placed the transformation

of *estar* as synchronous with *dar* and *ir,* a full century to 150 years later. Given this unexpected result, more research is needed to confirm and explore the role of *estar* in this four-verb class.

Also of note in comparison with earlier studies, *dar* underwent the same changes as the other verbs, but with a slightly different rate of change than first reported: here, the yod form did not approach equal use with the non-yod form until after the 1400s, but gained prominence at a faster rate than the others, also completing the change to yod dominance by the 1500s. Figure 2 shows these results for all four sets of data taken together.

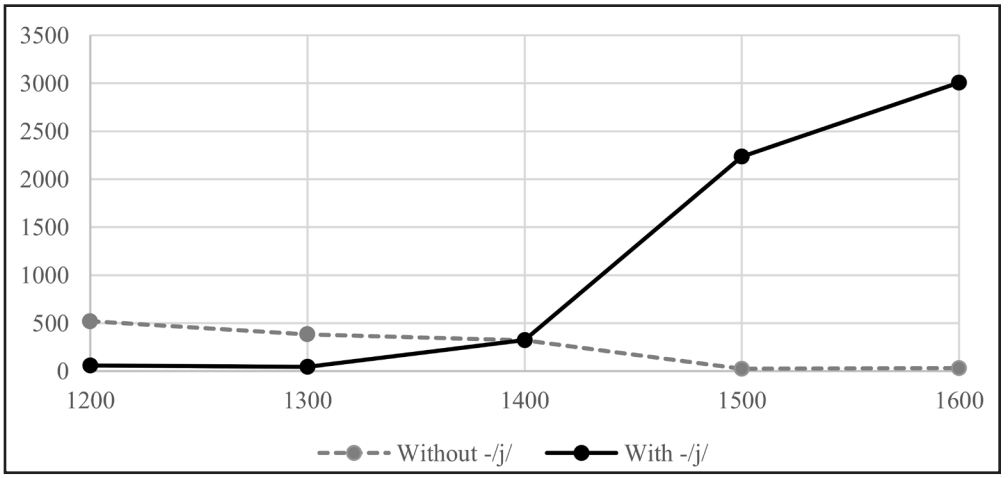

Figure 2: Tokens of the Four Closed-Class Verbs with and without /j/ by Century

Finally, when taken as a whole, the crisscross X shape of the competing verbs' rise and fall is clear. The entire class is graphed by century above in Figure 2. The death of the yod-less forms *so, do, esto,* and *vo* can be seen by their decline to almost 0 tokens, and the growth of the competing forms with yod can be seen starting slowly from the 1200-1400s, then finally elbowing out its competitors and growing exponentially in usage in the 1500s. To determine if the descriptive differences visible in the figures and tables above were statistically significant, a series of χ^2 and Fisher exact tests were run.

Results: Statistical Tests

The χ^2 (4) value for all four of the verbs in the class was 4891.98 $p \leq 0.0001$. There is therefore a significant difference in the use of the yod /j/ for this verb class over the centuries. Then, a second χ^2 test was performed on the *ser* set. As shown in Table 2, the χ^2 (4) value for the set was 3639.66 $p \leq 0.0001$. This result is also statistically significant.

Table 2: χ^2 Values for *So* vs. *Soy/Soi* from the Thirteenth to Seventeenth Centuries		
Years	**So vs. Soy**	
	So	*Soy/soi*
1200	433 *98.75* (1131.38)	57 *391.25* (285.55)
1300	333 *75.78* (873.17)	43 *300.22* (220.38)
1400	256 *109.83* (194.52)	289 *435.17* (49.10)
1500	19 *341.59* (304.65)	1676 *1353.41* (76.89)
1600	13 *428.05* (402.44)	2111 *1695.95* (101.57)
$\chi^2 = 3639.663$, df = 4, $p \leq 0.0001$		

While the eventual change from *so* to *soy* is established fact, this χ^2 test proves the difference is *statistically significant* and the degree of significance—in this case, $p \leq 0.0001$, or a 99.99% chance that the relationship is not due to chance. Establishing that the overall difference is significant set the groundwork for later χ^2 tests between the centuries to observe in which centuries the change occurred.

For this purpose, a series of χ^2 tests (*ser)* and Fisher exact tests (*estar, dar,* and *ir)* were run. In this case, tokens were examined century by century to identify statistically important changes to the verb paradigm. The results for the χ^2 analysis of the verb *ser* are shown in Table 3, while the Fisher Exact analyses for *estar, dar,* and *ir* are tabulated below in Tables 4, 5, and 6, respectively. A summary of these results taken together is shown below in Tabel 7.

Table 3: χ^2 Values for *Soy* vs. *Soy/Soi* by Century				
Years	**1300**	**1400**	**1500**	**1600**
1200	0.0438 0.8342 *Not sig.*			
1300		166.95 $p \leq 0.00001$ *Sig.*		
1400			805.057 $p \leq 0.00001$ *Sig.*	
1500				2.938 $p \leq 0.0865$ *Not sig.*

Table 4: Fisher Exact Analysis of *Esto* vs. *Estoy* by Century				
Years	**1300**	**1400**	**1500**	**1600**
1200	0.40 *Not sig.*			
1300		0.0016 *p ≤ 0.01*		
1400			0.000 *p ≤ 0.01*	
1500				1.0 *Not sig.*

Table 5: Fisher Exact Analysis of *Do* vs. *Doy* by Century

Years	**1300**	**1400**	**1500**	**1600**
1200	1.0 *Not sig.*			
1300		0.0909 *Not sig.*		
1400			0.000 $p \leq 0.01$	
1500				0.3169 *Not sig.*

Table 6: Fisher Exact Analysis of *Vo vs. Voy* by Century

Years	**1300**	**1400**	**1500**	**1600**
1200	0.4706 *Not sig.*			
1300		0.1074 *Not sig.*		
1400			0.000 $p \leq 0.01$	
1500				0.4370 *Not sig.*

Table 7: χ^2 Values for Whole Verb Class from the thirteenth to the seventeenth Centuries

Years	**Whole Verb Class**	
	Without /j/	*With /j/*
1200	520 *519.7* (0.00)	59 *59.3* (0.00)
1300	383 *383.28* (0.00)	44 *43.72* (0.00)
$\chi^2 = 0.0035,\ p \leq 0.953$		
1300	383 *280.42* (37.53)	44 *146.58* (71.70)
1400	321 *423.58* (24.84)	324 *221.42* (47.53)
$\chi^2 = 181.685,\ p \leq 0.00001$		
1400	321 *76.63* (779.33)	324 *568.37* (105.07)
1500	24 *268.37* (222.52)	2235 *1900.63* (30.00)
$\chi^2 = 1136.922,\ p \leq 0.00001$		
1500	24 *23.88* (0.00)	2235 *2235.12* (0.00)
1600	32 *32.12* (0.00)	3007 *3006.88* (0.00)
$\chi^2 = 0.0011,\ p \leq 0.973$		

Table 8: Significant Verbal Change by Century					
	Ser	**Estar**	**Dar**	**Ir**	**Whole Group**
1200-1300	nsd	nsd	nsd	nsd	nsd
1300-1400	**	*	nsd	nsd	**
1400-1500	**	*	*	*	**
1500-1600	nsd	nsd	nsd	nsd	nsd
*p≤0.01 **p≤0.0001 nsd = no significant difference					

Tables 2 and 3 show a significant difference in the trajectory of *so* and *soy/soi* from the thireenth–fifteenth centuries with a confidence interval of 99.99%. After the sixttenth century, change continued to occur, but it was no longer significant. This pattern of significant change in the thirteenth–fifteenth centuries is mirrored by the forms of *estar* with a 99% confidence interval, as demonstrated in Table 4. As illustrated in Tables 5 and 6, change did not become significant for *dar* or *ir* until the fifteenth century, but these changes happened faster, with significant change only occurring in a one-century window between the 1400 and 1500s. Both these were revealed to be significant with a Fisher exact (4) test at $p \leq 0.01$, or a 99% confidence interval. Changes occurring for these verbs in other centuries were not significant. Finally, Tables 7 and 8 show verbal movement of the four verbs taken together as a whole class.

Interpretation

Given these results, the null hypothesis H0 (There is no significant difference between tokens of first-person singular present indicative verbs with /j/ and without /j/ throughout the thirteenth–seventeenth centuries) was rejected. The χ^2 analysis proved the group of verbs, taken as a whole, endured significant change with 99.99% confidence. Therefore, H1 (There is a significant difference between tokens of first-person singular present indicative verbs with /j/ and without /j/ throughout the thirteenth–seventeenth centuries) was accepted, and each sub-hypothesis examined.

Given the results of the χ^2 test, $H1_A$ (There is a significant difference between tokens of the first-person singular present indicative form of the verb *ser* with /j/ and without /j/ throughout the thirteenth–seventeenth centuries) was rejected for the thirteenth and seventeenth centuries, but accepted for the fourteenth and fifteenth centuries ($p \leq 0.0001$). Likewise, $H1_B$ (There is a significant difference between tokens of the first-person singular present indicative form of the verb *estar* with /j/ and without /j/ throughout the thirteenth–seventeenth centuries)

was rejected for the thirteenth and seventeenth centuries but accepted for the fourteenth and fifteenth ($p \leq 0.01$).

Finally, the $H1_C$ (There is a significant difference between tokens of the first-person singular present indicative form of the verb *dar* with /j/ and without /j/ throughout the thirteenth–seventeenth centuries) was rejected for the thirteenth–fifteenth centuries and sixteenth but accepted between the fifteenth and sixteenth ($p \leq 0.01$). The $H1_D$ (There is a significant difference between tokens of the first-person singular present indicative form of the verb *ir* with /j/ and without /j/ throughout the thirteenth–seventeenth centuries) was accepted, but only for the fifteenth–sixteenth century ($p \leq 0.01$). $H1_E$ (There is a significant difference between tokens of *all four* first-person singular present indicative verbs with /j/ and without /j/ throughout the thirteenth-seventeenth centuries) was accepted for two centuries, the thirteenth–fifteenth, and rejected for the other two.

In general, these results support previous literature, with one exception. For instance, the idea that *ser* changed first, beginning in the 1200s, and took 200 years to complete the change, was supported by the results of this study. Likewise, it confirms that *ir* and *dar* endured significant change beginning a full century later, but at a faster rate, completing the change around the same time as *ser.* Most interestingly, the results of *estar* deviate from what was expected based on previous literature. Many descriptive studies have placed the addition of the yod to *estar* as much later than *ser,* either synchronously with *dar* and *ir* or perhaps even half a century later (Martínez-Gil 2012; Wanner 2006). This study, however, suggests an earlier start to the adoption of the yod in *estar,* fully a century earlier than expected and synchronously with *ser.* This novel result for the verb *estar* suggests, with the modern availability of online corpora, historical texts, and databases that was unprecedented when many of the topic's foundational articles in the 1970-80s were written, further exploration on this verb class is merited.

Conclusion

Suggestions for future investigations include removal of the *yo* collocate from tokens and the use of other corpora as data sources. This could allow for the discovery of texts with tokens of orthographic variations *doi, voi,* and *estoi.* Inclusion of other potential orthographic variations of the yod for each verb, i.e. *soe* in addition to *soi/soy,* would also be useful. Although current literature suggests the addition of the yod to *haber (ha + y = hay)* occurred due to fusion with the *y* ("there") particle and this verb class did not, an exploration into a relationship of when the two similar changes occurred could yield interesting results as to whether the different changes might have interacted with one another. Finally, this study analyzed the verbs by number (#) of tokens, but the corpus and other corpora

could also be analyzed by the number of texts in which the verbs occurred rather than the number of tokens in all texts.

These statistical analyses yielded some expected and some new results and should be replicated with larger data sets to test reliability. *Ser, dar,* and *ir* suffered significant change in accordance with previous investigations, while *estar* underwent change earlier than expected, beginning in the thirteenth century. Although the addition of the yod to *ser, estar, dar,* and *ir* was completed centuries ago, there is still work to be done to fully understand these verbs and the phonetic changes they endured.

Works Cited

Davies, Mark. (2002). *Corpus del Español: 100 Million Words, 1200s-1900s*. Web. 1 May 2016.

Díaz, Miriam. (2016). "Chapter Nine: Semantic Changes of Ser, Estar, and Haber in Spanish: A Diachronic and Comparative Approach." *Diachronic Applications in Hispanic Linguistics.* Ed. Eva Núñez Méndez. Newcastle upon Tyne: Cambridge Scholars Publishing. 303-04. Print.

Gago-Jover, Francisco. (1997). "Nuevos Datos sobre el Origen de Soy, Doy, Voy, Estoy." *La Corónica: A Journal of Medieval Spanish Language and Literature.* 24.2: 75-90. Print.

Gorog, Ralph de. (1980). "L'origine des Formes Espagnoles Doy, Estoy, Soy, Voy." *Cahiers de Linguistique Hispanique Médiévale.* 5.1: 157-62. Print.

Granvik, Anton. (2009). "Doy, estoy, hay, soy, y voy: La Combinación atípica de cinco monosílabos con una terminación extraparadigmática. Estado de la cuestión." *Estudios de Historiografía Lingüística.* Bastardín Candón, Teresa and Manuel Rivas Zancarrón, eds. Cadiz: Servicio de Publicaciones de la Universidad de Cádiz. 307-32. Print.

Gutiérrez-Rexach, Javier, ed. (2016). *Enciclopedia de lingüística hispánica.* London: Routledge. Print.

Lathrop, Tom. (2003). *The Evolution of Spanish.* 4th ed. Newark, Delaware: European Masterpieces. Print.

Martínez-Gil, Fernando. (2012). "Sobre la eclosión histórica de soy, doy, voy, estoy y hay: Una solución prosódica. " *Actas del VIII Congreso Internacional de Historia de la Lengua Española: Santiago de Compostela.* 1: 935-46. Print.

Nadeau, Jean-Benoit and Julie Barlow. (2013). *The Story of Spanish.* New York: St. Martin's Press. Print.

Nebrija, Antonio de. (1992). Esparza, Miguel Ángel and Ramón Sarmiento, eds. *Gramática castellana, introducción y notas de Miguel Ángel Esparza y Ramón Sarmiento.* Madrid: Fundación Antonio de Nebrija. Print.

Nebrija, Antonio de. (1492). *Gramática de la lengua castellana.* Asociación Cultural Antonio de Nebrija. Pub. 2007. Web. 16 Aug. 2017.

Lloyd, Paul M. (1987). *From Latin to Spanish, Vol. 1: Historical Phonology and Morphology of the Spanish Language.* Philadelphia: American Philosophical Society. Print.

Penny, Ralph. (2000). *Variation and Change in Spanish.* Cambridge: Cambridge UP. Print.

———. (2002). *A History of the Spanish Language.* 2nd ed. Cambridge: Cambridge UP. Print.

Pensado Ruíz, Carmen. (2000). "De nuevo sobre doy, estoy, soy, y voy." *Cuestiones de actualidad en lengua española.* Borego Nieto, Julio, Jesús Fernández González, Luis Santos Río, and Ricardo Senabre Sempere, eds. Salamanca: Ediciones Universidad de Salamanca. 187-95. Print.

Pharies, David A. (2007). *Breve historia de la lengua española.* Chicago: U of Chicago P. Print.

Rini, Joel. (1999). *Exploring the Role of Morphology in the Evolution of Spanish.* Amsterdam: John Benjamins Publishing Co. Print.

Santano Moreno, Julián. (2009). "Español soy, estoy, doy, voy: Un intento de explicación morfológica." *De morfología y sintaxis españolas: Dos estudios interpretativos.* Milan: Edizioni Universitarie di Lettere Economia Diritto. Print.

Staaff, Erik. (1907). *Étude sur l'Ancien Dialecte Léonais D'après des Chartes du XIIIe Siècle*. Uppsala: Almqvist & Wiksell. Print.

Wanner, Deteri. (2006). "An Analogical Solution for Spanish Soy, Doy, Voy, and Estoy." *Probus: International Journal of Latin and Romance Linguistics* 18.2: 267-308. Print.