1-1-2014

Variable Subject Doubling in Spoken Parisian French

Sara Zahler
Indiana University, szahler@indiana.edu

Variable Subject Doubling in Spoken Parisian French

Abstract

Subject doubling (SD) has long been a topic of interest for researchers. Numerous quantitative studies demonstrate highly variable rates of SD across dialects (e.g. Campion 1984, Auger & Villeneuve 2010, Nadasdi 2000, Sankoff 1982, Conveney 2003, 2005). However, few of the previous studies are variationist in nature, and those that are only examine one or two factors.

Consequently, the goal of this study is to undertake a comprehensive variationist analysis of SD in Parisian colloquial French (PCF), a relatively understudied dialect of French in variationist linguistics in general, and particularly with regard to SD. All tokens of preverbal 3rd person subjects that could be doubled (full noun phrases, strong pronouns and other types of pronouns, such as *lui*, *eux*, *personne*, etc.) were extracted from the from the Corpus de Français Parlé Parisien des Années 2000 (CFPP2000), a French corpus consisting of sociolinguistic interviews from various Parisian suburbs. 17 interviews spanning 14 Parisian suburbs and 25 speakers were used, resulting in 1097 occurrences. These tokens were subsequently coded for the following factors claimed to affect SD: specificity and definiteness of the subject; grammatical person; type and complexity of subject; presence and type of intervening elements; sentence type; verb type; polarity and use of ne; information status; age; sex; and suburb. The data were subsequently analyzed using GoldvarbX.

Several factors were significant and supported claims made in previous qualitative studies regarding subject doubling, particularly subject complexity, use of *ne*, intervening elements, sentence type, and specificity and definiteness of the subject. These results are similar to results found for other dialects, suggesting that many of the same constraints condition the use of SD in the different varieties of French. Nevertheless, grammatical person, information status, sex and age differed, indicating that although there are cross-dialectal similarities, there are certain dialect-specific uses of SD, specifically pertaining to social factors. Lastly, the importance of studying the role of frequency in SD is discussed.

Variable Subject Doubling in Spoken Parisian French

Sara L. Zahler.

1 Introduction

Subject doubling (SD), the co-occurrence of a subject clitic and a subject noun phrase (Nadasdi 1995), has long been a topic of interest for researchers. Numerous quantitative studies demonstrate highly variable rates of SD across dialects (e.g., Sankoff 1982, Campion 1984, Nadasdi 1995, Coveney 2003, 2005, and Auger and Villeneuve, 2010). However, most of the previous studies of SD in French either focus on dialects outside of Europe or minority languages and dialects within France, particularly first and second language varieties of Canadian French (Auger 1993, 1994, 1995, Nagy, Blondeau and Auger 2003, Auger and Villeneuve 2010). Additionally, while most investigations of SD discuss influencing factors, few of the studies are variationist in nature. Consequently, this study is the first quantitative investigation of SD in spoken Parisian French to examine a multitude of linguistic and social factors presented in the extant literature as influencing SD.

Specifically the goal of this investigation is to answer three main research questions: Firstly, what is the rate of SD in spoken Parisian French? Secondly, which linguistic and extralinguistic factors constrain SD in this variety? Lastly, how does the patterning of factors compare with the results found in previous research on other varieties of French focusing on this particular variable?

2 Background

Subject doubling (SD) is the co-occurrence of a subject clitic and a subject noun phrase (Nadasdi 1995), as in (1b), in contrast to the standard non-doubled subject, in which only the noun phrase (NP) is employed (1a).

(1) a. Dans le septième arrondissement **mes parents** sont arrivés. (7.1.2)
         'My parents arrived in the 7th arrondissement'.
     b. **Mes parents ils** se sont séparés quand j'avais trois ou quatre ans. (7.3.1)
         'My parents separated when I was three or four.'

This phenomenon has been attested in many spoken varieties of French and numerous studies have found variable percentages of SD across said varieties: 21% in middle class Parisian French (Ashby 1980); more than 80% in the speech of two speakers from Marseille (Sankoff 1982); 96.4% in adolescent speech from Villejuif, a Paris suburb (Campion 1984); 70% in the speech of two Montreal speakers (Auger 1991); 27% in Ontario (Nadasdi 1995); 24.4% in Picardie (Coveney 2003, 2005); 45% in the Saguenay region of Québec (Auger and Villeneuve 2010) and 55% in Montreal (Sankoff 1982). Based on these numbers, it is evident that SD is a highly variable phenomenon, and that even in varieties of French where it is dominant, SD is not categorical. This variability has furnished studies that discuss the different sociolinguistic factors that constrain the use of SD (e.g., Auger 1993, 1995, 1996, King and Nadasdi 1997, Coveney 2003, 2005, among others) or that investigate the rate of SD in various dialects of French and languages related to French. Nevertheless, despite the existence of several investigations concerning exactly this subject, little statistical work has been completed to corroborate theoretical explanations of SD variation.

With regard to variationist work concerning Parisian French, Ashby (1980) studied whether age, gender and profession correlate with the rate of SD in middle-class Parisian French: he did not include any linguistics factors in his analysis. Ashby found that men doubled subjects more

frequently than women while the middle age group had the highest rate of SD compared to the younger and older age groups. Barnes (1985) also investigated SD using recordings of spoken standard French from France. She investigated the effect of information status, that is, whether the subject was new to the discourse or previously mentioned. Her results showed that SD was more frequent with subjects referring to new information (61%) than with previously mentioned subjects (39%).

Several variationist studies of SD in non-Parisian varieties of French are pertinent to this study. Coveney (2003) studied the effect of *ne*-deletion and social factors on SD in a northern variety of French. Pooley (1996), Nadasdi (1995), Nagy, Blondeau and Auger (2003), and Auger and Villeneuve (2010) also investigate the effect of numerous linguistic and social factors on SD in several varieties of French (Chtimi, Ontarian French, L2 French by Montreal Anglophones, and Quebec French, respectively). The independent variables examined in this investigation are a combination of the constraints tested in these four studies. These factors are presented in the following section with relevant results included from the four studies mentioned.

3 Factors Studied

3.1 Strong Agreement

Among theoretical discussions of subject doubling, several factors have been proposed as possible explanations for subject doubling. One such account is that doubled subject clitics are manifestations of strong agreement between the verb and the subject. According to Auger (1996), certain linguistic constraints are related to verbal agreement, particularly noun definiteness and specificity. This tendency has been observed cross-linguistically: in Hungarian and Zulu, verbal agreement only occurs with definite objects (Marácz 1987, Wald 1979). Additionally, Vasseur (1996) claims that differences exist between specific subjects and non-specific subjects with regard to SD.

However, despite the fact that the relationship between noun specificity/definiteness and SD is frequently proposed, there are only three studies with somewhat divergent results that examine the variable effect of these two factors. Nadasdi (1995) and Nagy et al. (2003) found that specific and definite subjects favor SD. On the other hand, Auger and Villeneuve (2010) showed that although [+specific] subjects favored SD more than [-specific] ones, a sub-category of [-specific] subjects, generic subjects, were the most favorable. Furthermore, non-quantified [-definite] subjects co-occurred most with SD, despite [+definite] subjects favoring SD more than [-definite] ones. These results are due to the fact that specificity and definiteness coincide conceptually (e.g., generic subjects are [-specific], yet [+definite]). Consequently, these two factors were combined and coded as outlined in Table 1, with examples provided in English (Nadasdi 1995, cf. Thibault 1983) in order to differentiate the effect of specificity from that of definiteness in this study's data.

	[+definite]	[-definite]
[+specific]	Jean's brother	Those/the ones (fr. *ceux,* but
	André	not *ceux-là,* 'those there')
	Their house	Three of my cousins
	Those things there	
	(fr. *ces choses-là)*	
[-specific]	The average worker	Some data
	Today's youth	Someone (fr. *quelqu'un*)
	The people	Some funny stories
	Everyone	Whoever

Table 1: Coding for definiteness and specificity (cf. Thibault 1983).

3.2 Grammatical Person

Animacy is also cited as a linguistic factor that corresponds to verbal agreement. Barlow (1988) notes that in Arabic, Abkhaz, Fulani, Manam, Turkana and western Greenlandic the trait [±animate] is a factor in noun agreement. Auger (1996) also asserts that this constraint affects SD

in Quebec French: animate subjects favor SD. Auger and Villeneuve (2010) empirically confirm this hypothesis: their study demonstrates a strong effect of animacy on SD. However, Nagy et al. (2003) show a slight connection between inanimate subjects and SD. In Pooley (1996) animacy does not influence SD in either direction.

Plurality has been shown to affect SD as well: King and Nadasdi (1997) found that plural subjects correspond to a higher rate of SD than singular subjects (32% vs. 15%). As a result, subjects were coded for three categories of animacy (animate, inanimate material and inanimate non-material subjects) and subsequently subdivided into two categories of number (singular and plural), resulting in six combinations of animacy and plurality.

3.3 Subject Type

Given the convergent results from three previous studies (Nadasdi 1995, Nagy et al. 2003, Auger and Villeneuve 2010), the data were coded according to subject type. In each study mentioned above, strong pronouns and proper nouns were shown to favor SD while common nouns did not. As a result, the subjects were coded according to four categories (cf. Auger and Villeneuve, 2010): strong pronouns (e.g., *lui* 'he'*, eux* 'they') other pronouns (e.g., *quelqu'un* 'someone'), proper nouns (e.g., *Pierre, Papa, Paris*), and common nouns (e.g., *les examens* 'the exams'). Auger and Villeneuve (2010) included a category for verbs or clauses (e.g., *c'qu'on fait* 'what we do'). However, this category was not considered in the present study since sentences of the type given above are infrequent in the corpus used (2 tokens), neither of which occurred with a doubled subject.

3.4 Subject Complexity

Subject complexity may have an effect on SD: subject NPs that are longer or more complex due to the presence of relative clauses or complements require more processing. Subject NPs with complements or that are clausal in nature generally have space between the verb and head noun, potentially resulting in SD. Subject complexity has been previously examined by Campion (1984) and, Auger and Villeneuve (2010), whose results support this hypothesis. Auger and Villeneuve found that the most complex NPs, such as those with more than one complement or relative clauses, favor SD the most. Post-nominal adjectives, pre-nominal modifiers and prepositional complements are also favorable. On the other hand, NPs containing either no complement or post-nominal modifiers do not favor SD. These results indicate a general correspondence between rate of SD and subject complexity. While she did not label this factor as subject complexity, Campion found that subject NPs containing a relative clause favored SD. One difference is that post-nominal prepositional phrases neither favored nor disfavored SD. She did not examine the effect of post nominal modifiers.

For the sake of comparability, the coding for subject complexity for this study was adapted from Auger and Villeneuve (2010). Subjects were divided into NPs containing one of the 9 following possibilities: a post-nominal adjective, an apposition, coordination, a prepositional complement, a relative clause, any pre-nominal modifier, a post-nominal non-adjectival modifier, more than one complement, or no complement.

3.5 Intervening Elements

Intervening elements between the NP and the verb have also been observed to influence SD, potentially indicating a processing constraint. Nagy et al. (2003) found that all intervening elements favor SD, whereas their absence slightly disfavor it. Auger and Villeneuve (2010) demonstrated the same relationship. Consequently, all intervening elements will be coded according to the following classifications: adverbs (*aussi* 'also'); hesitations (*euh*); parentheticals *(ben)*; emphatic pronoun (*mon frère, lui* 'my brother, he'); clause (*Et ma mère, en me voyant, m'a dit* 'And my mother, upon seeing me, said to me'); *oui* 'yes' or *non* 'no'; more than one element; and absence of intervening elements.

3.6 Clause Type

Clause type will be coded following Auger and Villeneuve (2010): main, relative, subordinate introduced by *que* 'that' or *si* 'if', conditional, adverbial and other. Auger and Villeneuve demonstrated that subject doubling is slightly favored by main clauses while all subordinates disfavor SD to varying degrees. Nagy et al. (2003) provided corroborating results although their classification of clause type differed to some extent.

3.7 Verb Type

Verb type has been observed across studies to similarly affect SD in different varieties of French. Passive verbs disfavor SD as well as unaccusative verbs. On the other hand, unergatives and transitives favor SD (Nagy et al. 2003, Auger and Villeneuve 2010), as well as copula (Auger and Villeneuve, 2010). Verbs were classified into the following categories: copula, transitives, unergatives, unaccusatives, passives, pronominal verbs, modals, causal verbs, and verbs that take prepositional complements (e.g., *commencer à* 'to begin to'). Additionally, non-modal uses of *aller* 'to go'*, avoir* 'to have'*, faire* 'to do' and *pouvoir* 'to be able to' were coded as their own category due to their high frequency of use.

3.8 Polarity

Several studies have shown an inverse relationship between SD and the presence of the negative marker *ne* (Nadasdi 1995, Coveney 2003, Nagy et al. 2003; among others). Higher *ne*-retention as well as lower rates of SD are considered characteristic of formal speech, and therefore correspond. That is, higher rates of SD tend to co-occur with lower rates of SD. In order to test the connection between the two variables, the tokens were coded as either affirmative, negative with *ne* retention, or negative with *ne* deletion.

3.9 Information Status

As mentioned in the section on background information, information status has been shown to affect SD: subjects introducing new information are favorable whereas previously mentioned subjects are not. This has been interpreted as a relationship between focus and SD (Barnes 1985). Consequently, the same distinction was made in the coding for this study.

3.10 Social Factors

Age and sex were considered. Age was divided into three categories: 15-30, 31-55, 56+. Sex was divided into male and female. Contradictory results have been found regarding sex and age as determining factors: Ashby (1980) found that men and the middle age group favored SD, while women, youth and the oldest age group were unfavorable. Auger and Villeneuve (2010) exhibited completely different results: men and women produced similar rates of doubled subjects, while young speakers favor SD. However, when combining age and sex, young women favored doubled subjects more than all other groups.

4 Methodology

4.1 Corpus

All eligible tokens (cf. 4.2) were extracted from the *Corpus de Français Parlé Parisien des Années 2000* (CFPP2000), an oral corpus of 30 sociolinguistic interviews from thirteen Parisian neighborhoods (Branca-Rosoff, Fleury, Lefeuvre, and Pires 2012). Seventeen of the interviews from all thirteen neighborhoods featured in the corpus were used for this study.

4.2 Variable Context

All tokens containing a third person preverbal nominal subject that could be doubled were extracted: full NPs (2a); proper names (2b); strong pronouns such as *lui* 'he' (2c); and other pronouns such as *quelqu'un* 'someone'. While strong and weak *elle* differ in some varieties of French, both forms are typically pronounced [ɛl] in Parisian French and thus excluded. Additionally, tokens containing a post-verbal nominal subject were considered right-dislocated and were not included in the variable context.

(2) a. **Les résidences** ça s'fait dans les hôpitaux.
    'Residencies are done in hospitals.'
  b. Cette année, **Magalie** va à Lafontaine.
    'This year, Magalie is going to Lafontaine.'
  c. Mais **eux-autres** s'en rendent pas compte.
    'But they do not realize it.' (cf. Auger and Villeneuve 2010)

Tokens containing only a clitic pronoun were not considered part of the variable context, following the methodology employed in previous studies (Nagy et al. 2003, Auger and Villeneuve 2010). As such, hypothetical sentences such as (3a-b) were included, while sentences such as (3c) were not considered.

(3) a. **Pierre il** est professeur.
    'Pierre is a professor.'
  b. **Pierre** est professeur.
    'Pierre is a professor.'
  c. **Il** est professeur
    'He is a professor.'

Lastly, only 3rd person was considered since the only NPs that can co-occur with 1st or 2nd person subject clitics are the strong pronouns *moi* 'I/me', *toi* 'you sing.', *vous* 'you pl.' and *nous* 'we'. These strong pronouns cannot stand alone and are always used with their corresponding subject clitic. Consequently, they do not present variation (Nadasdi 1995, Nagy et al. 2003, Auger and Villeneuve 2010)

1097 tokens of eligible doubled and non-doubled subjects were extracted and coded according to the factors outlined above. The data were then analyzed with Goldvarb X. The results of the statistical analysis are presented in the following section.

5 Results

5.1 Linguistic Factors

The overall rate of SD in the CFPP2000 corpus is 22%. Table 2 shows the results of the statistical analysis for the linguistic factors that proved statistically significant. The application value is SD, such that weights above 0.50 indicate that a factor favors SD, while weights below 0.50 signify that a factor is unfavorable. 1060 tokens were used in the final statistical analysis since passive sentences were categorically not doubled (37 tokens). Also, appositions were excluded due to their very low frequency in the data (7 tokens). Consequently, these tokens were excluded.

Additionally, only eight of the nine linguistic factors were included in the final Goldvarb analysis due to interactions between subject type and subject animacy, definiteness, and specificity: strong pronouns are always animate, [+definite] and [+specific]. Thus, subject type was not considered. Of the eight factors that were included in the analysis, seven were statistically significant. The factor combining definiteness and specificity was not significant.

Polarity is the factor group with the highest range of effect (62). Importantly, while affirmative sentences neither favor nor disfavor SD, *ne* retention strongly disfavors SD. Only 2 of 74 tokens containing *ne* also contained a doubled subject. On the other hand, *ne* deletion strongly favors SD. These results coincide with previous studies and are not surprising. SD is often considered characteristic of colloquial speech (Sankoff 1982) and the link between *ne* retention and standard or formal varieties of French is well-attested in the sociolinguistic literature (Ashby 1981,

Coveney 1996, Armstrong 2001).

Subject complexity had a strong effect, with a range of 47. A general trend can be discerned based on the results: more complex subjects favor SD, such as post-nominal modifiers, relative clauses, multiple complements and prepositional complements. Conversely, prenominal modifiers, and coordination disfavor SD. This could be explained in terms of non-nominal material intervening between the head noun and the verb: NPs with prenominal modifiers and coordination do not have such intervening material and therefore, may not cause processing difficulties, as was hypothesized.

Factor	N	%SD	Weight
Polarity			
Negative w/out *ne*	24/66	36	0.71
Affirmative	205/913	23	0.53
Negative with *ne*	2/74	3	0.09
Range			*62*
Subject complexity			
Adj./ post nom. modifier	11/22	50	0.77
Relative clause/ multiple elements	21/53	40	0.71
Prepositional complement	16/67	24	0.60
No intervening element	180/873	21	0.48
Prenom. modifier./ coordination.	3/38	8	0.27
Range			*50*
Intervening Elements			
Emphatic./paren./*oui*/clause	22/40	55	0.82
Hesitation, adv., multiple el.	22/71	31	0.59
None	187/942	20	0.48
Range			*34*
Clause type			
Main	192/717	27	.58
Adverb	18/127	14	.40
Relative, *que/si*, prep, cond.	21/209	10	.30
Range			*28*
Grammatical Person			
Singular	173/592	29	0.62
Plural	58/460	13	0.35
Range			*27*
Verb type			
Avoir, aller, faire, pouvoir	50/171	29	0.64
Intransitive	46/191	24	0.55
Transitive, pronominal	56/249	23	0.51
Copular, modal, prep.,	79/441	18	0.41
Range			*23*
Information status			
Already mentioned	102/347	29	0.61
New referent	130/706	18	0.45
Range			*16*

p< 0.05, Total N: 231/1053, Total chi-squared= 246.7724
$\chi2$/cell= 0.7809, Log likelihood= -465.447

Input= 0.17 (22%)

Table 2: Linguistic factors affecting the use of subject doubling.

Sentences with elements intervening between the NP and the verb comprise 11% of the total data. All intervening elements favor SD: emphatic pronouns, parentheticals, *oui* 'yes', *non* 'no' and clauses demonstrate the highest rate of doubled subjects, while hesitations, adverbs and multiple elements slightly favor SD. The absence of an intervening element neither favors nor disfavors SD. This result indicates that there may be a processing effect of intervening elements that causes doubled subjects to be more frequent.

Results for the effect of clause type on SD correspond with those of previous research. As in Nagy et al. (2003) and Auger and Villeneuve (2010), main clauses favor SD, while all other subordinates are unfavorable. More specifically, relatives and conditional clauses strongly disfavor SD as in previous studies.

Verb type diverges from results found in previous studies: all intransitive verbs slightly favor SD and therefore were combined in the statistical analysis, whereas in Nagy et al. (2003) and Auger and Villeneuve (2010) unergative and unaccusative verbs, two types of intransitives, behaved differently: unergatives favored SD, while unaccusatives disfavored it. Thus, this distinction does not apply in Parisian French. Additionally, transitives favored SD in prior research but are neither favorable nor unfavorable in Parisian French (weight=0.51). Nevertheless, some similarities do exist between verb type in this variety of French and other varieties: copula disfavors SD and passives disfavor doubled subjects, categorically in the present study and variably so in other varieties. Lastly, the four most frequent verbs in the corpus, *avoir, aller, faire,* and *pouvoir*, favored SD the most. This result suggests that it may prove fruitful to consider frequency as a factor in future research on subject doubling. The implications of the importance of frequency will be addressed in the discussion section.

The final significant linguistic factor is information status. In this study, previously mentioned referents favor SD, while new referents are slightly unfavorable. These results differ from those found in Barnes (1985), who observed the reverse effect in the same variety of French. This dissimilarity may be attributable to differences in interpretation of 'new referent to the discourse' or to diachronic change.

Subject type	#tokens	%SD	%Total data
Strong pronouns	99	70%	12%
Proper nouns	35	32%	9%
Common nouns/ other pronouns	188	20%	79%

Table 3: Rate of SD according to subject type.

Although subject type was not included in the final statistical analysis, Table 3 reveals that the frequency of SD varies considerably across types. The results correspond to those found in Nadasdi (1995), Nagy et al. (2003) and Auger and Villeneuve (2010), displaying the same hierarchy of subject type in terms of SD rate: strong pronouns > proper nouns > common nouns/other pronouns. Specifically, 70% of strong pronouns occur with a doubled subject, suggesting a similarity to or potentially an effect of the categorical use of SD with 1st and 2nd person strong pronouns.

5.2 Extralinguistic Factors

Table 5 presents the statistical analysis of age and sex. Data on the speaker's sex and age was only available for 763 tokens due to corpus restrictions, and as such only these tokens were considered in the Goldvarb analysis. Age proved statistically significant: the oldest age group produced more doubled subjects than any other age group. The middle age group disfavored SD, while young speakers used it the least.

These results differ from those presented in previous studies. Ashby (1980) found that the middle age group favored SD, while youth and the oldest age group disfavored it. Auger and Villeneuve (2010) observed that youth favored SD, with young women producing the most doubled subjects. Since innovative forms are often associated with youth and women, Auger and Villeneuve (2010) suggest that their results may represent ongoing linguistic change. Conversely, the results from the current study indicate a possible decline in SD in spoken Parisian French.

Factor	N	%SD	Weight	%Total
Age				
56-	106	29	0.62	47
31-55	44	14	0.40	41
0-30	10	11	0.38	12
Range			*24*	
Sex				
Female	133	24	0.54	72
Male	27	13	0.39	28
Range			*15*	

p< 0.05, Total N: 160/763, Total chi-squared= 2.3974
χ2/cell= 0.3996, Log likelihood= -373.552
Input= .20 (21%)

Table 4: Extralinguistic factors affecting subject doubling.

There was a significant overall effect of gender in this study, unlike the results found in Auger and Villeneuve (2010). Women favor subject doubling (0.54), while men disfavor it (0.39). However, after combining sex and gender in the analysis, it becomes evident that these two factors act together in affecting SD: older women double their subjects the most (31%) while young men rarely produce SD (5%).

Table 6 presents a cross-tabulation of sex and age. Upon closer examination, the youngest age group has the smallest number of tokens. Consequently, it is possible that a larger group of young speakers would diverge less from the norm. Additionally, the overall rate of SD by young speakers is lower due to the young male group, while young women produce a rate of SD that is identical to the overall rate (21%). As such, SD use is not declining among all young speakers, and the trend observed above with regards to the effect of age is not as clear when gender is also considered. Furthermore, given that the speaker's schooling, socioeconomic status and neighborhood was not measured, it is unclear whether the sex and age groups are comparable.

	female		male		Σ	
	%SD	#tokens	%SD	#tokens	%SD	#tokens
56+	**31**	290	**21**	72	**29**	362
31-55	**15**	230	**11**	80	**14**	310
0-30	**21**	33	**5**	58	**11**	91
Σ	**24**	553	**13**	210	**21**	763

Table 6: Cross-tabulation of sex and age.

6 Discussion

6.1 Characterizing SD in Parisian French

The results presented in the previous section reveal several notable characteristics of the variety of French employed in this study. Firstly, in this study, as found in prior research on the subject, SD

shares an inverse relationship with *ne* retention, a characteristic of standard or more formal varieties of French. The low rate of SD in comparison to other varieties (22%) coupled with the high rate of *ne* retention in negative sentences (53%) indicate that the speech from the CFPP2000 corpus may be relatively formal or standard. This formality may be due to the topic of discussion of the corpus' interviews: Paris and its suburbs. The Parisian dialect is often viewed as the standard within France. It is possible that discussing his/her neighborhood or city causes the speaker to self-identify with Paris and consequently produce his/her perception of what a typical Parisian dialect is, i.e., a variety of French containing various formal characteristics, such as *ne* retention and lower rates of SD.

Secondly, in addition to the effect of *ne* retention on SD, Parisian French shares many similarities to other dialects previously investigated. The first resemblance pertains to clause type: all subordinates disfavor SD while main clauses present the highest rate of doubled subjects. Additionally, passive verbs and copula disfavor SD. Also, although not considered in the final probabilistic analysis, rates of SD according to subject type pattern identically to those of previous research. These consistencies are observed across dialects and studies and may be indicative of more general constraints on SD.

Furthermore, two other parallels between Parisian French and other varieties point to a processing constraint on SD across dialects of French. Firstly, more complex subjects favor SD. Additionally, all intervening elements between the NP and the verb favor SD. These effects are corroborated in all prior research that considered these factors. Nevertheless, a processing constraint cannot be the sole explanatory factor for SD in Parisian French, since sentences with no intervening material between the subject NP and the verb, and subject NPs without any complement do not disfavor subject doubling, as demonstrated by weights that approach 0.50 (0.48 for both factors).

Nevertheless, Parisian French differs with regard to various factors. The key difference pertains to factors related to strong agreement: while animacy, definiteness, and specificity have been shown to favor SD in other varieties, in Parisian French none of these variables statistically constrain SD use. This outcome is unexpected since strong agreement is one of the most frequent explanations of SD variation (Auger 1996). Consequently, it appears that doubled subjects do not function as an indicator of strong agreement in the spoken Parisian dialect. Other differences include the effects of information status, plurality, the distinction between unaccusative and unergative verbs and social patterning. Specifically, the extremely low rate of SD among young male speakers contrasts with previous studies.

6.2 Frequency

This study is the first to separate *aller, avoir, faire* and *pouvoir*, four frequent verbs in French, when coding for verb type. This category favored SD the most, indicating a possible frequency effect on SD that has not been previously researched. Frequency effects may have implications in the debate on whether subject clitics are grammaticalizing as morphological markers of verbal agreement for person and number in different varieties of French (Auger 1993, 1994, 1995, Roberge 1990, etc.), since frequency is widely accepted as playing an important role in grammaticalization (Bybee 2003, Haiman 1994, etc.). While this debate was not discussed in this paper, future research could examine frequency effects on SD as a potential argument for affixation of subject clitics.

7 Conclusion

This paper is the first quantitative variationist study of SD in Parisian French to include a variety of linguistic and extralinguistic factors in its analysis. The results reveal that Parisian French shares several similarities with other varieties of French that help to characterize SD as a trait of less formal/colloquial speech and as an effect of possible processing constraints. Nevertheless, this investigation also demonstrates that, although several theoretical accounts consider SD a manifestation of strong agreement between the verb and the noun, this function cannot be attributed to SD in Parisian French. Lastly, the results suggest that frequency effects, a factor not previously studied, may prove important in SD variation, and thusly, should be included in future studies of SD variation.

References

Armstrong, Nigel. 2001. *Social and Stylistic Variation in Spoken French: A Comparative Approach*. Amsterdam: John Benjamins.

Ashby, William. 1980. Prefixed conjugation in Parisian French. In *Italic and Romance: Linguistic Studies in Honor of Ernst Pulgram*, ed. H. J. Izzo, 195-207. Amsterdam: John Benjamins.

Ashby, William. 1981. The loss of the negative particle ne in French: a syntactic change in progress. *Language* 57:674–687.

Auger, Julie. 1993. More evidence for verbal agreement-marking in colloquial French. In *Linguistic Perspectives on the Romance Languages*, eds. W. J. Ashby, M. Mithun, G. Perissinotto and E. Raposo, 177–197. Amsterdam: John Benjamins.

Auger, Julie. 1994. Pronominal Clitics in Colloquial Québec French: A Morphological Approach. Doctoral dissertation, University of Philadelphia.

Auger, Julie. 1995. A morphological analysis of Quebec Colloquial French pronominal clitics. In *CLS 31-II: Papers from the Parasession on Clitics*, eds. A. Dainora et al., 32– 49.

Auger, Julie. 1996. Variation data and linguistic theory: Grammatical agreement and subject doubling. In A. *Proceedings of the Thirteenth Eastern States Conference on Linguistics '9,* eds. D. Green & V. Motapanyane, 1– 11.

Auger, Julie and Anne-José Villeneuve. 2010. La double expression des sujets en français saguenéen. In *Hétérogénéité et Homogénéité dans les Pratiques Langagières : Mélanges Offerts à Denise Deshaies*, eds. W. Remysen and D. Vincent, 67– 68. Québec : Presse d l'université Laval.

Barlow, Michael. 1988. A Situated Theory of Agreement. Doctoral dissertation, Stanford University.

Barnes, Betsy K. 1985. *The Pragmatics of Left-Detachment in Spoken Standard French*. Amsterdam: John Benjamins.

Branca-Rosoff, Sonia, Serge Fleury, Florence Lefeuvre, and Mat Pires. 2012. *Discours sur la ville : Corpus de Français Parlé Parisien des années 2000 (CFPP2000)*. Corpus, Université Paris 3 Sorbonne Nouvelle. URL `http://cfpp2000.univ-paris3.fr/`

Bybee, Joan. 2003. Mechanisms of change in grammaticization: The role of frequency. In *Handbook of Historical Linguistics*, eds. B. Joseph and R. Janda, 602-623. Oxford: Blackwell Publishers.

Campion, Elizabeth. 1984. Left Dislocation in Montréal French. Doctoral dissertation, University of Pennsylvania.

Coveney, Aidan. 1996. Variability in Spoken French: A Sociolinguistic Study of Interrogation and Negation. Exeter: Elm Bank.

Coveney, Aidan. 2003. Le redoublement du sujet en français parlé: Une approche variationniste. In *Structures Linguistiques et Interactionnelles du Français Parlé*, eds. A. B. Hansen and M. Hansen, 111–143. Copenhague: Museum Tusculanum Press.

Coveney, Aidan. 2005. Subject doubling in spoken French: A sociolinguistic approach. *The French Review* 79:96– 111.

Haiman, John. 1994. Ritualization and the development of language. In *Perspectives on Grammaticalization*, ed. W. Pagliuca, 3– 28. Amsterdam: John Benjamins.

Harris, Martin. 1978. *The Evolution of French Syntax*. London: Longman.

King, Ruth and Terry Nadasdi. 1997. Left dislocation, number marking and Canadian French. *Probus* 9:267–284.

Marácz, László. 1987. On the status of the Projection Principle in Hungarian grammar. In *Approaches to Hungarian; Volume 2*, ed. I. Kensei, 87–112. Szeged: Jate Szeged.

Nadasdi, Terry. 1995. Subject NP doubling, matching, and minority French. *Language Variation and Change* 7:1– 14.

Nagy, Naomi, Hélène Blondeau, and Julie Auger. 2003. Second language acquisition and "real" French: An investigation of subject doubling in the French of Montreal Anglophones. *Language Variation and Change* 15:731–03.

Pooley, Timothy. 1996. *Chtimi: The Urban Vernaculars of Northern France*. Clevedon: Multilingual Matters.

Roberge, Yves. 1990. *The Syntactic Recoverability of Null Arguments*. Kingston: McGill-Queen's University Press.

Sankoff, Gillian. 1982. Usage linguistique et grammaticalisation: Les clitiques sujets en français. In *La Sociolinguistique dans les Pays de Langue Romane*, eds. N. Dittmar and B. Schlieben-Lange, 81–85. Tübingen: G. Narr.

Thibault, Pierrette. 1983. Equivalence et Grammaticalisation. Doctoral dissertation, Université de Montréal.

Vasseur, Gaston. 1996. *Grammaire des Parlers Picards du Vimeu (Somme) : Avec Considération Spéciale du Dialecte de Nibas*. Abbeville, France : F. Paillart.

Wald, Benji. 1979. The development of the Swahili object marker: a study of the interaction of syntax and discourse. In *Syntax and Semantics, vol. 12: Discourse and Syntax,* ed. T. Givón, 505-24. New York: Academic Press

Department of Spanish and Portuguese
Indiana University
Bloomington, IN 47405-7103
szahler@indiana.edu