

À Paris/sur Paris: A Variationist Account of Prepositional Alternation
Before City Names in Hexagonal French

by

Emmanuelle Buillon

B.A., Université Sorbonne Nouvelle, 2018

A Thesis Submitted in Partial Fulfillment of the
Requirements for the Degree of

MASTER OF ARTS

in the Department of Linguistics

©Emmanuelle Buillon, 2021
University of Victoria

All rights reserved. This thesis may not be reproduced in whole or in part, by photocopy or other means, without the permission of the author.

We acknowledge with respect the Lekwungen peoples on whose traditional territory the university stands and the Songhees, Esquimalt and WSÁNEĆ peoples whose historical relationships with the land continue to this day.

À Paris/sur Paris: A Variationist Account of Prepositional Alternation
Before City Names in Hexagonal French

by

Emmanuelle Buillon
B.A., Université Sorbonne Nouvelle, 2018

Supervisory Committee

Dr. Alexandra D'Arcy, Supervisor
Department of Linguistics

Dr. Catherine Léger, Member
Department of French

Abstract

À is the prototypical preposition used before city names in French, yet there are reports that, since the mid-20th century, *sur* also appears in this context in vernacular varieties of European French. To date, research on the choice between *à* and *sur* has focused on semantic and pragmatic differences between the two, and has relied on made-up examples, small participant usage surveys, or empirical datasets that were not systematically analyzed. Moreover, the influence of social factors has received only scant attention. This thesis addresses these shortcomings by providing a quantitative, variationist and longitudinal account of *à/sur* alternations. It asks the following question: Which factors (linguistic and social) can account for prepositional variation before city names in Hexagonal French? The data was drawn from two publicly available corpora of spoken Hexagonal French, representing three locales: the Parisian city-centre, a group of suburban cities surrounding Paris, and the midsize provincial city of Orléans. The speakers (N = 151) were born between 1878 and 1994, providing a mixture of real- and apparent-time perspectives on variation. Following variationist methods, the analysis considers all contexts where *à/sur* variation is possible (N = 2542) and seeks to elucidate the variable grammar. Results indicate important differences between the three areas under study, both in terms of social patterns and in terms of linguistic constraints. In Paris, the use of *sur* is restricted to a few speakers and a few linguistic contexts, and is overall very infrequent. In the suburbs and Orléans, *sur* is more widely attested across speakers and contexts, but it remains a minority variant which seems to be on the decline, especially in the suburbs. Further, in Orléans, the variable grammar is less linguistically constrained, suggesting a trajectory of geographic diffusion. Overall, the quantitative findings support some of the semantic and pragmatic hypotheses proposed in earlier work while shedding light on how geographic, social and linguistic factors combine to explain a phenomenon of variation that has never been studied using variationist methods before.

Table of Contents

Supervisory Committee.....	ii
Abstract	iii
Table of Contents	iv
List of Tables.....	vi
List of Figures	viii
Acknowledgements	ix
Chapter 1. Introduction	1
Chapter 2. Literature review.....	7
2.1 <i>Sur</i> as a spatial preposition in the semantic literature.....	7
2.2 <i>Dans</i> as a spatial preposition in the semantic literature.....	11
2.3 The original locative function of <i>à</i>	14
2.4 <i>Sur</i> + city name: Semantic and pragmatic explanations.....	16
2.5 <i>Sur</i> + city name: Other potential factors of variation	20
Chapter 3. Methods	24
3.1 Data.....	24
3.1.1 Corpora.....	24
3.1.2 Speaker sample.....	26
3.2 Data extraction.....	27
3.2.1 Setting up the variable context	27
3.2.2 A reconsideration of <i>dans</i>	30
3.3 Coding.....	33
3.3.1 Social predictors	34
3.3.2 Linguistic predictors.....	38
Chapter 4. Results	43
4.1 Overall distribution for the combined corpora	43
4.2 Distributional results for the Parisian data.....	43
4.2.1 Overall distribution and longitudinal trajectory	43
4.2.2 Social factors	48
4.2.3 Linguistic factors.....	52
4.2.4 Summary of the distributional results for the CFPP2000 data.....	59

4.3 Distributional results for the ESLO data	60
4.3.1 Overall distribution and longitudinal trajectory	60
4.3.2 Social factors	63
4.3.3 Linguistic factors	66
4.3.4 Summary of the distributional results for the ESLO data	69
4.3.5 Summary of the distributional findings.....	69
4.4 Inferential statistics: Modelling variation in the PCN context	71
Chapter 5. Discussion.....	76
5.1 Implications of the study	76
5.2 Limitations and future research	81
References	83
Appendix A. Additional tables and figures	90
Appendix B. List of city names	93
Appendix C. List of motion verbs	100

List of Tables

Table 1: Speaker sample	26
Table 2: List of predictors and their reference levels	33
Table 3: Scale for occupation type	35
Table 4: Scale for educational background	35
Table 5: Distribution of <i>sur</i> by decade of birth in the CFPP2000 data	44
Table 6: Distribution of <i>sur</i> by decade of birth inside Paris and in the Parisian suburbs	47
Table 7: Distribution of <i>sur</i> by arrondissement inside Paris	48
Table 8: Distribution of <i>sur</i> by speaker's location in the Parisian suburbs.....	49
Table 9: Distribution of <i>sur</i> by gender in the CFPP2000 data	50
Table 10: Distribution of <i>sur</i> by socio-economic status in the CFPP2000 data	51
Table 11: Distribution of <i>sur</i> by type of situation in the CFPP2000 data	53
Table 12: Distribution of <i>sur</i> by type of activity in the CFPP2000 data.....	54
Table 13: Distribution of <i>sur</i> by type of trajectory in the CFPP2000 data	55
Table 14: Distribution of <i>sur</i> by type of verb in the CFPP2000 data.....	55
Table 15: Distribution of <i>sur</i> by decade of birth and type of verb in the Parisian suburbs.....	57
Table 16: Distribution of <i>sur</i> by city name inside Paris.....	57
Table 17: Distribution of <i>sur</i> by city name in the Parisian suburbs	58
Table 18: Distribution of <i>sur</i> by speaker's city and lexical city name in the Parisian suburbs	59
Table 19: Distribution of <i>sur</i> by decade of birth in the ESLO data	61
Table 20: Distribution of <i>sur</i> by speaker and decade of birth in the ESLO data	63
Table 21: Distribution of <i>sur</i> by gender in the ESLO data	63
Table 22: Distribution of <i>sur</i> by decade of birth and gender in the ESLO data.....	65
Table 23: Distribution of <i>sur</i> by socio-economic status in the ESLO data	65
Table 24: Distribution of <i>sur</i> by gender and socio-economic status in the ESLO data	66
Table 25: Distribution of <i>sur</i> by type of situation in the ESLO data	66
Table 26: Distribution of <i>sur</i> by type of activity in the ESLO data	67
Table 27: Distribution of <i>sur</i> by type of trajector in the ESLO data.....	67
Table 28: Distribution of <i>sur</i> by type of verb in the ESLO data	68
Table 29: Distribution of <i>sur</i> by city name in the ESLO data.....	68
Table 30: Summary of the distributional results in the three datasets	70
Table 31: Best mixed-effects model of predictors operating on <i>sur</i> : Paris	72
Table 32: Best mixed-effects model of predictors operating on <i>sur</i> : Parisian suburbs.....	73
Table 33: Best mixed-effects model of predictors operating on <i>sur</i> : Orléans	74

Table 34: Distribution of <i>sur</i> across speakers in all datasets	76
--	----

List of Figures

Figure 1: Map of the locations that were included in the CFPP2000 data.....	38
Figure 2: Distribution of <i>sur</i> by decade of birth in the CFPP2000 data.....	44
Figure 3: Distribution of <i>sur</i> by decade of birth inside Paris and in the Parisian suburbs	46
Figure 4: Distribution of <i>sur</i> by decade of birth and type of verb in the Parisian suburbs	56
Figure 5: Distribution of <i>sur</i> by decade of birth in the ESLO data	60
Figure 6: Distribution of <i>sur</i> by decade of birth and gender in the ESLO data	64

Acknowledgements

My deepest gratitude goes to my supervisor Dr. Alex D'Arcy, who shared invaluable knowledge and feedback with me as I was working on this thesis, as well as supported me with constant enthusiasm and understanding through all the steps, whether it was in person or online. I am also very grateful to her for giving me the opportunity to conduct the bulk of this research (as well as other research projects) at the Sociolinguistics Research Lab. I also wish to thank my second reader Dr. Catherine Léger for her involvement throughout this work, for her helpful comments, and for her precious insights into the prepositional system of French. Many thanks as well to Dr. Mark Lauersdorf for accepting to join the committee as the external examiner.

Thank you to Dr. Derek Denis as well as the attendees at NARNiHS 2021 for their help and suggestions regarding this project.

I am deeply grateful to the Department of Linguistics at UVic for the financial support I received throughout the years. I also want to thank all the faculty at the department, particularly Dr. Ewa Czaykowska-Higgins, who accompanied me through the beginning stages of my program, as well as Dr. Judith Nylvek and Dr. Dave McKercher for giving me the opportunity to present some of my research projects and interests to undergraduate students. Many thanks as well to Jenny, Maureen, and Emmanuelle, who all made me feel welcome and supported from the day I arrived.

Thank you to all the graduate students who I was lucky to meet and study alongside with during my time at UVic, in particular Ayden, Christy, Junyu, Keun, Mitchell, Mona, Rain, Tess, and Yi. Thank you for making this program such a fun experience! The warmest thanks to all my fellow lab members for making this space a supportive and inclusive environment that I am missing very much. I am thinking especially of Chloë, Christian, Kate, Lizzy, Rachel, and Sydney.

Of course, I want to give very special thanks to Caroline for being such an incredible lab manager, research collaborator, fellow student, but more importantly, friend. This time in our lives took quite an unexpected turn, but I am forever grateful that I had you by my side through it all.

Thank you to all my friends in Canada (in particular Lora) for their continued support from the moment I arrived to the moment I moved out. Thank you to my friends and family back in France for encouraging me from afar.

Finally, thank you to my parents and Mathilde for complaining about *sur* all these years ago, and for continuing to do so even as I was working on this project. Thank you for always supporting me despite the distance, and through the more uncertain times—I know this was not always easy.

Chapter 1. Introduction

The preposition *à* is the primary locative preposition used before city names in standard varieties of French, illustrated in (1). In some contexts, however, locative *à* can alternate in the standard language with locative *dans* (Spang-Hanssen, 1963, p. 197); an example is given in (2). As explained in sections 2.2 and 3.2.2, this latter preposition is used in specific semantic contexts emphasizing inwardness and containment. Moreover, it has been reported that a third preposition, namely *sur*, has been in use before city names since the mid-20th century, although the point of origin of the feature could be older than this, as we will see in later sections. The literature points out that the use of *sur* is restricted to spoken, typically vernacular, varieties of European French, such as Swiss, Belgian, and Hexagonal varieties (e.g. Grevisse & Goosse, 1993; Hernández, 2008; Huyghe, 2012; Lismoen, 2015). An example of this non-standard form is provided in (3).

- (1) Mon patron n'est jamais venu À Paris. (Frantext: Alphonse Daudet, *Le Nabab*, 1881)¹
 'My boss never came to Paris.'²
- (2) Ses beaux yeux devinrent fixes et ternes en pensant aux dangers qu'ils couraient
 DANS Paris (Frantext: Honoré de Balzac, *Une ténébreuse affaire*, 1846)
 'Her beautiful eyes stopped moving and became lifeless as she was thinking of the
 threats they were encountering in Paris'

¹ All the examples labelled as *Frantext* are data from the *Frantext* online corpus (ATILF, 1998–2019), a 260-million word public corpus of diachronic French texts, spanning from the 9th century to the 21st century.

² All translations from French are my own.

- (3) Vous venez de province et vous allez SUR Paris ? (Frantext: Michel Tournier, *Le Médianoche amoureux*, 1989)³

‘You’re coming from the provinces and you’re going to Paris?’

The literature suggests that *sur* is not restricted to the locative function, but is attested in other new functions as well (e.g. Cadiot, 1997). For example, in (4), it is durative, in competition with the standard preposition *pendant*, illustrated in (5).

- (4) c’est t-SUR deux jours je crois (ESLO2/RL437, b. 1970s)⁴

‘It’s over the course of two days I think’

- (5) j’ai fait grève PENDANT deux jours (ESLO1/PW85, b. 1930s)

‘I was on strike for two days’

What is unclear is whether the city name context was the starting point for this functional and contextual expansion, or whether this expansion has occurred in parallel across several contexts, such as the locative (3) and the durative (4) ones. Therefore, although the generalization of *sur* across linguistic contexts and functions in vernacular practice is worthy of attention in its own right, this study will focus on the preposition + city name context (henceforth abbreviated as the PCN context), which has received the most attention so far.

Notably, the bulk of this work has focused on the semantic and pragmatic differences between *à* and *sur* but has not considered *dans* as a possible variant, despite the fact that all appear before city names, as illustrated in (1)–(3) (see e.g. Hernández, 2008; Huyghe, 2012). It is also the case that analyses have either relied upon small sets of made-up examples (e.g.

³ This example is part of a dialogue within the novel from which it is excerpted, and is therefore meant to represent a vernacular variety.

⁴ All the examples labelled as *ESLO1* or *ESLO2* are data from the *Enquêtes socio-linguistiques à Orléans* online corpora (Eshkol-Taravella et al., 2012); details are provided in chapter 3. The label that follows the corpus label is the speaker’s identification, while ‘b.’ stands for ‘born’. Finally, with regards to this example, *t-* reflects a performance error from the speaker.

Huyghe 2012; Lismoen, 2015) or large datasets that operationalized raw frequency but did not systematically consider the effects of linguistic factors (e.g. Hernández, 2008). Work to date has also not explored whether social factors such as gender, socio-economic status, or age have an influence on variation between *à* and *sur*. Moreover, prepositional variation in this context is yet to be analyzed using variationist methods, meaning that little is known with respect to the ways in which internal (linguistic) and external (social) predictors combine to affect the probabilistic use of *sur* versus *à* (and possibly *dans*) in the vernacular. As a result, the operation of the variable grammar remains to be elucidated. This thesis therefore seeks to answer the following question: Which factors (linguistic and social) can account for prepositional variation before city names in Hexagonal French?

The major contribution of this study is to operationalize and test previous hypotheses regarding prepositional variation in the PCN context using quantitative variationist methods and a large amount of spoken data. First, it seeks to find which linguistic correlates may be operative. Previous research (e.g. Hernández, 2008; Huyghe, 2012) suggests that there are semantic and pragmatic differences between the use of each preposition in the PCN context. For example, as developed in section 2.4, Hernández (2008) argues that *sur* is used primarily to emphasize temporary situations, while *à* can be used in all contexts, regardless of whether the situation is temporary or not. A central tenet of variationist sociolinguistics is that when forms overlap in function, such differences may be neutralized in discourse, regardless of the linguistic feature under study. That is, speakers alternate between forms in the same context, resulting in a choice mechanism that is probabilistic rather than absolute. For example, in their examination of auxiliaries in Montreal French, Sankoff and Thibault (1977, p. 107) point out that:

“dans les langues naturelles, il n’y a pas de correspondance totale entre la forme et la fonction... [I]l n’y a pas de nécessité à avoir recours à des

auxiliaires différents pour marquer la nuance aspectuelle...[C]ette nuance est souvent marquée par des indices contextuels autres que l’auxiliaire et on trouve même des cas où il ne semble pas pertinent de marquer la distinction.”

[in natural languages, there is no full correspondence between form and function... There is no requirement to use different auxiliaries to express the aspectual nuance... This nuance is often expressed by contextual cues other than the auxiliary itself, and there are even cases where it does not seem relevant to express the distinction.]

Therefore, despite putative semantic and pragmatic differences between the prepositions under study, that they alternate in the PCN context suggests that they are interchangeable, as they share a locative function. This interchangeability is evidenced most concretely not by the variation demonstrated in (1), (2), and (3), but through the occurrence of supertokens (i.e. tokens where two different variants are used by the same speaker within the same conversational turn), as in (6) and (7):

(6) il est pas SUR Ivry i- il est À Vitry si vous voulez (CFPP2000/Assia Berrada, b. 1964)⁵

‘it [the school] is not in Ivry it’s in Vitry you see’

(7) auparavant j’étais SUR Lyon donc euh je suis arrivé à soixante-neuf à À Paris (CFPP2000/Younes Belkacem, b. 1949)⁶

‘before I was in Lyon so um I came to Paris in sixty-nine’

⁵ All examples labelled as *CFPP2000* are from the *Corpus de français parlé parisien des années 2000* online (Branca-Rosoff et al., 2012); details are provided in chapter 3.

⁶ Note that the reduplication of *à* in this example is a performance error from the speaker.

In these examples, the variants “alternate without change in representational meaning” (Poplack, 2011, p. 212), that is, semantic differences are neutralized. However, Poplack (2011) points out that a choice mechanism must still apply for speakers to select one variant over the other “to express the same meaning or function” (p. 213). The main goal of this thesis is to elucidate which factors combine to produce this choice mechanism in the PCN context.

In addition to linguistic factors that may be operative on this mechanism, the study will also reveal whether social factors like gender, age, or socio-economic status have an influence on attested variation. A diachronic perspective also allows us to see how the feature has evolved since the late 19th and early 20th centuries, providing the first step towards the analysis of the spread and potential grammaticalization of *sur* in the vernacular.

Grammaticalization can be defined as “the steps whereby particular items become more grammatical through time” (Hopper & Traugott, 2003, p. 2); prepositions are function words whose origins can often be found in lexical words or constructions (Hopper & Traugott, 2003, p. 4). As such, the spread of *sur* to contexts such as those illustrated in (3), (4), (6), and (7) could represent ongoing grammaticalization, as this function word generalizes across grammatical functions. Indeed, Hopper and Traugott point out that generalization, defined as “a balance between loss of older, typically more concrete, meanings, and development of newer, more abstract ones” (p. 101) is a frequent characteristic in early stages of grammaticalization. We will see in section 2.4 that *sur* evolved from being used in more functional and configurational contexts to more abstract contexts (Hernández, 2008). According to Hopper and Traugott’s (2003, p. 110) classification, this would make *sur* a primary preposition that is currently moving away from its “relatively concrete spatial meaning” and getting more grammatical through a process of generalization. However, *sur* does not seem to be associated (yet) with other properties that characterize grammaticalized

items, such as phonological reduction (Hopper & Traugott, 2003). It is therefore possible that *sur* is simply undergoing an extension in usage that is accompanied by the loss of its original semantic content, but not grammaticalization. The results should allow us to shed more light on this.

Finally, this study will also expand the work on patterns of variation in vernacular Hexagonal French, a variety that, according to Armstrong and Jamin (2002), has led to a “relatively small volume of research on variation and change” (p. 113). Although variationist work on vernacular varieties of Hexagonal French has increased over the last two decades (e.g. Ashby, 2001; Boughton & Pipe, 2020; Marchessou, 2018; Tristram, 2021), they remain less well-documented than variation and change in Laurentian varieties (e.g. Bigot & Papen, 2020 for Ontario French; Blondeau, 2001 for Montreal French; Blondeau et al., 2018 for Montreal and Ontario French; Sankoff & Blondeau, 2007 for Montreal French).

The thesis is structured as follows: first, chapter 2 presents a literature review that discusses how the three spatial prepositions under study have been classically defined in the field of semantics, how their variation in the PCN context has been explained in terms of semantic and pragmatic factors, and what other contextual factors may have a role in this variation. In the methods section, chapter 3, I motivate the value of variationist methods for bringing new light to the variation between *à*, *sur*, and *dans*, outline the corpora that form the empirical basis of my investigation and explain the speaker samples that were used to answer my research question, detail the data extraction process, and outline the factors this research has sought to test. The distributional and inferential results are then presented in chapter 4. Chapter 5 discusses the implications of the results and the limitations of the work, as well as potential ways to expand the research.

Chapter 2. Literature review

Research on prepositional variation in the PCN context is scant, and the majority of existing work has adopted a semantic approach. As such, before discussing this work, I first present a brief overview of the semantic literature on French spatial prepositions, focusing on each of the prepositions under study in this thesis. We will first see how *sur*, the innovative variant in the PCN context, has been described, before turning to the long-established *dans* and *à*. I then turn to the literature dealing with variation in the PCN context.

2.1 *Sur as a spatial preposition in the semantic literature*

Spatial prepositions can be defined as prepositions that answer one of the following questions: “Where?”, “From where?”, and “Through where?” (Lismoen, 2015, p. 7); as well as “To where?”. *Sur*, *à*, and *dans*, as well as *sous* ‘under’, *vers* ‘towards’, *près de* ‘near’, *devant* ‘in front of’, and so on, are French spatial prepositions, as they can all be used in contexts that relate to the concept of place.⁷

Vandeloise (1986) describes French spatial prepositions in terms of relationships between the *cible* and the *site*. The *cible* is defined as “un objet petit et mobile de position inconnue” [a small and mobile object whose location is unknown], whereas the *site* is characterized as “un objet plus grand et plus stable de position connue” [a bigger and more fixed object whose location is known] (p. 20). These two concepts are sometimes referred to in the English literature as *trajector* and *landmark* (e.g. Langacker, 1987), which I will use here. As an example, in (8), the preposition *dans* ‘in’ describes the spatial relationship between the trajector *la table* ‘the table’ and the landmark *la cuisine* ‘the kitchen’.

⁷ However, as pointed out by Cadiot (1997), Melis (2003), and Lismoen (2015), all three prepositions under study here have non-spatial uses as well, such as temporal uses and functional uses. Sections 2.1 through 2.3 focus exclusively on the spatial uses of these prepositions, which are also described as their more concrete uses (Cadiot, 1997).

(8) La table est DANS la cuisine.⁸

‘The table is in the kitchen.’

Vandeloise adopts a functional perspective to describe spatial prepositions. He explains this perspective as follows: “j’essaierai de décrire les mots spatiaux par rapport à des concepts fonctionnels liés à la connaissance extra-linguistique de l’espace que partagent les locuteurs d’une même langue” [I will strive to describe spatial words in terms of functional concepts that are related to the extra-linguistic knowledge of space that is shared by the speakers of a same language] (p. 22). In other words, he argues that spatial prepositions can be described using characteristics that are related to the physical placements of the trajector and the landmark, as they are perceived and experienced by the speaker. For *sur*, he proposes the following four characteristics (pp. 187–193):

- order on the vertical axis (the trajector is higher than the landmark);
- contact (the trajector is in direct or indirect contact with the landmark);
- size of the trajector (the trajector is smaller than the landmark);
- carrier-carriee relationship (the trajector is being carried by the landmark).

Vandeloise describes the relationship in the fourth characteristic in the following words: “Si *a est sur b*, l’action du site s’oppose à l’action de la pesanteur sur la cible.” [if *a is on b*, the landmark’s action counters the effect that gravity exerts on the trajector] (p. 193). A prototypical example with *sur* used as a spatial preposition can be found in (9). In this example, taking the perspective of the speaker who is looking at the scene, all four predicted characteristics of *sur* are met. The trajector (*le livre*) is higher than the landmark (*la table*),

⁸ Unless indicated otherwise, all examples are my own.

they are in direct contact, the trajector is smaller than the landmark, and the landmark carries the trajector.

(9) Le livre est SUR la table.

‘The book is on the table.’

Dendale and De Mulder (1997) expand on Vandeloise’s thesis and propose a few revisions. First, they choose to use the term *traits* “features” instead of *caractéristiques* “characteristics” to describe spatial prepositions. Much like phonological features, each feature can take a positive (+) or negative (–) value, and each preposition can be defined through a specific set of features taking specific values. Describing spatial prepositions in terms of combinations of features allows them to distinguish each preposition from one another. Therefore, if we translate Vandeloise’s characteristics for *sur* into binary features, the preposition could be described as in (10):

(10) [+ higher trajector], [+ contact], [+ smaller trajector], [+ landmark as carrier]

Dendale and De Mulder argue that the last feature (which was introduced in Vandeloise’s later works as the carrier-carriee relationship) is in fact superfluous to describe the uses of *sur*. Indeed, they explain that, in Vandeloise’s theory, this feature is not used to distinguish *sur* from other spatial prepositions, unlike the other features (Dendale & De Mulder, 1997, pp. 216–217). They also point out that combining the already existing features [+ higher trajector] and [+ contact] yields the same result of the landmark carrying the trajector, i.e. the [+ landmark as carrier] feature (p. 217). However, if only the remaining three features are kept, they argue that they are not enough to distinguish *sur* from *dans* in some contexts, such as in examples (11) and (12).

(11) La tasse est SUR l'armoire. (Dendale & De Mulder, 1997, p. 217)

'The cup is on the cupboard.'

(12) La tasse est DANS l'armoire (Dendale & De Mulder, 1997, p. 217)

'The cup is in the cupboard.'

In these examples, where the landmark is the cupboard and the trajector is the cup, *sur* and *dans* can both be described as in (13):

(13) [+ higher trajector], [+ contact], [+ smaller trajector]

Dendale and De Mulder (1997) therefore introduce a new feature, one that can distinguish *sur* from *dans*: the feature [+/- landmark as container]. Indeed, they argue that *sur* is used when the landmark does not contain the trajector, as is the case in (11), contrary to *dans* in (12), where the landmark ('the cupboard') contains the trajector ('the cup').

Therefore, following Vandeloise (1986) and Dendale and De Mulder (1997), the prototypical use of the spatial preposition *sur* can be described with the following features, as shown in (14):

(14) [+ contact], [+ higher trajector], [+ smaller trajector], [- landmark as container]

However, it should be noted that other uses of *sur* follow different combinations of features. For example, in (15), the trajector is perceived to be lower than the landmark by the speaker describing the situation.

(15) La mouche est SUR le plafond. (Dendale & De Mulder, 1997, p. 219)

'The fly is on the ceiling.'

Therefore, in this example, the combination of features for *sur* would be as given in (16):

- (16) [+ contact], [– **higher trajector**], [+ smaller trajector], [– landmark as container]

Such uses of *sur* are argued to be less prototypical by Dendale and De Mulder (1997), and in these contexts there is licit, grammatical variation with other prepositions like *à*, as can be seen in (17), where *au* is a contraction of the preposition *à* and the definite article *le*, or *contre*, as can be seen in (18). As a result, the PCN context is not the only spatial context where variation between *sur* and another preposition is permitted.

- (17) La mouche est AU plafond.

‘The fly is on the ceiling.’

- (18) Le tableau est SUR/CONTRE le mur. (Dendale & De Mulder, 1997, p. 219)

‘The painting is on the wall.’

The literature has therefore covered certain semantic characteristics that can be expected when *sur* is used prototypically as a spatial preposition. The innovative use of *sur* before city names has also been studied, but before moving on to this I will first discuss the semantic literature on *dans* and *à*, the other prepositions with which *sur* alternates in the PCN context.

2.2 Dans as a spatial preposition in the semantic literature

Vandeloise’s (1986) functional definition of the preposition *dans* used in spatial contexts is as follows: “a est dans... b si le site et la cible sont ... le premier et le deuxième élément de la relation contenant/contenu” [a is in... b if the landmark and the trajector are... the first and the second component of the container/content relationship] (p. 222). One of the main features of this container/content relationship is the fact that the landmark completely or partially includes the trajector, as seen in (19) and (20).

(19) les bijoux sont DANS le coffre-fort (Vandeloise, 1986, p. 210)

‘the jewels are in the safe’

(20) l’œuf est DANS le coquetier (Vandeloise, 1986, p. 213)

‘the egg is in the egg cup’

Another feature of this relationship is that, in some cases, the container exerts a force on the content, as can be seen in (21).

(21) le poisson est DANS la main (Vandeloise, 1986, p. 216)

‘the fish is in the hand’

In a posthumous article, Vandeloise (2017) provides a more extensive list of the features characterizing the container/content relationship:

- (a) The position of the content relative to the container does not change when the container is moving
- (b) The container prevents the content from transgressing its limits in more than one direction
- (c) Before containment, the content moves toward the container rather than the reverse
- (d) The container envelops the content
- (e) The container protects the content
- (f) The container hides the content

(Vandeloise, 2017, p. 11)

It should be noted that some uses of *dans* do not satisfy each one of these features. For example, the relationship between the trajector and the landmark in (22) satisfies only features (a), (b), and (d).

(22) La louche est DANS la soupe (Vandeloise, 2017, p. 11)

‘The ladle is in the soup’

Notably, Vandeloise points out that, due to features (a), (b), and (d), “the notion of container can extend progressively to the notion of place. English goes further in this extension than French. Localization has also some similarities with containment because an entity is affected by the place in which it is located.” (p. 12). This would explain why *dans* is used before city names in some contexts in French as well, as in (23).

(23) donc moi j’ai mon petit parcours je cours DANS Saint-Ouen j’ai mes repères tout ça
(CFPP2000/Catherine Lecuyer, b. 1973)

‘so I have my little route I run in Saint-Ouen I have my points of reference
and everything’

Spang-Hanssen (1963, p. 197) also points out that *dans* is used before city names to emphasize inwardness or when considering the extent of the city, as seen in (24) and (25).

(24) ça ressemblait au Rex, ce grand cinéma DANS Paris (par opposition à *dans la banlieue*), *Saints*, *Cesbron* (quoted by Spang-Hanssen, 1963, p. 198)

‘it looked like the Rex, this big movie theater inside Paris’ (as opposed to *in the suburbs*)

- (25) Aurélien... était capable de tourner ainsi des heures et des heures DANS Paris,
Aurélien, Aragon (quoted by Spang-Hanssen, 1963, p. 198)
 ‘Aurélien... was capable of wandering like this for hours on end inside Paris’

Therefore, *dans* seems to be used in contexts emphasizing inwardness and containment.

I now turn to the last preposition that alternates in the PCN context in Hexagonal French: the traditional, standard form, *à*.

2.3 The original locative function of à

Cervoni (1991, p. 5) points out that *à* is one of the most abstract prepositions in French (along with *de*), and that it is therefore difficult to interpret semantically. Goyens et al. (2002) add that “il n’existe pas de schème sous-jacent à la totalité des emplois de la préposition *à*” [there is no underlying framework that explains the entirety of the uses of the preposition *à*] (p. 278). They argue that this diversity of uses is due to the grammaticalization of *à*, leading it to become increasingly abstract as its uses have continuously expanded since the Old French period (pp. 275–277). Goyens et al. (2002) show that *à* has developed etymologically from three Latin prepositions: *ad*, *ab*, and *apud*, which progressively fused into one via grammaticalization. They argue that the primary origin of *à*, however, is *ad*, which was initially used to locate an object in directional contexts before expanding to positional contexts in Vulgar Latin, as shown in the examples in (26).

- (26) a ire AD urbem (Goyens et al., 2002, p. 279)
 ‘to go to the city’
 b esse AD urbem
 ‘to be in the city’

This historical use of *ad* is parallel to the locative use of *à* today in both directional and positional contexts, as seen in (27).

- (27) a Je vais À Paris la semaine prochaine.
 ‘I’m going to Paris next week.’
 b Je suis À Paris cette semaine.
 ‘I’m in Paris this week.’

As early as the Old French period, defined by Guiraud (1975) as the mid-9th century to the early 14th century, the preposition *à* itself grammaticalized and began performing new functions, such as introducing indirect objects. A modern example of this is given in (28).

- (28) J’ai donné ce livre À Lou.
 ‘I gave this book to Lou.’

Goyens et al. (2002) hypothesize that the initial locative function of *à* could become increasingly infrequent in the future: “le stade initial dont le processus de grammaticalisation serait parti (l’emploi locatif directionnel de *à*), qui est bien sûr encore attesté en français actuel, serait condamné à devenir de plus en plus marginal. C’est dans ce sens que nous parlons de ‘déplacement’ de *à* : le rôle basique de *à* ne coïnciderait plus avec son sens premier, locatif” [the initial stage from which the grammaticalization process may have started (the locative directional use of *à*), which is of course still attested in contemporary French, would be condemned to become more and more marginal. It is in that sense that we refer to a shifting of *à*: the basic role of *à* would no longer correspond to its original locative function] (p. 278). This prediction could explain why we now see variation between the long-established preposition *à* with *dans* and *sur* in locative contexts, especially the city name one,

as shown in (29). In this made-up example, the prepositions can be used interchangeably, albeit with small semantic nuances, as was explained in sections 2.1 through 2.3.

(29) Je vais À/DANS/SUR Paris tous les weekends.

‘I’m going to Paris every weekend.’

Therefore, to this day, *à* seems to be the prototypical preposition in use before city names (cf. (27)), as it corresponds to its original locative function, dating back to Latin (cf. (26)).

However, as we have seen in the introduction, it can now alternate with *dans* and *sur* in this context in some varieties of Hexagonal French, and it is this variation that will be the focus of this study. I now move on to discuss existing studies that have sought to explain this variation.

2.4 Sur + city name: Semantic and pragmatic explanations

Hernández (2008) studied the variation of *à* and *sur* before city names, drawing on semantic nuances as an explanatory mechanism. Following Cadiot’s (1997) classification of prepositions into “coloured” (i.e. more concrete, or lexical), “mixed”, and “colourless” (i.e. more abstract, or grammatical) categories, she argues that *à* is a colourless preposition which conveys mostly abstract meanings (cf. Cervoni, 1991), while *sur* can be classified as mixed because it can carry both concrete and abstract meanings. She describes more concrete uses of *sur*, as seen in (9) and repeated here as (30), as cases of a configuration in which the preposition encodes descriptive information on the layout of the scene. Examples like (29), repeated here as (31), on the other hand, can be described as localization, defined by Hernández (2008, p. 2) as an operation of abstract tracking that reflects the landmark’s specificity and the shared geographic knowledge of the speakers. Considering this extension of *sur* from configurational to localizing uses, she therefore argues that *sur* is becoming less concrete and is losing its configurational specificity (p. 6).

(30) Le livre est SUR la table.

‘The book is on the table.’

(31) Je vais SUR Paris la semaine prochaine.

‘I’m going to Paris next week.’

To further explore the differences between the two prepositions, she analyzed a corpus of classified ads published on Yahoo France in March 2006. She focused on the use of *à Paris* versus *sur Paris*, preceded by the following syntagms:

(32) Je suis ‘I am’

J’habite ‘I live’

Je travaille ‘I work’ À/SUR Paris ‘in Paris’

Je reste ‘I stay’

Je dors ‘I sleep’

As her search returned more than 130,000 tokens, Hernández focused on the raw frequencies of each context under study. Her results suggest that *sur* is more likely to appear in contexts where the speaker is referring to a temporary situation, as in (33a), in contrast with contexts referring to a more permanent situation, as in (33b). On the other hand, *à* appears in both temporary and permanent situations, as seen in the examples in (34), without seeming to favour one context over the other.

(33) a. Je suis SUR Paris pour faire un stage de trois mois. (Hernández, 2008, p. 10)

‘I am in Paris for a three-month internship.’

b. ?J’ai toujours vécu SUR Paris. (Hernández, 2008, p. 11)

‘I’ve always lived in Paris.’

- (34) a. Je suis né À Paris mais je suis vite parti. (Hernández, 2008, p. 11)
 ‘I was born in Paris but I quickly left.’
- b. J’ai toujours vécu À Paris. (Hernández, 2008, p. 11)
 ‘I’ve always lived in Paris.’

Hernández argues that using *sur* allows the speaker to emphasize temporary states, and therefore to present themselves as being mobile, reactive, and independent (p. 14). Similar findings were reported by Lismoen (2015). In order to further test Hernández’s hypothesis, she conducted surveys with 19 native French participants who had to choose between three prepositions (*à*, *sur*, and *dans*) in twelve sentences containing city names. In accordance with Hernández, she found that *sur* was more frequent before city names when the context implied a temporary situation or a mobile trajector.

Huyghe (2012) extended the scope of study to the use of *sur* in front of territory names, which include city names but encompass other types of locales as well, as in (35). In this example taken from Huyghe’s article, *sur* is in alternation with *à*. It should be noted, however, that all his examples are generated by him and do not come from empirical data.

- (35) Je suis SUR/À la plage.
 ‘I’m on the beach.’

Huyghe argues that such a use of *sur* is not consistent with the four characteristics presented by Vandeloise (1986), discussed in section 2.1. For example, because of the immateriality of the landmark when it refers to a territory, it is difficult to establish that there is a contact between the landmark and the trajector, or that the landmark carries the trajector. Huyghe therefore proposes that these contexts be considered separately, arguing that the use of *sur* is favoured when the landmark is presented as a “cadre de vie et d’action, à l’échelle de l’homme en particulier” [an environment dedicated to living and performing activities,

especially on a human scale] (p. 180), as seen in (35). Moreover, he argues that a sentence like (36) offers a description of the location that is less neutral than the one in (37), because it suggests and implies that the trajector is performing an activity, like a professional one (p. 181).

(36) Elle est SUR Paris.

‘She’s in Paris.’

(37) Elle est À Paris.

‘She’s in Paris.’

Similarly, the *Trésor de la langue française informatisé* (1994) dictionary points out that *sur* can be used when the following complement is the setting for an activity, as in the bolded text in (38).

(38) Discuter SUR le palier; bronzer, manger SUR la terrasse; piqueniquer SUR les bords de la Moselle; attendre quelqu’un SUR le trottoir; aller en vacances SUR la Côte d’Azur; **travailler SUR le secteur de, SUR Lunéville.** (*Trésor de la langue française informatisé*, 1994, “sur, prép.”)

‘To have a chat in the corridor; to sunbathe, to eat on the patio; to have a picnic by the Moselle; to wait for someone on the sidewalk; to go on vacation on the French Riviera; **to work in the Lunéville area, in Lunéville**’

Following this line of argumentation, *sur* + city name may be constrained by the animacy of the trajectory. Specifically, it would be predicted that *sur* is possible in this context with animate trajectors but that it is either ungrammatical or exceptional with inanimate ones, as in (39).

(39) ?Le parc de la Tête d’Or est SUR Lyon.

‘The Tête d’Or park is in Lyon.’

To summarize, it seems that the preposition *sur* is becoming less concrete as it is moving away from its prototypical configurational use. This semantic shift enables it to generalize to new contexts, such as the PCN one under study here. This extension in usage could potentially mean that, similar to *à*, *sur* is undergoing grammaticalization, but as pointed out in the introduction, this generalization is not, for now, accompanied by other prototypical hallmarks of grammaticalization.

Researchers have argued that the use of *sur* instead of *à* before city names could be explained by the fact that *sur* is used to perform pragmatic functions such as the expression of the trajector’s mobility or the reference to human activities. It may also convey an idea of approximation, where *sur* + city name is used to refer not only to the city itself, but to the general area that surrounds it. This interpretation is discussed by the semanticist Pierre Larrivé in Proust (2017). However, none of these semantic and pragmatic constraints have been tested quantitatively. Moreover, as we see in the next section, little research has examined other potential factors on variation.

2.5 Sur + city name: Other potential factors of variation

Language change can be observed in real-time by comparing the linguistic production of one specific community at several points in time, for example by doing a longitudinal study. But change can also be analyzed in apparent-time by looking at “the distribution of linguistic variables across age levels” (Labov, 1994, pp. 45–46). The apparent-time hypothesis holds that any speech differences between adjacent age groups reveal how the language has evolved, as it follows the uniformitarian principle that states that “knowledge of processes that operated in the past can be inferred by observing ongoing processes in the present”

(Christy, 1983). The general expectation is that, as a change progresses, use of the innovative or incoming form will increase as the age vector decreases. Therefore, age is a crucial factor in the study of language variation and change.

In the case of variation in the PCN context, only one study, Lismoen (2015), has sought to examine whether the age of the speaker affected their likelihood of using *sur* before city names. As is predicted in cases of ongoing change, Lismoen found that speakers in the youngest age group (below 30 years old) used *sur* more frequently than speakers in the oldest age group (up to 79 years old), though the difference was only slight. Lismoen also found that the youngest speakers used *sur* with a greater number of lexical verbs, that is, in more contexts. Although these results are somewhat preliminary, based on made-up examples and surveys conducted with a limited number of participants (N = 19) rather than empirical usage-based data, they are suggestive of an ongoing increase in the use of *sur* alongside ongoing generalization of this form across verbal contexts. In other words, they are suggestive of ongoing change in progress.

Relatedly, some authors (e.g. Fairon et al., 2018; Lismoen, 2015) point out that the *sur* construction may have first appeared with motion verbs used in a military context, such as in *marcher sur Paris* ‘to march on Paris’, before spreading to other verbs like positional ones. This suggests that the type of verb could be a factor of variation as well. It should also be noted that such military contexts have animate trajectors, which supports Huyghe (2012), according to whom *sur* is favoured when there is a reference to human activities and therefore animate trajectors.

Concerning region, Francard (2018) summarizes the different references that he found on the *sur* + city name phenomenon: he notes that its origins can be traced, depending on the source, to South-West France (Le Bidois, 1965), francophone Switzerland and North-East France (Thibault, 2004), and francophone Belgium (Dory, 1878). Considering when it was

published, this last reference indicates that *à/sur* alternation has been observed for at least 150 years. Therefore, it is not a recent innovation, although it is one that appears to be spreading slowly. The role of Switzerland as origin is, nonetheless, difficult to substantiate. The *Corpus oral de français de Suisse romande* (OFROM; Avanzi et al., 2012–2019), which includes 341 sociolinguistic interviews with Swiss participants in the francophone area of Switzerland, contains very few tokens of *sur* (e.g. *sur Neuchâtel*, N = 1 vs. *à Neuchâtel*, N = 93; *sur Genève*, N = 1 vs. *à Genève*, N = 78). Such low frequencies are more likely attributable to geographic spread than to geographic origin. It is interesting to note, however, that all the sources provided by Francard (2018) suggest that the origin of this variation occurred in a fairly circumscribed geographic area: metropolitan France and/or French-speaking areas in neighbouring countries (Belgium and Switzerland). This innovation does not seem to have been discussed for other varieties of French, such as Laurentian varieties and other colonial varieties, despite ongoing colonial expansion in some regions through the beginning of the 20th century (Aldrich, 1996).

In the literature, style is mentioned as a factor of variation only in passing. For example, Hernández (2008) based her analysis on a corpus of personal classified ads because they correspond to “un discours spontané avec tous les traits de l’oralité” [spontaneous speech that contains all the characteristics of orality] (p. 9). Similarly, Lismoen (2015) asked the participants of her survey to choose the preposition that they would use in a “contexte familier à l’oral” [a familiar oral context] (p. 25). Looking at the prescriptivist discourse, the Académie française website⁹ as well as the *Petit bon usage de la langue française* (Fairon et al., 2018) postulate that the *sur* + city name construction first appeared in popular or familiar language before spreading to mainstream media such as television or radio.¹⁰ If this was

⁹ http://www.academie-francaise.fr/questions-de-langue#84_strong-em-sur-paris-em-strong

¹⁰ When searching for ‘sur Paris’ in contemporary texts (from the 1980s onwards) on Frantext, most tokens refer to military contexts, which reinforces Lismoen (2015) and Fairon et al.’s (2018) hypothesis regarding the origins of *sur* in the PCN context. When the preposition is used in non-military contexts, it is found in dialogues, i.e.

indeed the trajectory of diffusion across genres, this would be an indication of change from below, which is a type of change that “appear[s] first in the vernacular, and represent[s] the operation of internal, linguistic factors” (Labov, 1994, p. 78). Change from above, which is defined as “borrowings from other speech communities that have higher prestige in the view of the dominant class” (Labov, 1994, p. 78), seems unlikely since there is no situation of contact with a speech community where *sur* (or a form from which it could be calqued) is prestigious.

In short, the existing literature on the PCN context has focused primarily on semantic and pragmatic nuances between the two (or three) prepositions under study, and has only briefly explored other potential factors of variation, namely age and region. Other social factors like gender or socio-economic status were, to my knowledge, not discussed in the literature. Moreover, where hypotheses have been supported by quantitative evidence, the data has been based on raw token counts and observations from small surveys. In the methods chapter, I explain how some of these shortcomings are addressed here through the use of a variationist framework that allows me to explore the variable grammar of the PCN context and find out which linguistic and social factors constrain this variation in two regions, and three locations, of France.

contexts meant to reflect orality. Therefore, outside the military construction, *sur* + city name does not seem to occur in formal written genres, which adds further support to a hypothesis of change from below. Moreover, the Académie française condemns the use of *sur* in the PCN context when it is not associated with a motion verb.

Chapter 3. Methods

To fully explore the constraints on variation in the PCN context and examine the evolution of this variable context (cf. potential ongoing grammaticalization), adopting a variationist framework is an ideal strategy. Indeed, the goal of variationist methods is to find how the dependent variable (here, the choice of the preposition before a city name) correlates with “patterns of social and/or linguistic phenomena” (Tagliamonte, 2006, p. 76). The vernacular, that is, casual, everyday speech, is seen as the ideal focus of variationist sociolinguistics (e.g. Milroy, 1992; Poplack, 1993). Indeed, Labov (1972, p. 208) argues that the vernacular provides the “fundamental relations which determine the course of language evolution”, because it is the first style that is acquired, and it is less altered by hypercorrection or style-shifting than other speech varieties. Labov (1994) also points out that, in the case of change from below—the more probable scenario for the feature under study—the incoming variant first appears in the vernacular before appearing in more formal styles, registers, and modalities. Therefore, I examined corpora of spoken (and speech-like) language in order to get a picture of two vernacular varieties of Hexagonal French, as it is here that innovative *sur* is more likely to appear in the relevant context. The analysis of a large and representative amount of data enabled me to model prepositional variation both descriptively and inferentially, to assess which factors operate on it and thus to uncover and assess how the variable grammar works for this feature.

3.1 Data

3.1.1 Corpora

The data that was used for the analyses presented here was drawn from two pre-existing corpora of spoken varieties of Hexagonal French: *Corpus de français parlé parisien* and *Enquêtes socio-linguistiques à Orléans*.

The *Corpus de français parlé parisien* (henceforth CFPP2000) is a collection of transcriptions of sociolinguistic interviews that were conducted in the mid-2000s with people living in Paris or adjacent cities (Branca-Rosoff et al., 2012). The sample is concentrated on participants who grew up in the Île-de-France region, making the data representative of French as it is spoken in and around the capital. The corpus comprises roughly 55 hours of spoken data from 60 speakers (35 women and 25 men, born between 1920 and 1994), the majority of which has been transcribed and tagged for part of speech (POS). Considering that all the data in this corpus was collected at one point in time, I will rely on the apparent-time hypothesis to consider the question of ongoing change in the Parisian area.

Enquêtes socio-linguistiques à Orléans (henceforth ESLO) is a resource composed of two collections of spoken French that were compiled in the French city of Orléans (Eshkol-Taravella et al., 2012). The first one (ESLO1) comprises data that was collected between 1968 and 1974, while the second one (ESLO2) comprises data that has been collected since 2008. The materials come primarily from sociolinguistic interviews that were carried out with over 400 participants (262 women and 165 men) living in the Orléans area. The majority of the data has been transcribed and is freely accessible, but unlike CFPP2000 from the Île-de-France region, it is not POS tagged. However, the diachronic and longitudinal aspects of this resource, in addition to the methods used to construct it, were useful for investigating change in the vernacular in real-time, as well as in apparent-time. This resource also allowed me to get a picture of a variety of French that is not Parisian. Considering the large amount of data and speakers of this resource, I chose to confine the analysis to a smaller speaker sample than was available in the entirety of the ESLO1 and ESLO2 collections. The sampling procedure is outlined in the next subsection.

3.1.2 Speaker sample

Table 1 outlines the speaker sample that was used for this research—151 speakers whose years of birth span from the 1870s to the 1990s. For the CFPP2000 corpus, I have extracted data from all the speakers that produced the variable ($N = 60$). Given the markedly larger number of speakers available in the Orléans corpora, I restricted extraction to a subset of the corpus (ESLO1: $N = 49$; ESLO2: $N = 42$). I made the decision to include, where available, a maximum of four speakers per cell (with two professional and two non-professional speakers where possible), in order to attempt to balance the data as much as possible. As is evident from Table 1, some of the cells are rather sparsely represented, particularly for speakers born in the late 19th century and the early 20th century, as well as those born in the 1990s; any results from such periods will necessarily have to be interpreted with caution, regardless of token frequency.

Table 1. Speaker sample

DoB	CFPP2000		ESLO1		ESLO2		Total N
	Women	Men	Women	Men	Women	Men	
1870s			1	—			1
1880s			1	—			1
1890s			4	—			4
1900s			4	4			8
1910s			4	4			8
1920s	4	—	4	4	3	—	15
1930s	2	1	4	4	3	1	15
1940s	10	7	4	4	4	4	33
1950s	3	2	2	1	2	4	14
1960s	3	1			3	3	10
1970s	7	6			4	4	21
1980s	4	8			3	4	19
1990s	2	—					2
Total N	35	25	28	21	22	20	151

(blank cells indicate periods not covered by a particular corpus)

The diachronic aspect of the study relies on the ESLO1 speakers, born between the 1870s and the 1910s. The synchronic aspect relies on the CFPP2000 and the ESLO2 corpora, which

contain data from speakers that were born between the 1920s and the 1990s. Therefore, the real-time analysis comes from a comparison between the ESLO1 and ESLO2 materials, while the apparent-time perspective relies on the CFPP2000 and ESLO2 data. This combination of perspectives provides a longitudinal view of the variable as represented by these materials, expanding the window beyond what is normally possible with strictly synchronic data.

3.2 Data extraction

3.2.1 Setting up the variable context

The variable context of this study, that is, the context in which several variants (here, prepositions) are in alternation, is defined as follows: locative *à/sur/dans* + city name.

Following variationist methods and the principle of accountability, which states that “every variant that is part of the variable context... must be taken into account” (Tagliamonte, 2006, p. 13), I needed to consider all contexts in which the variation of *à*, *dans* and *sur* before city names was possible, and I originally considered all three of these variants. Having said that, however, using a concordancer to exhaustively identify every instance of the variable context in each file was impractical, as it was difficult to predict if speakers would mention city names, and if so, which ones would occur. Therefore, I first compiled a list of city names that appeared in the data, using the POS tagged CFPP2000 corpus to search for proper nouns. I then used this list to search through both the CFPP2000 corpus and the Orléans sub-corpora as well, since these latter materials were not tagged. For ESLO1 and ESLO2, I added the names of the cities in the Orléans Métropole, an administrative group of cities that form the Greater Orléans area, since speakers were likely to mention these locales.¹¹ I used AntConc (Anthony, 2014) to search for and extract *à*, *dans*, or *sur* followed by each city name on the list, using the following search string: *(à/sur/dans) city name*.¹²

¹¹ I located the names via the following online site: <http://www.orleans-metropole.fr/> (date of access: September 9th, 2020).

¹² The complete list of city names that were searched for can be found in Table B1 in Appendix B.

I excluded a number of contexts from the outset. In particular, contexts in which the meaning of the preposition was not locative were excluded out of hand, because they were not part of the envelope of variation. These include *sur* when it means *à propos de* ‘about’, as in (40), and instances of *à* in the construction *c’est à*, where the meaning is ‘it belongs to’, as in (41). Similarly, when the preposition was part of a complex preposition, as in *jusqu’à* ‘until/up to’ or *par rapport à* ‘compared to’, it was excluded.

(40) SUR Rosny je . bon j’ai pas trop de recul effectivement parce que bon ça fait quand même que quatre ans que je suis là (CFPP2000/Isabelle Legrand, b. 1977)
 ‘about Rosny I . well I’m not really objective indeed because well I’ve actually been living here only for four years’

(41) il y a une les collections de Beaubourg quelque part elles sont à moi parce que c’est À Paris elles sont à moi enfin elles sont à moi elles sont collectivement (CFPP2000/Blanche Duchemin, b. 1983)
 ‘there’s a the art collection in Beaubourg in a way it’s mine because it belongs to Paris it’s mine I mean it’s mine collectively it is’

I also excluded tokens that occurred as a result of the speaker repeating verbatim something the interviewer or another speaker had just said, as in (42).

- (42) Interviewee: au départ c'est quelqu'un qui habitait dans notre immeuble qui avait donc
 quand même un bon rythme de vie et cetera et qui a déménagé près d'Orléans c'est ça
 Other speaker: À Compiègne
 Interviewee: À Compiègne n'importe quoi (CFPP2000/Blanche Duchemin, b. 1983)
 'At first it was someone who lived in our building so they had a good lifestyle, etc.
 and they moved near Orléans that's right
 In Compiègne
 In Compiègne I'm talking nonsense'

There were also cases in which the context under study occurred as part of a proper noun. Because no variation is possible (i.e. these tokens are categorical), they were set aside. An example is given in (43).

- (43) Sortir À Saint-Ouen (CFPP2000/Catherine Lecuyer, b. 1973)
 'Going out in Saint-Ouen' (name of a magazine)

The two final exclusion contexts involve expressions. Illustrated in (44), metalinguistic discourse about the variable context itself (or, as in the example, about a construction such as *monter à* that is part of said context) were not included, as they are performative and do not represent spontaneous, unreflective use. And as illustrated in (45), when the variable context is part of an expression that refers to a journey (i.e. *de XX à YY* 'from XX to YY'), the preposition is not variable, and such tokens were therefore not included.

- (44) il était monté À Paris comme on disait à ce moment-là (CFPP/Laurence Leblond, b. 1964)
 'he had "gone up" to Paris as people said back then'

- (45) sur la route de oui sur la route de Saint-Cyr-en-Val là À Orléans (ESLO1/GE133,
b. 1900s)

‘on the road from yes on the road from Saint-Cyr-en-Val there to Orléans’

Apart from these contexts, every other instance of locative *à/sur/dans* + city name was extracted from the speaker transcripts and compiled in a spreadsheet.

3.2.2 A reconsideration of *dans*

When taken all together, these methods resulted in a total of 2682 tokens for analysis (2542 tokens of *à* and *sur* + 140 tokens of *dans*). The overall distribution of the data revealed that, as predicted, *à* was the majority variant, accounting for a full 80.6% of the PCN tokens. *Sur* and *dans* were both minority variants, comprising 14.1% and 5.2% of tokens respectively. The extremely low frequency of *dans*, however, motivated a careful examination of the data. In particular, I wondered whether it should be maintained in the analysis. It was possible, for example, that it was restricted to specific kinds of contexts within the envelope of variation and thus was not on equal status as a full variant with *à* and *sur*. Such partitioning of grammatical variables is not uncommon (see e.g. D’Arcy & Tagliamonte, 2015).

Closer examination of the *dans* tokens (N = 140) revealed that the vast majority, 94.3%, collocated with just two lexical forms, *Paris* (70.0%) and *Orléans* (24.3%). As revealed in Table B1 in Appendix B, each of these city names represent around 27.5% of all the city names in the overall data. Additionally, looking at each corpus separately, I found that 94.8% of *dans* tokens (N = 96) in the CFPP2000 data were tokens of *dans Paris*, with *Paris* representing 43.6% of all city names (n *Paris* = 427, N = 980). In the ESLO data, 77.3% of *dans* tokens (N = 44) were in collocation with *Orléans*, which represents 43.4% of all city names in this dataset (n *Orléans* = 739, N = 1702). The remaining *dans* tokens were almost all tokens of *dans Paris* (15.9%), with *Paris* accounting for 18.2% of all city names (n *Paris* =

310). Therefore, the frequency of collocation of *dans* with *Paris* in the CFPP2000 data and with *Orléans* in the ESLO data is higher than the baseline frequencies of these city names in each of these corpora. In other words, there is a robust lexical skewing effect in the distribution and use of *dans* tokens, as well as near categoricity in the case of the Parisian data. It could therefore be posited that *dans* is mostly used when the city is big and acts as the major urban area in the speaker's daily life. This hypothesis is consistent with Vandeloise's (1986, 2017) remark that, in the container/content relationship specific to *dans*, the container (here, the city) may exert a force or influence on the content (here, the trajector): "x is dans (a spatial entity) y if x is affected by the zone of influence associated to y" (Vandeloise, 2017, p. 13).

Following this hypothesis, it would make sense for speakers to use *dans* primarily with city names that exert or have exerted an influence on them or the trajector they are discussing. This can be seen in (46), where the speaker is talking about regularly going to Paris as a child; the reference to habitual action may also support the use of *dans* in that context. In (47), however, the city name is mentioned in passing and has no obvious influence on the speaker's life; the use of *dans* in this context would sound almost ungrammatical.

(46) on prenait le train et puis on allait se promener DANS Paris (ESLO2/FH35, b. 1940s)
 'we would take the train and then we would go take a walk in Paris'

(47) ils y vont avec une association de de loisirs de Chartres alors j'ai été les conduire à À Chartres (ESLO1/DJ39, b. 1930s)
 ?ils y vont avec une association de de loisirs de Chartres alors j'ai été les conduire
 DANS Chartres
 'they go there with a leisure group from Chartres so I drove them to Chartres'

It also appeared that, in some instances, speakers used *dans* when referring specifically to the centre of the city, as opposed to the outskirts of it, as exemplified in (48) and (49) (note that this also applies to the use of *dans* in (46)). This follows Spang-Hanssen's (1963) remark that *dans* is used in the PCN context to insist on the inwardness of the city.

(48) comme l'endroit où j'habite par exemple porte des Lilas porte de Bagnolet vraiment la périphérie donc c'est plutôt dans ces zones-là oui c'est pas vraiment DANS Paris (CFPP2000/Gary Collard, b. 1985)

'like where I live for example porte des Lilas porte de Bagnolet really the outskirts so that's mostly in these areas yeah it's not really in Paris'

(49) jusque là je pense que on commence à se sentir pas DANS Paris et vraiment en banlieue (CFPP2000/Pierre-Marie Simo, b. 1975)

'until then I think that you start feeling like you're not in Paris and really in the suburbs'

In other words, the use of *dans* in these examples is similar to stressing the preposition *in* in English; that is, it functions pragmatically to signal meaning beyond the referential context encoded in the phrase. As it happens, a mere 5% of *dans* tokens in the dataset had the general locative meaning of 'in the city as a whole or its outskirts', as in (50). The remainder of *dans* tokens were either ambiguous and could be read as a general locative or a pragmatic city centre indicator (52.1%), or they had the more specialized meaning of 'in the city centre' (42.9%). Crucially, neither *à* nor *sur* can be used for this specialized meaning.

(50) moi ce que j'aime bien DANS Orléans c'est les les promenades les jardins les les les environs les environs sont bien jolis (ESLO1/WL512, b. 1890s)

'what I like in Orléans is the the promenades the gardens the the the outskirts the outskirts are really pretty'

On the basis of these distributional and functional results, I concluded that *dans* was indeed used in specialized contexts: with names of major cities (*Paris, Orléans*), and often to refer to the centre of the city specifically as opposed to the entire city, including its outskirts. In this latter function, it does not alternate with the other forms. Therefore, I chose to exclude *dans* from further consideration, and focus on the binary variation between *à* and *sur*. This left a total of 2542 tokens for analysis.

I now outline which social and linguistic predictors were considered for the analysis, and the details of how each factor was coded in the dataset.

3.3 Coding

On the basis of the literature, I originally considered twelve predictors, six linguistic and six social. These are outlined in Table 2. However, it should be noted that, for a variety of reasons, not all of these predictors ended up yielding clear results, and some are only briefly discussed in the analysis.

Table 2. List of predictors and their reference levels

Social predictors	Linguistic predictors
Age (decade of birth)	Type of situation (temporary/other)
Gender (woman/man)	Type of activity (human activities/other)
Occupation (scale from 1 to 5)	Type of trajector (human/non-human)
Education (scale from 1 to 6)	Lexical city name
Socio-economic status (scale from 2 to 11)	Lexical verb
Location (speaker's city or arrondissement)	Type of verb (motion/other)

3.3.1 Social predictors

As discussed in section 2.5, age—as a function of time—is operationalized here as an independent factor to explore the possibility of change in progress and to track the trajectory of the PCN context. To date, the effect of age has been tested for the PCN context in only one study (Lismoen, 2015). Therefore, I coded for the speakers' decade of birth (cf. Table 1) to explore the role of age in *à/sur* variation. Indeed, if a change is in progress with *sur* usage increasing over time, we can predict that younger speakers (i.e. those with more recent dates of birth) will use the incoming *sur* variant more frequently than older speakers (i.e. those with more distal dates of birth), revealing a trajectory of change in both real-time (ESLO1 vs. ESLO2) and apparent-time (CFPP2000 and ESLO2).

Gender has not yet been explored as a potential factor of variation for the feature under study. However, as pointed out by Labov (2001), gender is an important indicator of both stability and change: “gender is a powerful differentiating factor in almost every case of stable social stratification and change in progress that has been studied” (p. 262). Indeed, women generally deviate less from the standard than men for stable variables (i.e. variables that are not undergoing change), but are ahead of men when a change is in progress (Labov, 2001, p. 367). Therefore, including gender in the analysis will allow me to explore if it is a factor in ongoing variation and possible change. From the information made public for both CFPP2000 and ESLO, every speaker identified as either female or male, although this was potentially due to the way in which the demographic information was collected. In any case, I used a binary code for gender: woman, man.

Similar to gender, socio-economic status (SES) has not been explored as a factor of variation in this context, but is also an important sociolinguistic correlate. Indeed, Labov (2001) points out that when the linguistic variable is stable, it is “a monotonic function of social class” (p. 119), meaning that its frequency is stratified across social classes. In the case

of changes from below, the incoming feature originates in the central social classes, i.e. the upper-working and the lower-middle classes (Labov, 2001, p. 188). In the case of changes from above, the incoming feature originates in the higher-ranked classes, especially the “second highest status group” (Labov, 2001, p. 275). Therefore, studying SES as a factor of variation allowed me to test the hypothesis that the feature is undergoing a change from below, with the expectation that upper-working and lower-middle classes speakers would use it more frequently overall. Both the Île-de-France and the Orléans corpora gave information about the speakers’ occupation and educational backgrounds.

Using this information, I was able to code for occupation type and educational background. I initially used the 5-point scale outlined in Table 3 for occupation, and the 6-point scale outlined in Table 4 for education.

Table 3. Scale for occupation type

1	Unemployed
2	Manual workers (e.g. janitors)
3	Skilled manual workers (e.g. construction workers) and employees/salespeople
4	Non-manual workers (e.g. teachers, managers)
5	Professional workers (e.g. engineers, professors, higher managerial positions)

Table 4. Scale for educational background

1	Stopped before high school
2	High school diploma
3	Professional degree
4	Some university
5	Bachelor’s degree
6	Master’s degree and/or PhD

I then combined the occupation and education scores to create an SES score for each speaker. For example, if a speaker ranked 5 for occupation and 4 for education, their final SES score was 9. Therefore, SES scores ranged from 2 to 11. For analytical purposes, these were then merged into three categories: ‘upper class’ (SES score: 10 to 11), ‘middle class’ (SES score: 7

to 9), and ‘working class’ (SES score: 2 to 6). I also retained education and occupation levels as individual factors, so that I could establish the best SES model for the data rather than pre-determining it (i.e. education and/or occupation vs. combined education and occupation information).

Finally, I coded for location. The two corpora represent two geographic regions: Île-de-France (CFPP2000), which comprises Paris and some of its adjacent cities (Suresnes in the west, Saint-Ouen in the north, Montreuil and Rosny-sous-Bois in the east, Ivry-sur-Seine and Le Kremlin-Bicêtre in the south), and Orléans (ESLO), a smaller city located 130 kilometers south of Paris.

Île-de-France is the largest urban centre in France with 11.7 million inhabitants (Branca-Rosoff et al., 2012), and it can be contrasted with Orléans, which counted 96,000 inhabitants in 1968 (when data collection for ESLO1 began), and 113,000 inhabitants in 2008 (for the ESLO2 data) (Insee, 2019). Additionally, since the CFPP2000 corpus contains data from speakers who live both inside Paris and also in surrounding cities, I was able to distinguish speakers from Paris and those from its suburbs. This allowed me to explore whether the data distributed differently within a region with distinct urban settings (Paris/urban, Parisian suburbs/semi-urban) and across regions (Île-de-France, Orléans).

Since the Île-de-France region is more populated than Orléans, it would be expected that *sur* would be more frequent there, following classic models of geolinguistic diffusion like the gravity model (Bailey et al., 1993) or the cascade model (Trudgill, 1974). Both models predict that incoming linguistic features spread from more populated to less populated urban centres.¹³ Moreover, Labov (2007) points out that, in the case of a change from below—which I assume *sur* to be—the change first appears through generational incrementation in the

¹³ Statistics related to race and ethnicity are prohibited in France, but it is likely that the Île-de-France region, as France’s most important urban centre, is more ethnically diverse than Orléans. This could have an impact on the linguistic features used by the speakers in each region, but the lack of information about this in the corpora meant that I was not able to code for, and test, the effects of ethnicity on variation.

speech community of a major urban centre and may then diffuse to other smaller urban centres through contact between adult speakers. However, in cases of such diffusion, he argues that the linguistic variable “expands in a weaker, not a stronger, form” (p. 6), and that “transmission across adults... loses much of the fine structure of the linguistic system being transmitted” (p. 35). Therefore, in the case of the PCN context in Hexagonal French, we can predict that the change originated in the Parisian area, which is not only the most populated area of France but also its cultural, political and economic centre. It then likely spread to smaller areas such as Orléans, but with an altered or weakened variable grammar and constraints. Tagliamonte and Denis (2014) indeed point out that diffusion arises “via contact when an innovation evolves in one community on the model of another community and where distinct constraints on the system obtain—the result of imperfect intergenerational learning” (p. 92).

The CFPP2000 corpus also allowed me to make more fine-grained distinctions between the locations of each speaker. Indeed, in addition to making a distinction between Paris and its suburbs, the CFPP2000 materials provide information on the speakers’ location at the time the interviews were recorded: for Paris, their *arrondissement*,¹⁴ and for the suburbs, their city. Figure 1 presents a map of the different locations that were included in the corpus. All these locales (e.g. 3rd *arrondissement*, Montreuil, etc.) were given a separate code, in order to find out whether the data distributed differently across neighbourhoods and cities in the Parisian corpus.

¹⁴ Paris is divided into twenty administrative districts, called *arrondissements*.

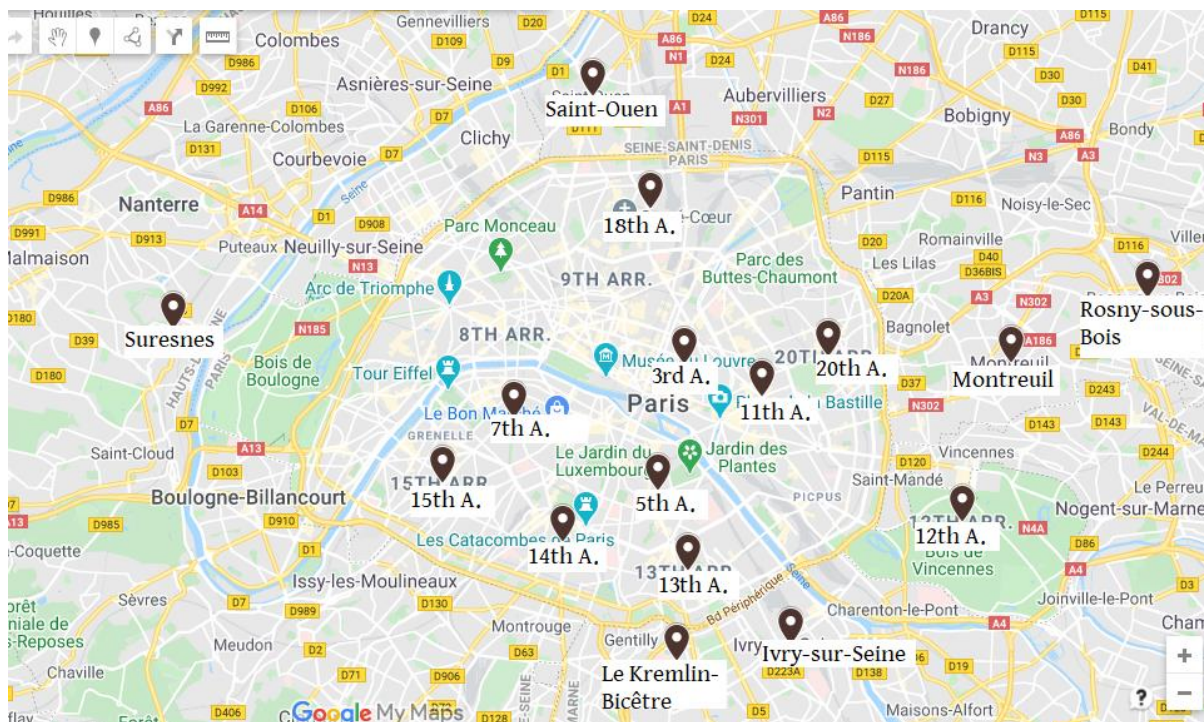


Figure 1. Map of the locations that were included in the CFPP2000 data (screenshot from Google Maps)

To summarize, I coded for the following social factors: age, gender, SES, and location.

Occupation and education were also retained as individual factors in case they turned out to be relevant for the analysis. I now move on to the linguistic factors that I coded for.

3.3.2 Linguistic predictors

As seen in Table 2, the first factor to be tested, type of situation, was the one hypothesized by Hernández (2008), according to whom *sur* + city name is more frequently used by speakers to refer to temporary situations in contexts where mobility is implied. Hernández points out that several linguistic clues can lead to such an interpretation. For example, the speaker might frame their utterance with a reference to a short duration, as in (51), or hint at the fact that the situation is somehow exceptional, as in (52).

(51) Je suis SUR Paris pour faire un stage de trois mois.

‘I am in Paris for a three-month internship.’

- (52) Une fois n'est pas coutume, je reste SUR Paris ce soir. (Hernández, 2008, p. 10)
 'Just this once, I'll stay in Paris tonight.'

Nonetheless, this predictor was not always straightforward to code, as some contexts were ambiguous or did not include reference to duration or mobility at all, as in (53). In this example, it is difficult to determine from context whether the speaker is talking about a situation that has been occurring for a long time, or that is relatively new. Such tokens were coded as ambiguous.

- (53) parce que c'est vrai que À Paris c'est cher quand même tous les cafés tout ça ça revient cher (CFPP2000/Katia Teixeira, b. 1994)
 'because it's true that in Paris it's expensive really all the cafes and everything it's expensive'

The second factor, type of activity, was suggested by Huyghe (2012). It predicts that *sur* + city name is favoured when the landmark is presented as a place dedicated to human activities, such as professional activities. In practical terms, when coding for this factor I looked for clues that referred to such activities, such as a human subject or a reference to the job environment, as in (54). I coded such tokens as 'human activities'. The rest of the tokens either did not refer to any kind of activity or were ambiguous, as in (55), and I coded them as 'no reference to activities'.

- (54) et mon mari a effectivement la possibilité de de retravailler SUR Paris (ESLO1/EH5, b. 1960s)
 'and my husband indeed got the opportunity to work in Paris again'

(55) il y a des choses que maintenant à voir À Orléans qui n'existaient pas avant
(ESLO1/JR327, b. 1930s)

‘there are things now that you can see in Orléans that didn’t exist before’

To further test Huyghe’s hypothesis, I also coded for the type of trajector. This was a binary group, contrasting human (as in (54)) and non-human (as in (55)) trajectors. My hypothesis was that non-human trajectors would be a disfavourable context for the use of *sur*.

I also coded for the city name with which the preposition appeared. This allowed me to explore whether *sur* was favoured with specific city names, for example bigger cities as opposed to smaller ones. It was also possible that the city name could capture a frequency effect, as it could be hypothesized that *Paris* and *Orléans* would be discussed at length in their respective corpora, and perhaps their increased frequency would lead to a higher rate of *sur*, compared to other less frequent city names. Indeed, Bybee and Hopper (2001) point out that “accounts of grammatical... structure must take note of how frequency and repetition affect and, ultimately, bring about form in language” (p. 3). Note that, for this factor, only the city names that had over 15 tokens in each dataset were given an individual code. The rest of them were grouped into an ‘other’ category. For example, in the ESLO data, *Nantes* and *Lille* were coded as ‘other’, as they had 13 tokens and 5 tokens respectively.

Finally, I considered the lexical verb itself, to see if particular collocational types favoured the use of *sur*, bearing in mind the possible effects of frequency.¹⁵ Type of verb was also considered as a factor of variation, which allowed me to test whether *sur* was favoured with motion verbs rather than other ones, as suggested by Hernández (2008), Lismoen (2015) and Fairon et al. (2018). To code for this factor, I considered whether the verb implied an idea

¹⁵ Due to the inconclusive patterns it yielded, the lexical verb predictor was not included in the final analysis. The distributional results for this factor can be found in Tables A7 and A8 in Appendix A.

of movement or not¹⁶. For example, *aller* ‘to go’, *partir* ‘to leave’, or *revenir* ‘to come back’ were all coded as motion verbs, while *être* ‘to be’, *travailler* ‘to work’, or *rester* ‘to stay’ were all coded as ‘other’. The list of verbs that were coded as motion verbs can be found in Appendix C.

Note that the type of verb and lexical verb factors were not relevant for all the tokens. Indeed, in some cases, the prepositional phrase was associated with something other than a verb phrase, such as a noun phrase or an adjectival phrase, as seen in (56) and (57). Such tokens were coded as ‘N/A’ for these two factors.

(56) oui parce que j’étais déjà dans un collège privé SUR Bagnolet en fait

(CFPP2000/Isabelle Legrand, b. 1977)

‘yes because I was already attending a private middle school in Bagnolet actually’

(57) c’était un un ensemble d’usines et d’ateliers qui était très fréquentes À Montreuil

dans les années cinquante (CFPP2000/André Morange, b. 1951)

‘it was a group of factories and workshops that were very common in Montreuil in the fifties’

Moreover, just like for the city name factor, only the verbs that had over 15 tokens in each dataset were given an individual code for the lexical verb factor. The rest of them were grouped into an ‘other’ category, such as *se promener* ‘to take a walk’ in the ESLO data (4 tokens).

In short, these variationist methods enabled me to explore the alternation between *à* and *sur* in the PCN context in two distinct collections of data representing two geographic areas of Metropolitan France. In order to investigate the variable grammar of this feature, a

¹⁶ Note that motion verbs include verbs that encode an idea of direction (such as *aller* ‘to go’), and those that do not (such as *marcher* ‘to walk’). Future work could explore whether this distinction is also a factor on variation.

variety of factors was coded for: decade of birth, gender, occupation, education, socio-economic status, and location for the social factors, and type of situation, type of activity, type of trajector, lexical city name, lexical verb, and type of verb for the linguistic factors. Using the variable rule program Rbrul (Johnson, 2008), distributional and inferential analyses were performed, and they are presented in the next section.

Chapter 4. Results

4.1 Overall distribution for the combined corpora

When taking all the instances of *à* and *sur* together ($N = 2542$), *sur* occurs in the PCN context 14.9% of the time, which represents 379 tokens; *à* is by far the majority variant, accounting for the remaining 2163 tokens, at 85.1% overall. Strikingly, an identical distribution is found both for the CFPP2000 data (*sur* 14.9%, $n_{sur} = 132$, $N_{sur+à} = 884$) and the ESLO data (*sur* 14.9%, $n_{sur} = 247$, $N_{sur+à} = 1658$) when they are considered separately. Although such results could lead us to think that both locations have matching distributional conditions and variable grammars, it should be kept in mind that the two corpora represent different regions, demographics, and time frames (from the 1870s to the 1980s for Orléans, and from the 1920s to the 1990s for Paris and its suburbs). There are good reasons, therefore, to expect differences to emerge across the datasets. In order to explore the distribution of the data in a more detailed way, I examine the results for each corpus separately from this point forward, and report the distribution for linguistic and social factors for each locale. This allows me to compare and contrast the variable grammars in each area, starting with Paris, then moving on to Orléans.

4.2 Distributional results for the Parisian data

4.2.1 Overall distribution and longitudinal trajectory

As noted above, *sur* occurs 14.9% overall in the CFPP2000 materials ($N_{sur+à} = 884$). The literature suggests that this preposition has been used in the PCN context in Hexagonal French since at least the mid-20th century (cf. chapters 1 and 2). As the incoming variant, we expect to see an incremental increase of *sur* over apparent-time. Looking at Figure 2 and Table 5, we can make two primary observations. First, *sur* is attested in the Parisian data from speakers born in the 1940s. Although this would seem to align with claims in the literature, there are

very few tokens of either *à* or *sur* + city name from speakers born in the previous two decades—just 27 for the 1920s and 32 for the 1930s. As such, the results should be treated with caution and the 1940s not assumed to be the point of actuation (recall, for example, that it was attested in Belgian French in the late 19th century); it is entirely likely that—through a combination of low overall frequency of occurrence (a hallmark of incipient forms) and sparsity of data—the CFPP2000 dataset simply failed to capture *sur* in this specialized context of occurrence among speakers born earlier in the 20th century (I return to this point in footnote 20).

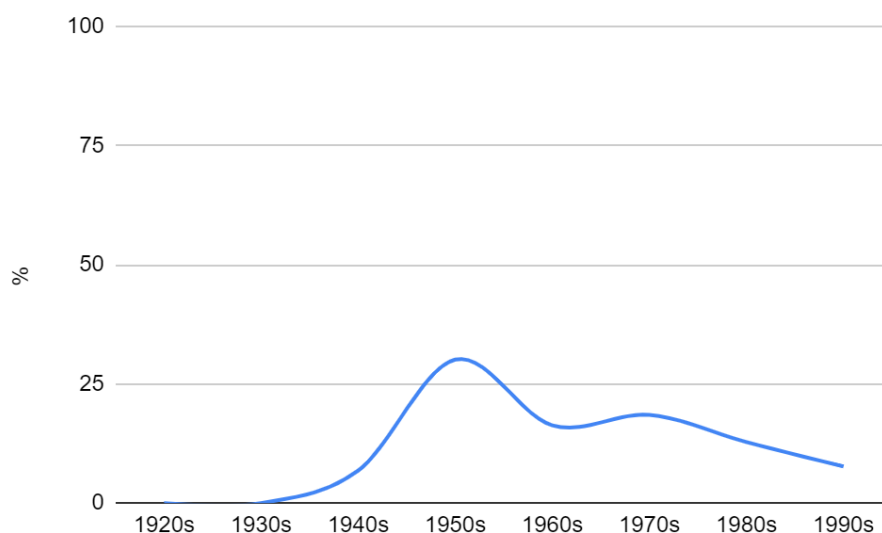


Figure 2. Distribution of *sur* by decade of birth in the CFPP2000 data

Table 5. Distribution of *sur* by decade of birth in the CFPP2000 data

	% (<i>sur</i>)	n (<i>sur</i>)	n (<i>sur</i> + <i>à</i>)
1920s	0.0	0	27
1930s	0.0	0	32
1940s	6.8	15	220
1950s	30.1	44	146
1960s	16.3	8	49
1970s	18.5	43	233
1980s	12.8	21	164
1990s	7.7	1	13
Total		132	884

Second, the trajectory in Figure 2 is not one of monotonic increase over time, the predicted pathway of change for incoming forms (Labov, 2001). After an immediate period of intense and rapid rise in frequency, there is a clear peak in the use of *sur* among speakers born in the 1950s, a peak that is immediately followed by a prolonged period of regular and incremental decline. Thus, it seems that the PCN context is reverting toward the prototypical and long-entrenched *à*. That is, *sur* may represent a short-term innovation that has reversed, rather than a long-term change that will persevere. In the next subsections, we will also see that the use of *sur* inside Paris is restrained to specific linguistic contexts, which means that it might not be able to replace *à* in all environments, and this could partly explain why we do not find a monotonic increase of *sur* in the Parisian data.

Of course, Figure 2 reflects the distribution of the feature in the whole of the CFPP2000 corpus, which contains data from two locations: the city of Paris itself, and its surrounding suburbs. This is not trivial: social, demographic, ethnic, and cultural differences differentiate Paris from its suburbs (see Cheshire, 2020). The overall distribution of *sur* in each of these locations suggests that the trajectory in Figure 2 actually masks a split between these two areas: the overall frequency of *sur* drops to a mere 5.84% in the city ($N_{sur + \grave{a}} = 394$), but rises to 19.6% in the suburbs ($N_{sur + \grave{a}} = 490$). A simple chi-square test with Yates correction indicates that this difference between the urban and peri-urban contexts is significant ($X^2 (df 1, N = 884) = 34.2951, p = 0.00001$).

With this in mind, Figure 3 and Table 6 present the apparent-time trajectories for Paris and its suburbs separately. The pattern within the city is erratic, with *sur* appearing to rise and fall over time. As it happens, only 8 of the 40 urban speakers in the corpus use *sur*, while the remaining 32 use *à* categorically. In other words, the city-centre trajectory in Figure 3 is driven by a minority of speakers, suggesting that the feature is not entrenched in the Parisian vernacular overall. It is also difficult to ascertain why certain speakers use the feature while

others do not; some are men, some are women, some are older, some are younger, they are spread across a range of arrondissements and they come from a range of backgrounds (see section 4.2.2). And notably, the 1970 and 1980 cohorts are the only ones for which *sur* is attested in the corpus by more than one speaker. As such, the distribution of *sur* in Paris in Figure 3 may reflect idiolectal propensities rather than a meaningful temporal pattern. All of these factors combine to suggest that *sur* has never had more than limited use in the city and that it is very much a minority variant. As a result, I concentrate on the variable grammar of those 8 urban speakers who do use *sur*,¹⁷ and exclude the data from the categorical *à* users. This leaves 102 tokens for analysis, among which 22.5% are tokens of *sur*.

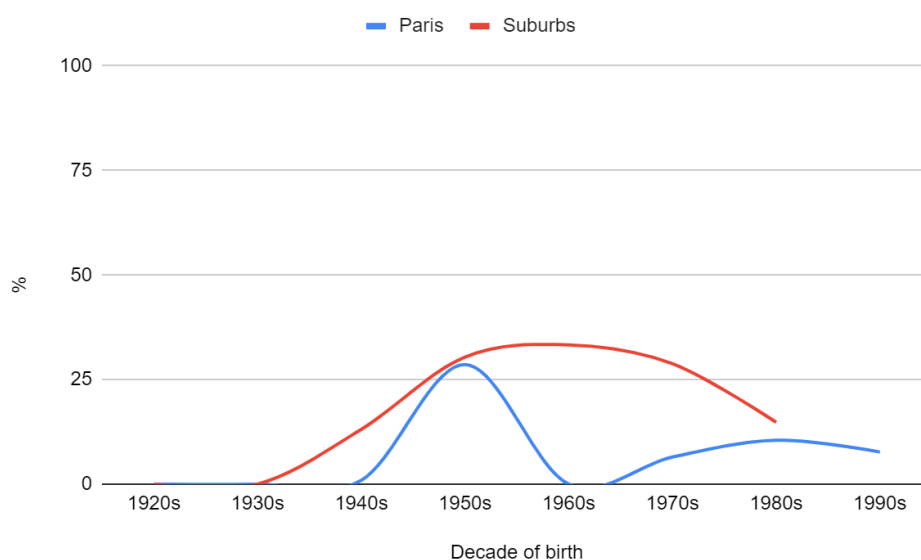


Figure 3. Distribution of *sur* by decade of birth inside Paris and in the Parisian suburbs

¹⁷ Of these 8 speakers, 2 are men (1 born in the 1940s, and 1 in the 1980s) and 6 are women (1 born in the 1950s, 3 born in the 1970s, 1 in the 1980s, and 1 in the 1990s).

Table 6. Distribution of *sur* by decade of birth inside Paris and in the Parisian suburbs

	Paris			Suburbs		
	% (<i>sur</i>)	n (<i>sur</i>)	n (<i>sur</i> +à)	% (<i>sur</i>)	n (<i>sur</i>)	n (<i>sur</i> +à)
1920s	0.0	0	17	0.0	0	10
1930s	0.0	0	21	0.0	0	11
1940s	0.9	1	113	13.1	14	107
1950s	28.6	6	21	30.4	38	125
1960s	0.0	0	25	33.3	8	24
1970s	6.5	7	108	28.8	36	125
1980s	10.5	8	76	14.8	13	88
1990s	7.7	1	13			
Total		23	394		109	490

In the suburbs, the situation is markedly different. The peri-urban context is represented in the corpus by 20 speakers, and a majority of them (13 speakers) use *sur*.¹⁸ This suggests that in the Parisian suburbs, unlike in the city itself, *sur* is a robust and viable locative in the PCN context. In other words, *sur* is not an urban form so much as it is a peri-urban one. This consistent representation across individuals also enables me to feel more confident about the temporal pattern in Figure 3. The first attestations occur among speakers born in the 1940s, and then the use of *sur* rises to a high of approximately 30% in the PCN context.¹⁹ It then exhibits three decades of relative stability, from the 1950s to the 1970s. *Sur* now seems to be on the decline, as speakers go back to the conservative (and potentially now socially marked; see section 5.1) locative, *à*.

Given the differences across the CFPP2000 datasets, from this point forward I treat each Parisian subset (urban vs. suburban) separately. As the analysis will show, the contrasts

¹⁸ Of these 13 speakers who use *sur*, 6 are men (1 born in the 1940s, 1 in the 1950s, 4 in the 1980s), and 7 are women (2 born in the 1940s, 1 in the 1950s, 1 in the 1960s, and 3 in the 1970s).

¹⁹ The view provided by Figure 3 suggests that *sur* may have appeared in the suburbs somewhat earlier than it did within the city (i.e. speakers born in the 1940s versus those born in the 1950s, although the Parisian subset does have one 1940s speaker who uses *sur* once). Given the sparsity of the data, which was noted above for the aggregate view, and is an issue that is exacerbated here, such an interpretation is not well-founded. Indeed, that the frequency of *sur* should rise from 0.9% to nearly 30% overall in Paris in less than a decade (note that speakers in the 1940s cell were born as late as 1948, and those in the 1950s cell as early as 1955), is a strong indication that *sur* was already in use prior to this point in Paris and that the corpus simply fails to capture more than one attestation.

in the variable grammar for the PCN context go beyond frequency of occurrence, distribution by speaker and (possible) trajectory of change, affecting other constraints on variation as well. Additionally, because I am concerned with variation between \grave{a} and *sur*, from this point forward I have again removed the speakers from the data where \grave{a} is categorical. In Paris, as noted above, this brings the combined number of *sur* and \grave{a} tokens to 102, where the overall frequency of *sur* is 22.5% (N speakers = 8). In the suburbs, the dataset consists of 382 \grave{a} and *sur* tokens, where the overall frequency of *sur* is 28.5% (N speakers = 13). I note that the suburban distribution can likely be taken as more or less representative of the community norm, whereas in the urban data, it represents only a minority of speakers overall and not the distributional norm for the city itself.

In the next two sections, I discuss the effects of the remaining social factors in both locales, before moving on to the linguistic constraints, as outlined in section 3.4.

4.2.2 Social factors

The CFPP2000 materials provide information on the speakers' location at the time the interviews were recorded. Tables 7 (Paris) and 8 (suburbs) report the results for *sur* according to this information.

Table 7. Distribution of *sur* by arrondissement inside Paris

	% (<i>sur</i>)	n (<i>sur</i>)	n (<i>sur</i> + \grave{a})	Speaker n
7 th arrondissement	40.0	2	5	2
11 th arrondissement	12.5	1	8	1
13 th arrondissement	22.7	5	22	2
14 th arrondissement	33.3	6	18	1
18 th arrondissement	7.7	2	26	1
20 th arrondissement	30.4	7	23	1
Total		23	102	8

Looking at the last two columns in Table 7, we can see that the data in the majority of the cells is sparse. Although the distributional data suggests that use of *sur* is potentially robust in

some arrondissements (e.g. 7th, 14th, 20th), most are represented by a single speaker; no arrondissement is represented by more than two speakers. As such, I set this factor aside for the city.

Table 8. Distribution of *sur* by speaker's location in the Parisian suburbs

	% (<i>sur</i>)	n (<i>sur</i>)	n (<i>sur</i> +à)	Speaker n
Saint-Ouen (north)	51.7	30	58	3
Ivry-sur-Seine (south)	50.0	52	104	4
Rosny-sous-Bois (east)	45.2	14	31	1
Montreuil (east)	12.3	8	65	2
Le Kremlin-Bicêtre (south)	5.6	3	54	2
Suresnes (west)	2.9	2	70	1
Total		109	382	13

Looking at the distribution of *sur* by location in the suburbs (Table 8), there seems to be a split between the three top cities, where frequencies are all above 45%, and the three bottom cities, where frequencies are all below 13%. Although the number of speakers per cell is again relatively low—especially for Rosny-sous-Bois and Suresnes—it is notable that the locations are generally represented by more individuals and higher token counts than in Paris. Thus, the results in Table 6 can be taken to suggest a potential effect of location with regard to *sur* usage. However, there is no obvious geographic pattern that could explain the contrast between Saint-Ouen and Ivry-sur-Seine on the one hand, and Le Kremlin-Bicêtre and Montreuil on the other hand. For example, Saint-Ouen is located in the north of Paris while Ivry-sur-Seine is in the south. Therefore, it is possible that the explanation can be found in the social and demographic characteristics of each particular city, but this is beyond the scope of this study. More crucially, however, the evidence in Table 8 supports the interpretation that *sur* in the PCN context is a productive part of the vernacular in the Parisian suburbs, while it remains markedly more restricted in the city itself.

Table 9 reports use of *sur* according to gender. In Paris, men have higher rates of use than women, but it should be noted that only 2 men out of the original 17 (11.8%) use *sur* in this dataset, against 6 women out of the original 23 (26.1%). This suggests that the use of the feature among men is rare inside Paris, but when they do use it, they may use it frequently (though it is difficult to say, since 23 of the 24 male tokens in Table 9 come from a single individual). That it is primarily attested among women, and largely confined to a single man, suggests that it is associated with women rather than with men in Paris. In parallel, the suburban data displays an association of *sur* with women, who use *sur* at twice the frequency men do. Higher frequencies of innovative variants among women are consistent with known patterns of linguistic change generally (e.g. Labov, 2001), though as discussed above, the apparent-time trajectory problematizes an interpretation of *sur* as ‘incoming’ (though it certainly can be categorized as a relatively recent innovation, a category of change also generally lead by women).

Table 9. Distribution of *sur* by gender in the CFPP2000 data

	Paris			Suburbs		
	% (<i>sur</i>)	n (<i>sur</i>)	n (<i>sur</i> +à)	% (<i>sur</i>)	n (<i>sur</i>)	n (<i>sur</i> +à)
Women	19.2	15	78	35.0	83	237
Men	33.3	8	24	17.9	26	145
Total		23	102		109	382

Given evidence for an effect of a social factor—gender—it is likely that SES is also relevant.

The results are presented in Table 10.²⁰

²⁰ Since education and occupation were retained as individual factors while coding, I considered how they influenced the data when taken separately. The distributional results for these two factors can be found in Tables A1 through A4 in Appendix A, where the same patterns are visible as in this amalgamate SES measure. Notably, the differences in rates between groups of speakers are generally more marked in the suburbs than inside Paris.

Table 10. Distribution of *sur* by socio-economic status in the CFPP2000 data

	Paris			Suburbs		
	% (<i>sur</i>)	n (<i>sur</i>)	n (<i>sur</i> +à)	% (<i>sur</i>)	n (<i>sur</i>)	n (<i>sur</i> +à)
Upper class	16.7	1	6	2.9	2	70
Middle class	23.6	17	72	61.5	64	104
Working class	20.8	5	24	20.7	43	208
Total		23	102		109	382

A similar pattern is evident in both locations. As predicted in the case of a change from below, which tends to follow the curvilinear principle (Labov, 2001), *sur* is more frequent in the middle class than it is in the upper and working classes. Moreover, in both locations, *sur* is least frequent within the upper class. The effect of SES is greater in the suburbs, however, where the distributional differences are more marked. Indeed, in the suburbs, the upper class displays a nearly categorical avoidance of *sur* whereas for the middle class group it is notably the majority preposition in the PCN context. In the suburbs, the only other contexts where *sur* is the majority variant are among speakers living in Saint-Ouen and Ivry-sur-Seine (cf. Table 6). The findings for SES indicate that *sur* is more strongly socially stratified by socio-economic factors in the suburbs than it is inside Paris, where the differences between SES groups are more subtle.

To summarize the findings for social factors, within Paris it is difficult to disentangle whether the patterns emerge from bona fide social constraints or whether they are an effect of individual differences. In the suburban data there are more tokens in each cell and the use of *sur* is more frequent, both at the speaker level and at the community level. This enables us to see clear social stratification: *sur* is more frequent among women and within the middle class. There also appear to be effects of location. The overarching takeaway is that the urban/peri-urban distinction is not trivial. *Sur* is restricted in Paris, limited to a handful of speakers who make use of this form. In the Parisian suburbs, on the other hand, it has diffused across speakers and, in that sense, it is a more productive resource in the vernacular.

4.2.3 Linguistic factors

As discussed in chapter 3, most of the linguistic factors are based on the existing literature on the subject. The goal is to explore whether these claims about *sur* are supported by quantitative analysis of vernacular data.

The first factor to be considered is type of situation, which contrasts temporary and long-term situations. Hernández (2008) hypothesized that *sur* is more frequent when the context refers to a temporary situation (i.e. when the mobility of the speaker is implied) and less frequent when the context emphasizes a long-term situation. Although I had originally included a third category, ‘ambiguous’, when the context did not provide any clue as to the longevity of the situation the speaker was describing, these tokens ended up patterning similarly to the long-term ones, suggesting a binary distinction between contexts that refer to unambiguously temporary situations, such as example (58), versus all other contexts, such as examples (59) and (60).

(58) euh le samedi je sors avec des amies soit on mange ensemble soit on on fait euh enfin on va SUR Paris au cinéma ça dépend (CFPP2000/Assia Berrada, b. 1964)

‘um on Saturdays I go out with friends we either have dinner together or we do um I mean we go to Paris to the movies it depends’

(59) bon parce que comme les parents habitent À Garches je me disais aussi que ça pouvait être pratique (CFPP2000/Dominique Valin, b. 1971)

‘well because the parents live in Garches I was also telling myself that it could be convenient’

(60) euh AU Kremlin il y a pas beaucoup de enfin il y a des des pigeons (CFPP2000/Killian Belamy, b. 1986)

‘um in Le Kremlin there isn’t a lot of I mean there are pigeons’

As a result, from this point I collapse the long-term and ambiguous temporality tokens into a single category. The distributional results for the CFPP2000 data are reported in Table 11. As hypothesized by Hernández (2008), *sur* is more frequent when the context refers to a temporary situation, an effect that is evident in both the city and its suburbs.

Table 11. Distribution of *sur* by type of situation in the CFPP200 data

	Paris			Suburbs		
	% (<i>sur</i>)	n (<i>sur</i>)	n (<i>sur</i> +à)	%	n (<i>sur</i>)	n (<i>sur</i> +à)
Unambiguously temporary situations	35.3	12	34	40.8	51	125
Other situations	16.2	11	68	22.6	58	257
Total		23	102		109	382

The next factor is activity type, which contrasts contexts referring to human activities with those that do not. This contrast is illustrated in examples (61) and (62). In (61), and when considering the entirety of the sentence rather than only the verb and the PCN context—a procedure that was applied throughout coding all the tokens for this factor—the landmark is envisioned as a setting for human activities (the pool and the market). In (62), no explicit reference to such activities is made. Huyghe (2012) predicted that contexts like (61) are more susceptible to give rise to *sur* than contexts like (62).

- (61) les filles ont piscine enfin quand elles sont avec moi À Garches et que la piscine est juste à côté du marché (CFPP2000/Dominique Valin, b. 1971)
 ‘the girls go to the swimming pool when they’re with me in Garches and the pool is just next to the market’

- (62) mais il y avait des bois À Montreuil à l'origine puisque ça qu'il s'appelle
Montreuil-sous-Bois (CFPP2000/Mireille Masson, b. 1944)
'but originally there were woods in Montreuil since it's called Montreuil-sous-Bois'
(“Montreuil-under-the-woods”)

The distributional results can be found in Table 12. In both locations, the predicted direction of effect is evident yet it is also marginal, especially in the city. Chi-square tests with Yates correction indicate that neither is significant at $p < .05$ (Paris: X^2 (df 1, $N = 102$) = 0, $p = .939735$; suburbs: X^2 (df 1, $N = 382$) = 2.5, $p = .110911$), suggesting that this is not an active part of the variable grammar.

Table 12. Distribution of *sur* by type of activity in the CFPP2000 data

	Paris			Suburbs		
	% (<i>sur</i>)	n (<i>sur</i>)	n (<i>sur</i> +à)	% (<i>sur</i>)	n (<i>sur</i>)	n (<i>sur</i> +à)
Human activities	24.0	6	25	32.9	56	170
Other activities	22.1	17	77	25.0	53	212
Total		23	102		109	382

Moving to type of trajector, this factor tests the potential effect of human versus non-human trajectors, a contrast illustrated in (63) and (64). It was deduced from Huyghe (2012), who suggested that *sur* is used in contexts where human activities are emphasized.

- (63) je suis arrivée À Montreuil en quatre-vingt-onze (CFPP2000/Mireille Masson,
b. 1944)
'I arrived in Montreuil in ninety-one'
- (64) quand on va sur les autres commerces SUR Saint-Ouen y a quand même des
commerces qui sont fermés (CFPP2000/Catherine Lecuyer, b. 1973)
'when we go to the other stores in Saint-Ouen there's still stores that are closed'

Table 13 reports the results for the CFPP2000 data. Inside Paris, there is virtually no difference between the two contexts, while the suburban data reveals a possible effect. A chi-square test with Yates correction suggests that the effect is not significant, however (χ^2 (df 1, $N = 382$) = 2, $p = .15399$).

Table 13. Distribution of *sur* by type of trajector in the CFPP2000 data

	Paris			Suburbs		
	% (<i>sur</i>)	n (<i>sur</i>)	n (<i>sur</i> +à)	% (<i>sur</i>)	n (<i>sur</i>)	n (<i>sur</i> +à)
Human	22.7	15	66	32.0	63	197
Non-human	22.2	8	36	24.9	46	185
Total		23	102		109	382

The next linguistic constraint is type of verb. The prediction for this factor is that motion verbs such as *aller* ‘to go’ or *venir* ‘to come’ are more favourable to the newer variant than are other types of verbs, or non-verbal contexts. Indeed, it has been argued that the *sur* + city name construction emerged in military contexts with the expression *marcher sur* + city, which encodes a meaning of movement (e.g. Hernández, 2008; Lismoen, 2015; Fairon et al., 2018).

Table 14. Distribution of *sur* by type of verb in the CFPP2000 data

	Paris			Suburbs		
	% (<i>sur</i>)	n (<i>sur</i>)	n (<i>sur</i> +à)	% (<i>sur</i>)	n (<i>sur</i>)	n (<i>sur</i> +à)
Modifies a motion verb	50.0	9	18	34.3	23	67
Modifies something else	16.7	14	84	27.3	86	315
Total		23	102		109	382

The results in Table 14 suggest that the effect of this putative source context remains entrenched, a regular finding in the grammaticalization literature (cf. Poplack & Tagliamonte, 1999; Tagliamonte et al., 2014, on the grammaticalization of *going to*). Particularly evident in

Paris, *sur* is more frequent with motion verbs than it is in any other contexts; initial testing suggests that the effect is not significant in the suburbs (X^2 (df 1, $N = 382$) = 1, $p = .313626$).

To further explore the type of verb constraint and its evolution over time, I cross-tabulated speaker's decade of birth and verb type. Given the nature of the urban dataset, this analysis was feasible only for the suburban data. Figure 4 and Table 15 present these results. We can see that, through to speakers born in the late 1960s, there is a consistent tendency for *sur* to be more frequent with motion verbs than elsewhere. However, by speakers born in the 1970s, the contextual distinction has levelled. This may be a consequence of ongoing obsolescence, ongoing grammaticalization, or both.²¹

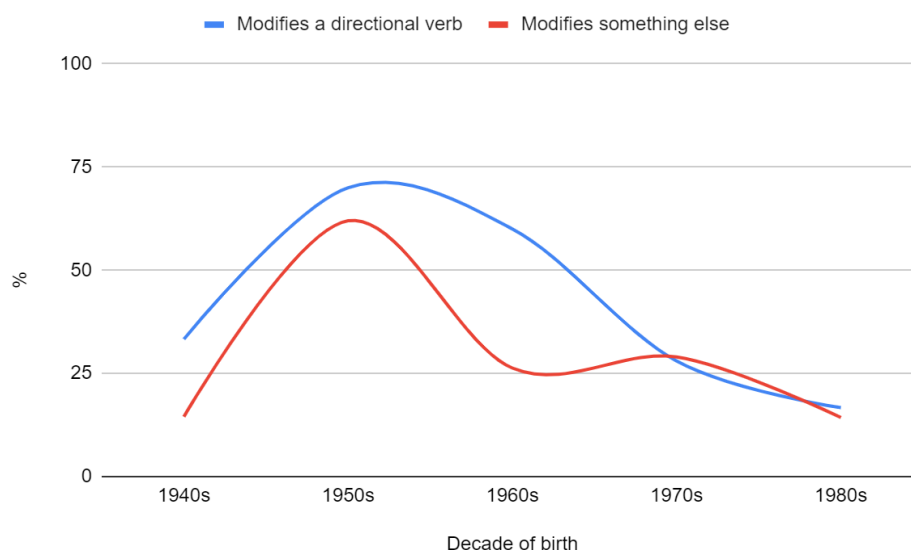


Figure 4. Distribution of *sur* by decade of birth and type of verb in the Parisian suburbs

²¹ As discussed in the methods, I also checked for a potential effect of the lexical verb itself (see Table A7 in Appendix A), but no individual form seems to be driving the patterns we see.

Table 15. Distribution of *sur* by decade of birth and type of verb in the Parisian suburbs

	Modifies a motion verb			Modifies something else		
	% (<i>sur</i>)	n (<i>sur</i>)	n (<i>sur</i> +à)	% (<i>sur</i>)	n (<i>sur</i>)	n (<i>sur</i> +à)
1940s	33.3	3	9	14.5	11	76
1950s	70.0	7	10	62.0	31	50
1960s	60.0	3	5	26.3	5	19
1970s	28.0	7	25	29.0	29	100
1980s	16.7	3	18	14.3	10	70
Total		23	67		86	315

Finally, I consider the potential effect of the city name itself in the PCN construction. As noted in the methods, only those city names that had more than 15 tokens in each dataset were coded separately; the remainder were grouped as ‘other’. As shown in Table 16, which reports the results for the Parisian data, the only individual city to be mentioned at a frequency above 15 was *Paris* itself. All other cities were marginal (for example, *Angers* and *Belloy-en-France* occurred just once), and this ‘other’ category is essentially a knock-out context: use of *sur* with a city other than Paris is exceptionally rare. With *Paris*, in contrast, the innovative preposition has a rate of occurrence of roughly 34%. In short, it seems that within the city itself, *Paris* is effectively the only city name with which *sur* appears.

Table 16. Distribution of *sur* by city name inside Paris

	% (<i>sur</i>)	n (<i>sur</i>)	n (<i>sur</i> +à)
<i>Paris</i>	33.8	22	65
Other	2.7	1	37
Total		23	102

The results for the suburban data are reported in Table 17, and they reveal that this dataset is more lexically robust than the urban one. Six different city names have token frequencies of at least 15 in the corpus. Three city names in particular display extremely high rates of *sur*: *Saint-Ouen*, *Ivry(-sur-Seine)*, and *Paris*. The effect of *Paris* is perhaps to be expected, on the basis of Table 13—there is a clear collocational preference for this string. Notably, *Ivry(-sur-*

Seine) and *Saint-Ouen* are the same two suburban locations where frequency of *sur* was markedly high (cf. Table 8).

Table 17. Distribution of *sur* by city name in the Parisian suburbs

	% (<i>sur</i>)	n (<i>sur</i>)	n (<i>sur</i> +à)
<i>Saint-Ouen</i>	58.8	20	34
<i>Ivry(-sur-Seine)</i>	47.4	27	57
<i>Paris</i>	42.6	43	101
<i>Montreuil</i>	14.3	6	42
<i>Kremlin(-Bicêtre)</i>	3.7	1	27
<i>Suresnes</i>	0.0	0	24
Other	12.4	12	97
Total		109	382

As noted in chapter 3, I wondered if *sur* specifically occurred with the name of the major urban city in the speaker's life (here, *Paris*), as well as the name of the city in which they live. Looking at the cross-tabulation of speaker's location and city name, reported in Table 18, we find that this hypothesis is supported: *sur* is found almost exclusively either in the context of the speaker's own city or with *Paris*. This is particularly evident for speakers living in Ivry-sur-Seine and Saint-Ouen, which are the locations in which the use of *sur* is frequent (cf. Table 8). In the other suburban cities, where *sur* is infrequent overall, the effect is less clear.

Table 18. Distribution of *sur* by speaker's city and lexical city name in the Parisian suburbs

Lexical city name	Speaker's city	Ivry-sur-Seine	Rosny-sous-Bois	Saint-Ouen	Le Kremlin-Bicêtre	Montreuil	Suresnes	Total
<i>Ivry(-sur-Seine)</i>	% (<i>sur</i>)	47.4						57
	n (<i>sur+à</i>)	57						
<i>Saint-Ouen</i>	% (<i>sur</i>)			58.8				34
	n (<i>sur+à</i>)			34				
<i>Kremlin (-Bicêtre)</i>	% (<i>sur</i>)	0.0			3.9			27
	n (<i>sur+à</i>)	1			26			
<i>Montreuil</i>	% (<i>sur</i>)	0.0				14.6		42
	n (<i>sur+à</i>)	1				41		
<i>Suresnes</i>	% (<i>sur</i>)						0.0	24
	n (<i>sur+à</i>)						24	
<i>Paris</i>	% (<i>sur</i>)	78.1	60.0	81.8	7.7	4.8	0.0	101
	n (<i>sur+à</i>)	32	10	11	26	21	1	
Other	% (<i>sur</i>)	0.0	38.1	7.7	0.0	33.3	4.4	97
	n (<i>sur+à</i>)	13	21	13	2	3	45	
Total		104	31	58	54	65	70	382

4.2.4 Summary of the distributional results for the CFPP2000 data

To summarize the findings from the CFPP2000 materials, it is clear that Paris and its suburbs exhibit quite distinct patterns of use for *sur* in the PCN context. For the urban centre, *sur* is found only among a handful of speakers and a handful of contexts (temporary situations, motion verbs, and in collocation with *Paris*). The effects of social factors in this locale are not clear, but the findings suggest that, inside Paris, *sur* is not an incoming variant within the PCN context overall: it does not increase monotonically over time and in social and linguistic contexts. *À* remains the majority variant, and *sur* occupies a niche in the system, appearing only in specific linguistic domains, notably in collocation with motion verbs with a rate of 50%. In the suburbs, *sur* seems to be more clearly socially stratified, with significant effects of both gender and SES, and it is also linguistically constrained by at least two factors (type

of situation and city name). Its use has also persevered more robustly than inside Paris, though beginning with speakers born in the 1970s, it may be retreating here as well. I now turn to the distributional results for Orléans, to see what further light they can shed on variation in the PCN context.

4.3 Distributional results for the ESLO data

4.3.1 Overall distribution and longitudinal trajectory

In the Orléans data ($N_{sur+\hat{a}} = 1658$), the newer variant occurs 14.9% of the time overall, which represents 247 tokens. Once again, this is the same rate that obtained for both corpora together (see section 4.1) and for the aggregate CFPP2000 data (see section 4.2.1). On the surface, this would seem to suggest that the feature is acting similarly, but as we will see in the factor-by-factor distribution, its variable grammar in Orléans displays differences from what was found inside Paris and in the Parisian suburbs, starting with the longitudinal trajectory. This is presented in Figure 5 and Table 19.

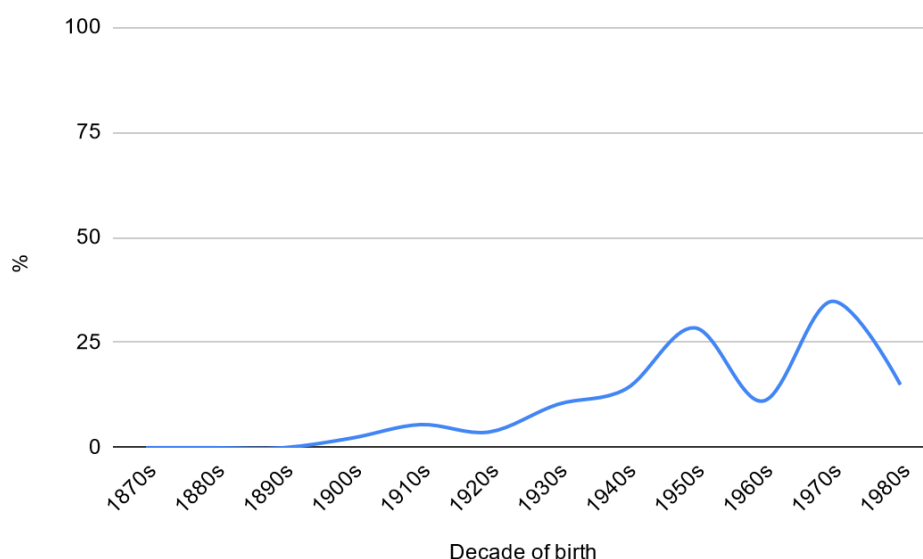


Figure 5. Distribution of *sur* by decade of birth in the ESLO data

Table 19. Distribution of *sur* by decade of birth in the ESLO data

	% (<i>sur</i>)	n (<i>sur</i>)	n (<i>sur</i> +à)
1870s	0.0	0	2
1880s	0.0	0	2
1890s	0.0	0	32
1900s	3.3	2	61
1910s	5.5	7	128
1920s	3.7	6	161
1930s	10.3	20	195
1940s	14.0	47	336
1950s	28.5	53	186
1960s	11.1	23	208
1970s	34.8	65	187
1980s	15.0	24	160
Total		247	1658

Whereas the Parisian corpus included only speakers born in the 1920s and later, the Orléans corpus includes material from speakers born in the late 19th century, starting in the 1870s. Note that the longitudinal trajectory in the Parisian data (Figure 2) suggested that *sur* was relatively recent, with the first tokens appearing among speakers born in the 1940s, but I was cautious about interpreting this finding, considering the low token counts for the two previous decades and historical attestations that suggest it was a 19th century innovation. Figure 5 shows that *sur* is attested in the PCN context in Orléans starting among speakers born in the first decade of the 20th century.²² In other words, these materials provide evidence that the feature is markedly older than the CFPP2000 data would suggest, adding at least forty years to its timeline. Indeed, there is reason to be cautious here as well: again, the data from speakers born in the decades leading up to 1900 is sparse. It is possible that *sur* is older yet,

²² Interestingly, one of these early users of *sur* (born in the 1900s) grew up in Paris, and only moved to Orléans when he was 42 years old. Another 1900s speaker was born in Tourcoing (in the North of France, near the Belgian border), and arrived in Orléans when she was 53. Therefore, we can postulate that the feature was already in use both in the Parisian region and the North of France at the beginning of the 20th century, an interpretation that is consistent with historical attestations (cf. Dory, 1878 for francophone Belgium).

and that this corpus fails to capture incipient use prior to the first tokens among those born at the turn of the century.

Just like in the capital and its suburbs, the use of *sur* seems to have risen quickly among speakers born in the 1950s, with a clear peak in this decade. The recurrence of this decade as a point of expansion across all three subsets of the data (city of Paris, Parisian suburbs, Orléans) suggests a bona fide trend in the history of this form. After this short period, however, its use is again unstable and its trajectory is quite turbulent: the high in the 1950s is followed by an immediate drop in the 1960s, an even higher peak in the 1970s, where the feature reaches a rate of almost 35%, before dropping again in the most recent data.

For this corpus, and as I did for CFPP2000, I considered the distribution of *sur* by speaker. This sheds further light on the trajectory of this form. Indeed, as was found in the urban Parisian data, not all speakers use *sur*. Table 20 (which does not include the data from speakers born from the 1870s to the 1890s, as they all use *à* categorically) reveals that the variant is used sporadically by individuals in the early decades of the 20th century. It is then progressively implemented over the following decades, until, among those born in the 1960s and 1970s, it is used by 100% of the sampled speakers. Notably, the rate of speaker use drops almost by half in the 1980s, and this perspective on individual speaker use corroborates the interpretation that was given for Figure 5: *sur* may be falling out of use. This same pattern is evident in the data from CFPP2000.

Table 20. Distribution of *sur* by speaker and decade of birth in the ESLO data

DoB	Speakers, N	Speakers who use <i>sur</i> , %	Overall frequency of <i>sur</i> , %	Total N
1900s	9	33.3	4.1	73
1910s	7	14.3	5.2	116
1920s	11	27.3	3.7	161
1930s	12	50.0	10.3	195
1940s	16	50.0	14.0	336
1950s	9	66.7	28.5	186
1960s	6	100.0	11.1	208
1970s	8	100.0	34.8	187
1980s	7	57.1	15.0	160
Total	85			1622

As with the Parisian data, in order to focus on the variable grammar of speakers who used *sur*, I have excluded the tokens from all the speakers who used *à* categorically (see section 4.2.1). This brings the new total of tokens to 1145, among which 21.6% are tokens of *sur*. There is thus a remarkable similarity across all three datasets: for speakers who use *sur*, whether it appears idiolectal (Paris city) or incorporated in local vernaculars (Parisian suburbs, Orléans), its frequency seems to be in the range of 22–29% overall and declining in recent decades.

I now go over the distributional results for the remaining social factors, before moving on to those for the linguistic factors.

4.3.2 Social factors

I begin the discussion of social factors with gender. The distribution is found in Table 21.

Unlike in the Parisian suburbs (I set aside the city data on this, for the reasons discussed in section 4.2.2), it is men, not women, who use *sur* more frequently. This difference is significant (X^2 (df 1, $N = 1145$) = 22, $p = < .00001$).

Table 21. Distribution of *sur* by gender in the ESLO data

	% (<i>sur</i>)	n (<i>sur</i>)	n (<i>sur</i> + <i>à</i>)
Women	15.8	90	570
Men	27.3	157	575
Total		247	1145

However, the story is somewhat more complex than this result suggests. A cross-tabulation of gender with decade of birth, shown in Figure 6 and Table 22, sheds more light on gender patterns in this dataset. With the exception of the 1910s (a cell which contains only 18 tokens) and a slight drop between the 1940s and the 1950s, women display an incremental and monotonic increase of *sur* over time, a trend that is consistent through to the youngest speakers, born in the 1980s. The trajectory for men is more difficult to interpret. There are no tokens from men born in the 1910s and 1920s, which obscures part of the story. For the decades in which tokens are available, the results are inconclusive: men sometimes pattern similarly to women (1900s and 1960s), they are sometimes ahead (1940s, 1950s and 1970s), and sometimes behind (1930s and 1980s). One thing that should be noted, however, is that the apparent decline in the use of *sur* that was visible in the aggregate data of Figure 5 seems to be driven entirely by men, while rates of use by women remain on the rise. Thus, the feature might not be waning in Orléans after all. Overall, however, due to these gaps and irregular frequencies of use among men, the gender patterns in this dataset remain difficult to interpret.

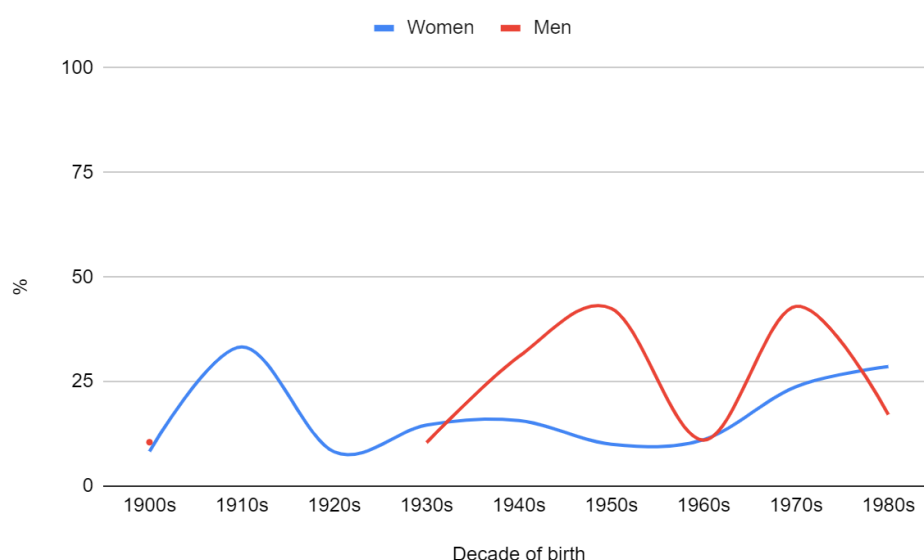


Figure 6. Distribution of *sur* by decade of birth and gender in the ESLO data

Table 22. Distribution of *sur* by decade of birth and gender in the ESLO data

	Women			Men		
	% (<i>sur</i>)	n (<i>sur</i>)	n (<i>sur</i> +à)	% (<i>sur</i>)	n (<i>sur</i>)	n (<i>sur</i> +à)
1900s	8.3	1	12	10.5	2	19
1910s	33.3	6	18			
1920s	8.2	6	73			
1930s	14.6	15	103	10.4	5	48
1940s	15.7	16	102	31.0	31	100
1950s	10.0	5	50	42.5	48	113
1960s	11.1	10	90	11.0	13	118
1970s	23.8	19	80	43.0	46	107
1980s	28.6	12	42	17.1	12	70
Total		90	570		157	575

Moving to socio-economic status, reported in Table 23, we find that, as with the CFPP2000 data, *sur* is most frequent among middle class speakers, again following the curvilinear principle.²³ However, this tendency is only marginal (the frequencies are all quite similar), and a chi-square test with Yates correction reveals that the differences between these rates are not significant (X^2 (df 2, N = 1145) = .77, p = .680407).

Table 23. Distribution of *sur* by socio-economic status in the ESLO data

	% (<i>sur</i>)	n (<i>sur</i>)	n (<i>sur</i> +à)
Upper class	19.6	37	189
Middle class	22.6	107	473
Working class	21.3	103	483
Total		247	1145

To summarize the findings for social factors in the ESLO data, the results for both gender and SES are inconclusive: the gender patterns are difficult to interpret, while SES exhibits a familiar yet non-significant effect. Moreover, looking at the cross-tabulation of gender and

²³ The separate results for the education and occupation factors in Orléans can be found in tables A5 and A6 in Appendix A. The results do not pattern clearly, but there seems to be a tendency for speakers that were given median scores for these two factors to be slightly more favourable to the use of *sur*.

SES in Table 24, the results are further obscured by the fact that women and men pattern in opposite ways: the men follow the curvilinear principle, while among the women, the middle class is actually the group where *sur* is the least frequent. This, combined with the temporal trajectory in Figure 6, suggests that the social patterns associated with *sur* in the PCN context may be shifting over time. Further investigation, with a different dataset, is needed.²⁴

Table 24. Distribution of *sur* by gender and socio-economic status in the ESLO data

	Women	Men
Upper class	15.6%	23.2%
Middle class	10.2%	28.8%
Working class	18.6%	26.9%

I now move on to distributional results for the linguistic constraints.

4.3.3 Linguistic factors

The results for the first linguistic factor, type of situation, are presented in Table 25. In the Parisian data, *sur* was more frequent with temporary situations than with other types of situations. Here, however, there is virtually no difference between the two contexts, suggesting that this predictor has no significant effect in the Orléans materials.

Table 25. Distribution of *sur* by type of situation in the ESLO data

	% (<i>sur</i>)	n (<i>sur</i>)	N (<i>sur</i> +à)
Unambiguously temporary situations	21.9	113	517
Other situations	21.3	134	628
Total		247	1145

²⁴ Section 2.5 briefly reviews some of the prescriptivist commentary that has been made by the Académie française on the use of *sur* in the PCN context but to my knowledge, there has been no study focusing on speakers' perception and language attitudes associated with this feature. Therefore, although it seems that the Orléans results may point to shifting social ideologies, such an interpretation of the current results remains speculative. Future research could explore this question in more detail.

The results for type of activity are displayed in Table 26. They show that *sur* is more frequent in contexts emphasizing human activities than in those that make no reference to such activities. This is consistent with the results for CFPP2000, both urban and peri-urban.

Although the Parisian results did not reach the level of significance, those for Orléans do (X^2 (df 1, N = 1145) = 4.2, p = .039868).

Table 26. Distribution of *sur* by type of activity in the ESLO data

	% (<i>sur</i>)	n (<i>sur</i>)	n (<i>sur</i> +à)
Human activities	24.0	142	592
Other activities	19.0	105	553
Total		247	1145

The results for type of trajector are displayed in Table 27. As with the results in the CFPP2000 materials, there is no effect of human versus non-human trajectors; the frequencies are effectively identical.

Table 27. Distribution of *sur* by type of trajector in the ESLO data

	% (<i>sur</i>)	n (<i>sur</i>)	n (<i>sur</i> +à)
Human	21.4	123	575
Non-human	21.8	124	570
Total		247	1145

The next linguistic constraint is type of verb; the results are presented in Table 28. There is no effect of motion verbs. This is the same result that was found for the Parisian suburbs, contrasting these two locales (Orléans and the suburbs) with the city centre, where *sur* was significantly more frequent with motion verbs than in other contexts.²⁵

²⁵ Just like for the suburban data, I did a cross-tabulation of decade of birth and type of verb (see Figure A1 in Appendix A). In Orléans, unlike in the suburbs, there is no clear longitudinal patterning for this factor. For example, in the 1950s, motion verbs are a more favourable context to *sur* than other contexts, but in the 1970s the opposite obtains.

Table 28. Distribution of *sur* by type of verb in the ESLO data

	% (<i>sur</i>)	n (<i>sur</i>)	n (<i>sur</i> +à)
Modifies a motion verb	21.7	52	240
Modifies something else	21.5	195	905
Total		247	1145

The final linguistic constraint is city name; the distributional results are presented in Table 29.

The first section of the table (in blue) displays results for *Orléans* and other local city names.

The second section (in green) gives results for other city names in France. Finally, the ‘other’ category contains city names which appeared fewer than 15 times in the data.

Table 29. Distribution of *sur* by city name in the ESLO data

	% (<i>sur</i>)	n (<i>sur</i>)	n (<i>sur</i> +à)
<i>Orléans</i>	30.8	143	465
<i>Saint-Jean-de-Braye</i>	22.2	6	27
<i>Fleury-les-Aubrais</i>	19.2	5	26
<i>Olivet</i>	19.0	11	58
<i>Saint-Jean-le-Blanc</i>	6.7	1	15
<i>Saint-Jean-de-la-Ruelle</i>	5.3	1	19
<i>Paris</i>	26.0	50	192
<i>Tours</i>	10.7	6	56
<i>Blois</i>	0.0	0	31
Other	8.9	24	270
Total		247	1145

The most frequent city name overall (*Orléans*) is also the name that co-occurs with *sur* the most frequently. Considering that *Orléans* is the main urban centre for the speakers in this dataset, this result can be compared with what was found for *Paris* in the CFPP2000 data. The other city name that represents a significant part of the data is *Paris* itself (n = 192). With this

city name, the rate of *sur* is 26%, which is more than the average frequency of the variant in the overall data (21.6%).²⁶

From these findings, it could be argued that the use of *sur* is more frequent with city names that are familiar to the speaker (e.g. *Orléans*, as well as other cities that are part of the Greater Orléans area, such as *Saint-Jean-de-Braye*), or that re-occur frequently in speech due to their status (e.g. *Paris*). As a reminder, a similar phenomenon was found in the suburban data, in which speakers favoured the use of *sur* with both *Paris* and the name of their own city. For the Orléans data, however, such a finding cannot be confirmed because the corpus does not provide precise information on the speaker's location within the Orléans area.

4.3.4 Summary of the distributional results for the ESLO data

To summarize the results from the Orléans data, the *sur* variant displays an unstable trajectory over time, a trajectory that seems to be an effect of patterns of usage among men. There is no consistent effect of SES or gender. Regarding the linguistic predictors, it was found that, with the exception of type of activity and city name, most have no visible effect. In other words, consistent with a pattern of geographic diffusion, linguistic constraints in Orléans seem to be weaker than in Paris and its suburbs.

4.3.5 Summary of the distributional findings

Table 30 presents a summary of the distributional results by factor in the three datasets.

Comparing the columns, it is clear that there are more inconsistencies across Paris, its suburbs, and Orléans than there are consistencies (where the data allows meaningful analysis).

Gender, location and time are difficult to assess, given token distributions and interactions with other predictors. The only effect that appears across all three locales is the effect of city

²⁶ Note, however, that a chi-square test with Yates correction indicates that the difference between these two frequencies is not significant: X^2 (df 1, N = 1337) = 1.9, p = .167963.

name in the PCN construction. Given the nature of the effect, which foregrounds *Paris* and, if not in this city, other major centres in the speaker's life, there is some predictable fluctuation but none that changes the overall interpretation of this factor. Type of trajectory does not seem to have an effect anywhere, another consistency. Among the linguistic predictors that seem to constrain variation, however, the hallmark of diffusion is evident. Indeed, as we move outward from the city centre of Paris (where, nonetheless, use of *sur* is sporadic rather than embedded in the vernacular), fewer constraints appear to obtain: type of situation does not persevere beyond the Parisian suburbs, and verb type obtains only in urban Paris.

Table 30. Summary of the distributional results in the three datasets

	Paris	Parisian suburbs	Orléans
Location	Inconclusive	✓ (Saint-Ouen + Ivry-sur-Seine > other cities)	✓
Gender	Inconclusive	✓ (women > men)	✓ (men > women, but confounding longitudinal patterns)
SES	✓ (curvilinear)	✓ (curvilinear)	✗
Type of situation	✓ (temporary > other)	✓ (temporary > other)	✗
Type of activity	✗	✗	✓ (human activities > other)
Type of trajectory	✗	✗	✗
Type of verb	✓ (motion > other)	✗	✗
City name	✓ (<i>Paris</i> > other)	✓ (<i>Paris</i> + speaker's city > other)	✓ (<i>Orléans</i> + Greater Orléans cities + <i>Paris</i> > other)

All these findings need to be confirmed by a mixed-effect analysis, which is presented in the next subsection.

4.4 Inferential statistics: Modelling variation in the PCN context

A mixed-effect analysis enables us to determine which factors (either categorical or continuous) are selected as significant in each of the three locales, and to elucidate which model best fits the data, thereby exposing the variable grammar. As noted in chapter 3, to perform this analysis, I used the program Rbrul (Johnson, 2008), which operates in the R environment.

Both the social and the linguistic predictors were modelled identically for each locale, to the extent possible, though certain deviations were required on the basis of the datasets. I did this to facilitate comparison between the locales. The social predictors were modelled as follows. For every model, individual speaker was run as a random effect and SES was run as a categorical ternary factor (lower, middle, upper). Gender as a categorical binary factor and decade of birth as a continuous factor were added as potential predictors for the suburban and the ESLO datasets, but not for the Paris one, since the relevant cells were too sparsely populated for inferential modelling. Regarding linguistic predictors, they were all run as categorical factors, but some recoding had to be done for the city name factor in the suburbs and Orléans. Indeed, in order for this factor to yield clearer results, city names that had similar distributional patterns were grouped together. In the suburban dataset, I made a binary distinction between *Paris*, *Ivry-sur-Seine*, and *Saint-Ouen* on the one hand, and all the other city names on the other hand. In Orléans, a ternary distinction was made: *Orléans*, *Fleury-les-Aubrais*, *Saint-Jean-de-Braye*, and *Olivet* as one group, *Paris* by itself, and then the remaining city names.

The best model for Paris is reported in Table 31, where the only factors that were selected as significant are linguistic factors. Based on what was found in the distributional analyses, it is no surprise that the effect of city name is the strongest (range = 66), followed by type of verb and type of activity. A more surprising result is that type of situation is not

selected, even though it displayed a seemingly robust effect in the distributional analysis (see Table 9). Cross-tabulations of this factor with the others in the model (not shown here) reveal that type of situation interacts in this data with two factors: type of activity and city name. These interactions can probably be explained by the fact that the number of tokens is very low (fewer than 10) in some of these cells—and specifically in the cells where *sur* is expected to be more, not less, frequent. As pointed out by Guy (1980), “an acceptable, reproducible level of accuracy is not obtained until each factor has at least 10 tokens representing it” (p. 20). Elsewhere, type of situation has a consistent effect, with temporary situations being a favourable context. Therefore, the fact that this predictor is not selected as part of the best model is probably a consequence of data sparsity rather than it not having a bona fide effect on the use of the variant. Thus, I take type of situation to be a critical part of the system in Paris, despite the results obtained in Table 31.

Table 31. Best mixed-effects model of predictors operating on *sur*: Paris

Application value [sur]	Factor	Log odds	Tokens	Response proportion	Factor weight
City name $p < 0.01$	<i>Paris</i>	1.587	65	0.338	0.83
	Other	-1.587	37	0.027	0.17
Type of verb $p < 0.01$	motion	0.844	18	0.500	0.70
	Other	-0.844	84	0.167	0.30
Type of activity $p < 0.1$	human activities	0.65	25	0.240	0.66
	no reference to activities	-0.65	77	0.221	0.34
Not significant:	type of situation, type of trajector, SES				
Model	deviance 83.524	AIC 93.524	df 5	intercept -1.454	speaker random standard deviation: 0

From these findings as well as what the distributional analysis showed, we can conclude that, among the few Parisian speakers who use *sur* as part of their idiolect, the variable grammar of this feature is robustly constrained by linguistic factors, and overarchingly so by the name of the city itself: *Paris*. *Sur* is also favoured with motion verbs, in contexts referring to human activities, and for temporary situations.

Table 32 reveals that, in the Parisian suburbs just like inside Paris, the first factor to be selected in the model is city name. Type of situation also emerges as a robust effect, but unlike in the city centre, neither type of activity nor type of verb is selected. Corroborating the distributional analyses, two social factors are selected as significant: gender, with women favouring *sur*, and SES, with the middle and working class groups favouring it as well. Thus, the variable grammar in the suburbs evidences social factors that were unobservable in the city centre due to the nature of the data, while being constrained by fewer linguistic factors.

Table 32. Best mixed-effects model of predictors operating on *sur*: Parisian suburbs

Application value [sur]	Factor	Log odds	Tokens	Response proportion	Factor weight
City name $p < 0.01$	<i>Paris + Ivry-sur-Seine + Saint-Ouen</i>	0.775	192	0.469	0.69
	Other	-0.775	190	0.100	0.32
Type of situation $p < 0.01$	Temporary	0.638	125	0.408	0.65
	not temporary	-0.638	257	0.226	0.35
SES $p < 0.05$	middle class	1.337	104	0.615	0.80
	working class	0.379	208	0.207	0.59
	upper class	-1.756	70	0.029	0.15
Gender $p < 0.1$	Women	0.508	237	0.350	0.62
	Men	-0.508	145	0.179	0.38
Not significant:	decade of birth, type of activity, type of verb, type of trajectory				
Model	deviance 321.705	AIC 335.705	df 7	intercept -1.639	
	speaker random standard deviation: 0.575				

Finally, in Orléans, only two linguistic factors are selected as significant: city name and type of activity. They follow the same directions of effect as in Paris and its suburbs: *Paris* remains a key context, alongside other more local city centres, as well as contexts referring to human activities. Regarding social factors, gender is significant, with men favouring *sur*, as well as decade of birth. Notably, SES is not selected, confirming that the effect of social class is not robust in Orléans.

Table 33. Best mixed-effects model of predictors operating on *sur*: Orléans

Application value [sur]	Factor	Log odds	Tokens	Response proportion	Factor weight
City name $p < 0.01$	<i>Orléans + Saint-Jean-de-Braye + Fleury-les-Aubrais + Olivet</i>	0.666	576	0.286	0.66
	<i>Paris</i>	0.457	192	0.260	0.61
	Other	-1.123	377	0.085	0.25
Decade of birth $p < 0.05$		0.154			
Gender $p < 0.05$	Men	0.303	575	0.273	0.58
	Women	-0.303	570	0.158	0.43
Type of activity $p < 0.1$	human activities	0.158	592	0.240	0.54
	no reference to activities	-0.158	553	0.190	0.46
Not significant:	type of situation, type of trajector, SES				
Model	deviance 1015.178	AIC 1051.178	df 7	intercept -2.938	
	speaker random standard deviation: 0.769				

To conclude on the inferential analyses across the three datasets, the city name factor is the only one to be selected in all locations, with a similar effect: *Paris* and the local city names are a favouring context for the use of *sur*, while other city names are not.

With the four linguistic predictors selected across the models, the picture that emerges is that the variable grammar of Paris is the most complex in terms of linguistic constraints. In the suburbs and Orléans, only two linguistic predictors appear to operate. Such a result supports the hypothesis that *sur* has diffused from one big urban centre (Paris) to smaller and more distant ones, with some of its linguistic constraints weakening or falling out of operation in the process.

The same does not obtain for social factors. Indeed, inside Paris, the variable grammar is linguistically constrained, with no social factor having a robust effect among *sur* users. In the suburbs and in Orléans, it is sociolinguistically constrained, with significant effects of both linguistic and social factors. As discussed in the next section, this seems to reveal an important contrast between Paris on the one hand, and Orléans and the suburbs on the other

hand, a contrast that could be explained by a variety of geographic, social, economic and ethnic factors.

Therefore, these results confirm that, in Paris, the variant is only used by a handful of speakers and confined to a few linguistic contexts. In the suburbs and in Orléans, its use is more widely spread across speakers. It is also socially stratified but not as strongly constrained by linguistic factors than inside Paris, meaning that it appears in a wider variety of contexts, a possible effect of geographic diffusion.

Chapter 5. Discussion

5.1 Implications of the study

An unexpected but critical finding of this study is how irregular the use of *sur* is across speakers, a phenomenon that is summarized in Table 34. The first observation that can be made is that only about half of the speakers who use the PCN construction in these materials use *sur* before city names—the other half use *à* categorically, suggesting that *sur* + city name is either not part of their idiolect, or that they simply did not produce it during their interview.

Table 34. Distribution of *sur* across speakers in all datasets

Location	Speakers who use <i>sur</i> , %	Speakers, N
All	43.7	151
Paris	20.0	40
Suburbs	65.0	20
Orléans	49.5	91

Moreover, there is a clear split in speaker use between the three areas under study. In Paris, only 8 out of the 40 speakers (20% of them) use *sur*. This is in strong contrast with the Parisian suburbs, where the majority of the speakers (65%) do. This finding is especially striking considering that Paris and its suburbs are part of the same geographical area, suggesting that there is a real contrast between the urban and peri-urban settings.

This split can likely be explained by differences in economic, social and ethnic factors. Indeed, Cheshire (2020) points out that “the centre of Paris is the preserve of the richer sections of the population, and contrasts vividly with the bleaker environment of the *banlieues* [suburbs]” (p. 22). This phenomenon can be traced back to French housing policies from the 1940s and 1950s, which sought to provide low-cost housing to immigrant populations at the edge of big cities, and contributed to “separating the richer and poorer sections of the population of Paris” (Cheshire, 2020, p. 22). Moreover, in their study on youth language attitudes, Secova et al. (2018) show that their teenage participants, all living in the poorer

suburban areas of Paris, are not only aware of such a split, but that it is part of their identity in that they explicitly seek to distinguish themselves from Parisians: “[i]n terms of identity, most participants in this study similarly reject what they see as the affluent ‘Parisian’ persona... The distinction is clearly felt between ‘posh’ people living in inner Paris and those in the *banlieues*, or more restrictively the *cités* (housing estates)... The differentiation is made not only along geographical and social lines, but also, to a great extent, along ethnic lines” (pp. 253–254). Although both Cheshire (2020) and Secova et al. (2018) specifically study the language of young people (born in the 1990s and 2000s), it can be assumed that older generations are also aware of this ideological and identity-based split between inner Paris and its suburbs.

With regard to *sur*, several hypotheses can be made to explain the different patterns of use that were found between the two areas: either suburban speakers are trying to distinguish themselves from Parisians by using the less prototypical preposition, or Parisians are trying to distinguish themselves from suburban (and provincial) populations by avoiding *sur* and using the more prototypical, and standard, preposition *à* almost categorically instead. Of course, these two hypotheses could also be true simultaneously. But as a reminder, it was also found that the use of *sur* is on the decline among younger generations in the suburbs (see Figure 3). It should be noted that the use of *sur* in the PCN context when it is not associated with a motion verb is criticized and condemned by the Académie française, and this could reveal a phenomenon of stigmatization of the variant, not only by prescriptivist institutions but by the Parisian society at large. Therefore, if it is indeed the case that the feature is stigmatized (to any extent) by Parisians, then this recent decline could suggest that younger suburban speakers are looking for class and geographic mobility, leading them to shift toward the speech of Parisians (cf. Labov, 2006). Alternatively, this could be an effect of the power and age imbalance of the interview setting in which their data was recorded. In any case, it would

be worth trying to find out how speakers of various social and economic contexts perceive the use of *sur*, and if they associate it with specific groups of the population, which would in turn explain these diverging patterns of usage. Moreover, as already pointed out in section 2.5, style may also be a factor of variation, which was not explored in the present study due to the data coming exclusively from transcriptions of sociolinguistic interviews aiming to record casual speech styles. Some sources, like the Académie française and Fairon et al. (2018), argue that the variant has spread to the mainstream media. Such claims could be verified by looking at data from a variety of sources, both written and oral (e.g. newspapers, radio and television programs, songs, etc.).²⁷

To further complicate the findings for the suburbs, it was found that two groups of cities display robust differences: Ivry-sur-Seine, Saint-Ouen and Rosny-Sous-Bois all have high rates of *sur* (between 45% and 52%), while the remaining cities (Montreuil, Le Kremlin-Bicêtre and Suresnes) have rates below 15%, which is below the average frequency inside Paris (22.5%). However, it should be remembered that the urban frequency is artificial, reflecting the use of a handful of *sur* users, and that, due to the low number of participants in some of the suburban locations (1 or 2 in the majority of them), it is difficult to disentangle how much of this variation is due to individual differences, versus differences in social, economic and ethnic contexts between these cities.

Regarding the third location, Orléans, it was found that around 50% (45 out of 91) of the speakers in the dataset use *sur*. This is more than what was found in Paris, but less than in the suburbs. Table 16 revealed that the feature seems to have been regularly implemented in Orléans through the decades, starting with only a handful of speakers in the earlier stages, and eventually reaching 100% of the speakers in the more recent decades. This finding

²⁷ As anecdotal evidence, I have mostly encountered the *sur* variant in casual registers where the speaker is using vernacular language (e.g. interviews, game shows, YouTube videos, rap songs, etc.), but I have rarely witnessed it in formal registers (either written contexts like press articles, or spoken contexts like news broadcasts).

corroborates a theory of change through diffusion, where the early adopters of the feature may have been mobile speakers who had been in contact with the point of origin of *sur* (potentially the Parisian region), and brought it back to Orléans with a weakened variable grammar until the feature was eventually transmitted to subsequent generations. This hypothesis is also supported by the fact that fewer linguistic predictors are selected in the statistical models (distributional, inferential) for Orléans than the one for Paris.

By further examining linguistic constraints, it was found that, inside Paris, the use of *sur* has specialized in a handful of contexts, especially motion verbs. This goes against what we would predict in cases of grammaticalization, where we would expect the variant to spread to more contexts (i.e. to generalize to all verb types), instead of being constrained to a niche in the system. In the suburbs and in Orléans, however, the verb type factor is not selected as significant. This, combined with the fact that *sur* is used with more city names there than in Paris, could point to a more grammaticalized preposition in these two locations, as it means that *sur* is more lexically productive. However, this loss of constraint could also be a consequence of diffusion. The effects of these two phenomena (grammaticalization vs. diffusion) are difficult to disentangle at present, but as already pointed out in the introduction, although extension of usage is a hallmark of grammaticalization, this seems to be the only property of this phenomenon that *sur* is exhibiting, at least from what we can gather from this study's design.

The city name predictor revealed an interesting phenomenon as well. Indeed, it was found that the name of the place in which the speaker lives, as well as *Paris* in the suburbs and Orléans, are consistently a very favouring context for the use of *sur*. Of course, this finding can be partly explained by a frequency effect, as the local city name and the capital *Paris* are more likely to appear in speech on a daily basis than other city names. However, it can also suggest that speakers are not only more likely to implement innovative variants in

familiar and regular contexts (*Paris*), but also contexts that are directly linked to their sense of identity and community (local city names). This goes back to Secova et al.'s (2018) observations on youth language and identity.

The initial research questions asked which factors could explain the variation between *à* and *sur*, and whether the hypotheses made by previous research were supported by quantitative findings. It was found that location was the first major factor of variation, as important differences were exposed across the three areas under study. A range of linguistic factors, some of them introduced by previous research—Hernández (2008) for type of situation, Huyghe (2012) for type of trajector, and Hernández (2008), Lismoen (2015) and Fairon et al. (2018) for type of verb—were found to be significant in all or some of the subsets of the data, confirming the hypotheses of these researchers. Additionally, the effects of social factors such as gender in the suburbs and Orléans, and SES in the suburbs, revealed that the variable grammar in these locations is not only linguistically but also socially constrained.

Finally, the longitudinal perspective provided by the mixture of real- and apparent-time analysis revealed that the implementation of the feature was already underway at the end of the 19th century, but that it does not appear to have progressed monotonically in any of the locations. Instead, it followed a trajectory of peaks and drops, with a notable rise in the 1950s across all locations, as well as in the 1970s (among women) in Orléans. This twenty-year delay between the Parisian region and Orléans further supports a hypothesis of diffusion of the feature from one locale to the other, with an acceleration in the second half of the century. Notably, the mid-20th century is the period that most previous literature gives as the starting point for the use of *sur* before city names in Hexagonal French (Grevisse & Goosse, 1993; Hernández, 2008; Huyghe, 2012; Lismoen, 2015). Again, the present study reveals that the use of *sur* is actually at least fifty years older than this, and possibly even older. A quick

search on Google Books Ngram Viewer reveals that rare attestations of *sur* before city names can be found as early as the 1810s, in military contexts and with motion verbs such as in (65), further supporting some of the findings and interpretations discussed here. Notably, such attestations are similar to what was found in Frantext for written sources from the 1980s onward (see footnote 11), suggesting that the feature may well have emerged in such contexts and remains restricted to such constructions in formal contexts, even though it has expanded to vernacular genres, where it has generalized across contexts, since then.

- (65) [Il] entre dans le golfe de Juan, et marche SUR Paris (Google Books Ngram Viewer: *Instruction sur les opérations de l'assemblée extraordinaire du Champ de Mai*, 1815)
 'He enters the gulf of Juan and walks over Paris'

5.2 Limitations and future research

An important limitation of this study is the fact that the majority of the Parisian data had to be excluded from the analysis on the basis that the speakers used *à* categorically. This left very few tokens for analysis for Paris, and therefore the picture of the variable grammar is potentially incomplete in this location. One wonders if results would have been different if other participants had been included in the corpus, and future work could explore this question by reproducing this study with another sample of speakers. It would also be worthwhile to interview speakers of different backgrounds about their own use of *sur*, for example through surveys. This is what Lismoen (2015) did, but with only 19 participants. This type of study could therefore be reproduced with a larger speaker sample. A linguistic attitude study could also be performed in order to find out if the variant is indeed stigmatized by certain groups of speakers.

Moreover, as explained above, due to the available data, style was not included in the analysis as a potential factor of variation, but future work could examine its role in order to

confirm if the change affecting *sur* is a change from below, with a spread from vernacular to more formal contexts. Certainly the SES evidence presented here supports such an interpretation, as does its highly constrained use among speakers within Paris, but further corroboration would nonetheless prove fruitful, if only to ascertain the ways in which social and linguistic predictors diffuse across styles and genres.

Another limitation is that *dans* tokens were excluded on the grounds that the use of this preposition before city names seems to have a function of emphasizing the centre of the city, and that variation with *dans* is therefore not possible in all linguistic environments. The same phenomenon occurs with the preposition *vers* ‘towards’, which is used with directional verbs and therefore alternates with *à* and *sur* in such contexts. This deserves a study in itself, to confirm on the basis of empirical evidence whether the variable grammar is indeed partitioned this way.

Finally, as mentioned in the introduction, the use of *sur* before city names is possibly part of a wider phenomenon in which the contextual and functional uses of this preposition are expanding, for example to indicate a duration as shown in (4) and repeated here in (66).

(66) c’est t-SUR deux jours je crois (ESLO2/RL437, b. 1970s)

‘It’s over the course of two days I think’

Therefore, future work could focus on this generalization of *sur* in vernacular practice, and confirm that the change seen in the PCN context is not in fact an isolated phenomenon, but rather part of a wider change that is currently underway in Hexagonal French.

References

Corpora

ATILF. (1998–2019) *Base textuelle Frantext*. ATILF–CNRS & Université de Lorraine.

Retrieved from <https://www.frantext.fr/>

Avanzi, M., Béguelin, M.-J., & Diémoz, F. (2012-2019). *Présentation du corpus OFROM – Corpus oral de français de Suisse romande*. Université de Neuchâtel. Retrieved from

<http://www.unine.ch/ofrom>

Branca-Rosoff, S., Fleury, S., Lefevre, F., & Pires, M. (2012). *Discours sur la ville*.

Présentation du Corpus de français parlé parisien des années 2000 (CFPP2000).

Retrieved from <http://cfpp2000.univ-paris3.fr/CFPP2000.pdf>

Eshkol-Taravella, I., Baude, O., Maurel, D., Hriba, L., Dugua, C., & Tellier, I. (2012). Un grand corpus oral « disponible » : le corpus d’Orléans 1968-2012, *Ressources linguistiques libres*, 52(3), 17–46. [https://halshs.archives-ouvertes.fr/halshs-](https://halshs.archives-ouvertes.fr/halshs-01163053/document)

[01163053/document](https://halshs.archives-ouvertes.fr/halshs-01163053/document)

Laboratoire ligérien de linguistique–UMR 7270 (LLL) (2018). ESLO-MD : *Enquêtes socio-linguistiques à Orléans : Corpus micro-diachronie* [Corpus]. ORTOLANG (Open Resources and TOols for LANGuage). Retrieved from

<https://hdl.handle.net/11403/eslo-md/v1>.

Trésor de la langue française informatisé (1994) ATILF–CNRS & Université de Lorraine. Retrieved from <http://www.atilf.fr/tlfi>

Texts

Académie française (n.d.). *Questions de langue. Sur Paris ?*

http://www.academie-francaise.fr/questions-de-langue#84_strong-em-sur-paris-em-strong

- Aldrich, R. (1996). *Greater France. A history of French overseas expansion*. London: Macmillan.
- Anthony, L. (2014). AntConc 3.4.3m [software]. Tokyo: Waseda University.
Available at: [http:// www.antlab.sci.waseda.ac.jp/antconc_index.html](http://www.antlab.sci.waseda.ac.jp/antconc_index.html).
- Armstrong, N., & Jamin, M. (2002). Le français des banlieues: Uniformity and discontinuity in the French of the Hexagone. In K. Salhi (Ed.), *French in and out of France. Language policies, intercultural antagonisms and dialogue* (pp. 107–136). Bern, Switzerland: Peter Lang.
- Ashby, W. (2001). Un nouveau regard sur la chute du *ne* en français parlé tourangeau: S’agit-il d’un changement en cours? *Journal of French Language Studies*, 11(1), 1–22. <https://doi.org/10.1017/S0959269501000114>
- Bailey, G., Wikle, T., Tillery, J., & Sand, L. (1993). Some patterns of linguistic diffusion. *Language Variation and Change*, 5(3), 359–390.
<https://doi.org/10.1017/S095439450000154X>
- Bigot, D., & Papen, R. (2020). Observing variation and change in Ontario French through internal, external and identity factors. *Journal of French Language Studies*. Advance online publication, 1–23. <https://doi.org/10.1017/S0959269520000186>
- Blondeau, H. (2001). Real-time changes in the paradigm of personal pronouns in Montreal French. *Journal of Sociolinguistics*, 5(4), 453–474.
<https://doi.org/10.1111/1467-9481.00160>
- Blondeau, H., Mougeon, R., & Tremblay, M. (2018). Analyse comparative de *ça fait que*, *alors, donc* et *so* à Montréal et à Welland: mutations sociales, convergences, divergences en français laurentien. *Journal of French Language Studies*, 29(1), 35–65. <https://doi.org/10.1017/S0959269518000169>

- Boughton, Z., & Pipe, K. (2020). Phonological variation and change in the regional French of Alsace: Supralocalization, age, gender and the urban–rural dichotomy. *Journal of French Language Studies*, 30(3), 327–353.
<https://doi.org/10.1017/S0959269520000125>
- Bybee, J., & Hopper, P. (2001). *Frequency and the emergence of linguistic structure*. Amsterdam/Philadelphia: John Benjamins.
- Cadiot, P. (1997). Les paramètres de la notion de préposition incolore. *Faits de langues*, 9, 127–134. https://www.persee.fr/doc/flang_1244-5460_1997_num_5_9_1148
- Cervoni, J. (1991). *La préposition. Étude sémantique et pragmatique*. Paris/Louvain-la-Neuve: Éditions Duculot.
- Cheshire, J. (2020). Taking the longer view: Explaining multicultural London English and multicultural Paris French. *Journal of Sociolinguistics*, 24(3), 308–327.
<https://doi.org/10.1111/josl.12385>
- Christy, C. (1983). *Uniformitarianism in linguistics*. Amsterdam/Philadelphia: John Benjamins.
- D’Arcy, A., & Tagliamonte, S. A. (2015). Not always variable: Probing the vernacular grammar. *Language Variation and Change*, 27(3), 255–285.
<https://doi.org/10.1017/S0954394515000101>
- Dendale, P., & De Mulder, W. (1997). Les traits et les emplois de la préposition spatiale *sur*. *Faits de langues*, 9, 211–220.
https://www.persee.fr/doc/flang_1244-5460_1997_num_5_9_1158
- Dory, I. (1878). *Wallonismes*. Liège: Vaillant-Carmanne.
- Fairon, C., Simon, A. C., & Grevisse, M. (2018). *Le petit bon usage de la langue française*. Louvain-la-Neuve/Paris: De Boeck Supérieur.

- Francard, M. (2018, September 28). Superfétatoire, ce SUR? *Le Soir*. Retrieved from <https://www.lesoir.be/>
- Goyens, M., Lamiroy, B., & Melis, L. (2002). Déplacement et repositionnement de la préposition *à* en français, *Linguisticae Investigationes*, 25(2), 275–300.
<https://doi.org/10.1075/li.25.2.06goy>
- Grevisse, M., & Goosse, A. (1993). *Le bon usage : grammaire française. Treizième édition*. A. Goosse (Ed.). Paris: Duculot.
- Guiraud, P. (1975). *L'ancien français*. (5th ed.). Paris: Presses Universitaires de France.
- Guy, G. G. (1980). Variation in the group and the individual: The case of final stop deletion. In W. Labov (Ed.), *Locating language in time and space* (pp. 1–36). New York: Academic Press.
- Hernández, P. C. (2008). La décoloration de la préposition *sur* : une explication en termes d'intégration conceptuelle. *Formes symboliques* 7, 1–19.
<http://www.formes-symboliques.org/2008/02/19/la-decoloration-de-la-preposition-sur-une-explication-en-termes-dintegration-conceptuelle-patricia-hernandez/>
- Hopper, P., & Traugott, E. (2003). *Grammaticalization*. Cambridge: Cambridge University Press.
- Huyghe, R. (2012). *Sur* et les noms de territoires. *Cuadernos de Filología Francesa*, 23, 169–186. https://www.academia.edu/6725101/Sur_et_les_noms_de_territoires
- Johnson, D. E. (2008). Getting off the GoldVarb standard: Introducing Rbrul for mixed-effects variable rule analysis. *Language and Linguistics Compass*, 3(1), 359–383.
<https://doi.org/10.1111/j.1749-818X.2008.00108.x>

- Insee. (2019). Historique des populations communales. Recensements de la population 1876–2017. Populations légales 2017. Retrieved from <https://www.insee.fr/fr/statistiques/3698339#consulter>
- Labov, W. (1972). *Sociolinguistic patterns*. Philadelphia: University of Pennsylvania Press.
- Labov, W. (1994). *Principles of linguistic change. Volume 1: Internal factors*. Oxford: Blackwell.
- Labov, W. (2001). *Principles of linguistic change. Volume 2: Social factors*. Oxford: Blackwell.
- Labov, W. (2006). *The social stratification of English in New York City* (2nd ed.). Cambridge: Cambridge University Press.
- Labov, W. (2007). Transmission and diffusion. *Language*, 83(2), 344–387.
<https://doi.org/10.1353/lan.2007.0082>
- Langacker, R. (1987). *Foundations of cognitive grammar. Volume 1*. Stanford: Stanford University Press.
- Le Bidois, R. (1965, June 26). De quelques emplois de “sur”. *Le Monde*. Retrieved from <https://www.lemonde.fr/>
- Lismoen, I. (2015). “*Je suis sur Paris cette semaine*” : une analyse de l’emploi de la préposition sur devant les noms de villes. (Unpublished master’s thesis). Norwegian University of Science and Technology, Norway.
- Marchessou, A. (2018). Strasbourg, another setting for sociolinguistic variation in contemporary French. *Journal of French Language Studies*, 28(2), 265–289.
<https://doi.org/10.1017/S095926951800008X>
- Melis, L. (2003). *Les prépositions en français*. Paris: Ophrys.
- Milroy, J. (1992). *Linguistic variation and change*. Oxford: Blackwell.

- Poplack, S. (1993). Variation theory and language contact. In D. Preston (Ed.), *Variation theory and language contact: American dialect research* (pp. 251–286). Amsterdam/Philadelphia: John Benjamins.
- Poplack, S. (2011). Grammaticalization and linguistic variation. In B. Heine and H. Narrog (Eds.), *Handbook of grammaticalization* (pp. 209–224). Oxford: Oxford University Press.
- Poplack, S., & Tagliamonte, S. (1999). The grammaticization of *going to* in (African American) English. *Language Variation and Change*, 11(3), 315–342.
<https://doi.org/10.1017/S0954394599113048>
- Proust, J. M. (2017, August 2). À tous ceux qui disent “sur Paris”. Slate. Retrieved from <http://www.slate.fr/story/148410/sur-vous-etes-sur>
- Sankoff, G., & Blondeau, H. (2007). Longitudinal change across the lifespan: /r/ in Montreal French. *Language*, 83(3), 560–588. <https://doi.org/10.1353/lan.2007.0106>
- Sankoff, G., & Thibault, P. (1977). L’alternance entre les auxiliaires “avoir” et “être” en français parlé à Montréal. *Langue française*, 34, *linguistique et sociolinguistique*, 81–108. https://www.persee.fr/doc/lfr_0023-8368_1977_num_34_1_4818
- Secova, M., Gardner-Chloros, P., & Atangana, F. (2018). ‘Il parle normal, il parle comme nous’: Self-reported usage and attitudes in a banlieue. *Journal of French Language Studies*, 28(2), 235–263. <https://doi.org/10.1017/S0959269518000078>
- Spang-Hanssen, E. (1963). *Les prépositions incolores du français moderne*. Copenhagen: G.E.C. Gads Forlag.
- Tagliamonte, S. A. (2006). *Analysing sociolinguistic variation*. Oxford: Cambridge University Press.

Tagliamonte, S. A. & Denis, D. (2014). Expanding the transmission/diffusion dichotomy:

Evidence from Canada. *Language*, 90(1), 90–136.

<http://www.jstor.org/stable/24672021>

Tagliamonte, S., Durham, M., & Smith, J. (2014). Grammaticalization at an early stage:

Future *be going to* in conservative British dialects. *English Language and Linguistics*,

18(1), 75–108. <https://doi.org/10.1017/S1360674313000282>

Thibault, A. (2004). *Dictionnaire suisse romand*. Chêne-Bourg: Zoé.

Tristram, A. (2021). Variation and change in future temporal reference with *avoir* and *être*.

Journal of French Language Studie, 31(1).

<https://doi.org/10.1017/S0959269520000174>

Trudgill, P. (1974). Linguistic change and diffusion: Description and explanation in

sociolinguistic dialect geography. *Language in Society* 3(2), 215–246.

<https://doi.org/10.1017/S0047404500004358>

Vandeloise, C. (1986). *L'espace en français*. Paris: Éditions du Seuil.

Vandeloise, C. (2017). Three basic prepositions in French and in English: A comparison.

Corela, HS-23. <https://doi.org/10.4000/corela.5033>

Appendix A

Additional tables and figures

Table A1. Distribution of sur by education inside Paris

	% (<i>sur</i>)	n (<i>sur</i>)	n (<i>sur</i> +à)
6	16.7	1	6
5	23.6	17	72
4	12.5	1	8
3	25.0	4	16
Total		23	102

Table A2. Distribution of sur by education in the Parisian suburbs

	% (<i>sur</i>)	n (<i>sur</i>)	n (<i>sur</i> +à)
6	2.9	2	70
5	29.5	23	78
4	54.4	43	79
3	31.4	11	35
2	45.2	14	31
1	18.0	16	89
Total		109	382

Table A3. Distribution of sur by occupation inside Paris

	% (<i>sur</i>)	n (<i>sur</i>)	n (<i>sur</i> +à)
5	16.7	1	6
4	20.0	9	45
3	27.9	12	43
1	12.5	1	8
Total		23	102

Table A4. Distribution of sur by occupation in the Parisian suburbs

	% (<i>sur</i>)	n (<i>sur</i>)	n (<i>sur</i> +à)
5	8.9	8	90
4	53.1	69	130
3	27.0	24	89
1	11.0	8	73
Total		109	382

Table A5. Distribution of *sur* by education in the ESLO data

	% (<i>sur</i>)	n (<i>sur</i>)	n (<i>sur</i> +à)
6	18.8	42	224
5	7.1	4	56
4	25.7	95	369
3	23.4	22	94
2	20.5	80	390
1	33.3	4	12
Total		247	1145

Table A6. Distribution of *sur* by occupation in the ESLO data

	% (<i>sur</i>)	n (<i>sur</i>)	n (<i>sur</i> +à)
5	14.6	24	164
4	19.4	76	391
3	28.6	112	392
2	16.9	21	124
1	18.9	14	74
Total		247	1145

Table A7. Distribution of *sur* by lexical verb in the CFPP2000 data

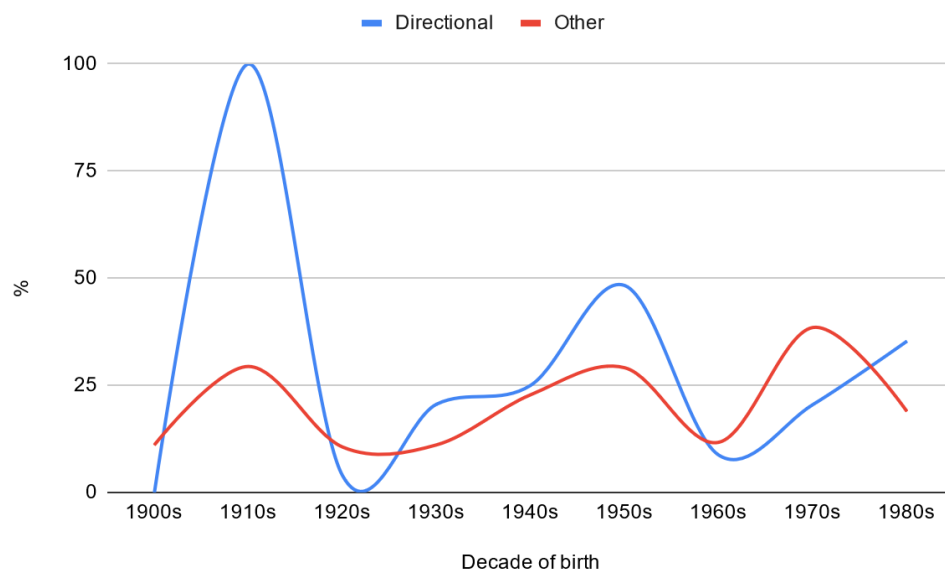
	Paris			Suburbs		
	% (<i>sur</i>)	n (<i>sur</i>)	n (<i>sur</i> +à)	% (<i>sur</i>)	n (<i>sur</i>)	n (<i>sur</i> +à)
<i>aller</i> 'to go' (M)				23.1	9	39
<i>habiter</i> 'to inhabit'	14.3	1	7	5.6	1	18
<i>être</i> 'to be'	0.0	0	12	33.3	18	54
<i>vivre</i> 'to live'	0.0	0	5	8.3	1	12
Other	35.5	11	31	39.1	27	69
N/A	23.4	11	47	27.9	53	190
Total		23	102		247	382

('M' stands for 'motion')

Table A8. Distribution of *sur* by lexical verb in the ESLO data

	% (<i>sur</i>)	n (<i>sur</i>)	n (<i>sur</i> +à)
<i>revenir</i> ‘to come back’ (M)	46.7	7	15
<i>travailler</i> ‘to work’	44.4	16	36
<i>venir</i> ‘to come’ (M)	39.1	9	23
<i>rester</i> ‘to stay’	40.0	6	15
<i>arriver</i> ‘to arrive’ (M)	22.6	7	31
<i>être</i> ‘to be’	21.2	32	151
<i>partir</i> ‘to leave’ (M)	14.3	3	21
<i>aller</i> ‘to go’ (M)	16.5	17	103
<i>habiter</i> ‘to inhabit’	8.1	3	37
<i>naître</i> ‘to be born’	0.0	0	19
<i>vivre</i> ‘to live’	0.0	0	14
Other	18.4	18	98
N/A	22.2	129	582
Total		247	1145

(‘M’ stands for ‘motion’)

Figure A1. Distribution of *sur* by decade of birth and type of verb in the ESLO data

Appendix B

List of city names

Table B1 presents the list of city names that were searched for during the extraction process.

The second column displays the number of tokens for each city name in the overall data

(before excluding the *dans* tokens). The third column shows the overall percentage that each

city name represents in the dataset. When the second and third columns are blank, this means

that the city name did not appear in the variable context under study.

Table B1. List of city names that were searched for

City name	n	%
<i>Agen</i>		
<i>Aix-en-Provence</i>	3	0.11
<i>Alfortville</i>		
<i>Amiens</i>	2	0.07
<i>Amsterdam</i>	1	0.04
<i>Angers</i>	2	0.07
<i>Annecy</i>	1	0.04
<i>Antony</i>	2	0.07
<i>Argenteuil</i>	1	0.04
<i>Asnières</i>		
<i>Aubervilliers</i>	2	0.07
<i>Aulnais</i>		
<i>Auxerre</i>		
<i>Bagneux</i>	2	0.07
<i>Bagnolet</i>	11	0.41
<i>Barcelone</i>	1	0.04
<i>Beaumont</i>		
<i>Beaugency</i>	3	0.11
<i>Beauvais</i>	1	0.04
<i>Belgrade</i>	1	0.04
<i>Belloy-en-France</i>	1	0.04
<i>Benghazi</i>	1	0.04
<i>Berck</i>	1	0.04

<i>Berlin</i>	5	0.19
<i>Biarritz</i>	1	0.04
<i>Blois</i>	32	1.19
<i>Boigny-sur-Bionne</i>		
<i>Bois-Colombes</i>	8	0.30
<i>Bondy</i>	5	0.19
<i>Bordeaux</i>	7	0.26
<i>Bou</i>		
<i>Boulogne</i>	10	0.37
<i>Bourg-en-Bresse</i>	1	0.04
<i>Bourges</i>	15	0.56
<i>Bourron-Marlotte</i>	1	0.04
<i>Brest</i>		
<i>Bricy</i>	1	0.04
<i>Bruxelles</i>	6	0.22
<i>Busy</i>		
<i>Bussy-Saint-Georges</i>		
<i>Caen</i>		
<i>Cannes</i>	3	0.11
<i>Cap Ferrat</i>		
<i>Casablanca</i>	1	0.04
<i>Cergy</i>		
<i>Chambord</i>	5	0.19
<i>Chamonix</i>	1	0.04
<i>Champigny</i>		
<i>Champigny-sur-Marne</i>	17	0.63
<i>Chanteau</i>	3	0.11
<i>Charbonnière</i>	1	0.04
<i>Chartres</i>	18	0.67
<i>Château-Chinon</i>	1	0.04
<i>Châteaudun</i>	2	0.07
<i>Châteauneuf-sur-Loire</i>	2	0.07
<i>Châteauroux</i>	5	0.19
<i>Châtenay-Malabry</i>	2	0.07
<i>Châtillon</i>		
<i>Châtillon-Montrouge</i>		
<i>Chavilles</i>	1	0.04

<i>Chécy</i>	14	0.52
<i>Choisy</i>	5	0.19
<i>Clamart</i>	2	0.07
<i>Clermont-Ferrand</i>	5	0.19
<i>Clichy</i>	1	0.04
<i>Cologne</i>	1	0.04
<i>Combleux</i>	1	0.04
<i>Combreux</i>	1	0.04
<i>Compiègne</i>	1	0.04
<i>Copenhague</i>	1	0.04
<i>Créteil</i>	6	0.22
<i>Deauville</i>	2	0.07
<i>Drancy</i>	5	0.19
<i>Draveil</i>		
<i>Dreux</i>	1	0.04
<i>Dublin</i>	1	0.04
<i>Dunkerque</i>	1	0.04
<i>Ecry</i>	1	0.04
<i>Epinay-sur-Seine</i>	2	0.07
<i>Evry</i>	2	0.07
<i>Fécamp</i>	4	0.15
<i>Fez</i>		
<i>Fleury-les-Aubrais</i>	33	1.23
<i>Fontainebleau</i>	2	0.07
<i>Fontenay-sous-Bois</i>	4	0.15
<i>Garches</i>	15	0.56
<i>Grenoble</i>	6	0.22
<i>Hong Kong</i>	1	0.04
<i>Ingrannes</i>	5	0.19
<i>Ingré</i>	10	0.37
<i>Issy</i>		
<i>Issy-les-Moulineaux</i>	2	0.07
<i>Ivry-sur-Seine</i>	61	2.27
<i>Joinville</i>	1	0.04
<i>Juvisy-sur-Orge</i>	2	0.07
<i>Kremlin-Bicêtre</i>	28	1.04
<i>La-Chapelle-Saint-Mesmin</i>	1	0.04

<i>La Rochelle</i>	5	0.19
<i>Las Vegas</i>	2	0.07
<i>Le Caire</i>	1	0.04
<i>Le Havre</i>	1	0.04
<i>Le Mans</i>	2	0.07
<i>Lille</i>	5	0.19
<i>Lisbonne</i>		
<i>Londres</i>	11	0.41
<i>Longjumeau</i>	5	0.19
<i>Lons-le-Saunier</i>		
<i>Lorient</i>	1	0.04
<i>Loury</i>	5	0.19
<i>Luxembourg</i>		
<i>Lyon</i>	7	0.26
<i>Madrid</i>	1	0.04
<i>Maisons-Alfort</i>	6	0.22
<i>Malakoff</i>	9	0.34
<i>Manchester</i>		
<i>Mantes</i>		
<i>Mantes-la-Jolie</i>		
<i>Marboué</i>	1	0.04
<i>Mardié</i>		
<i>Marigny-les-Usages</i>		
<i>Marseille</i>	9	0.34
<i>Megève</i>	1	0.04
<i>Mehun-sur-Yèvre</i>	2	0.07
<i>Melbourne</i>	1	0.04
<i>Melun</i>	3	0.11
<i>Metz</i>	2	0.07
<i>Meudon</i>	4	0.15
<i>Miami</i>	1	0.04
<i>Milan</i>	1	0.04
<i>Mitry</i>		
<i>Mitry-Mory</i>		
<i>Mont-Saint-Aignan</i>	4	0.15
<i>Montargis</i>	3	0.11
<i>Montgeron</i>	1	0.04

<i>Montigny</i>	1	0.04
<i>Montluçon</i>	1	0.04
<i>Montpellier</i>	12	0.45
<i>Montréal</i>	5	0.19
<i>Montreuil</i>	120	4.47
<i>Montrouge</i>		
<i>Mulhouse</i>		
<i>Mureaux</i>	1	0.04
<i>Nancy</i>	3	0.11
<i>Nanterre</i>	3	0.11
<i>Nantes</i>	13	0.48
<i>Neuilly</i>	4	0.15
<i>Neuilly-Plaisance</i>	2	0.07
<i>Neuville-aux-Bois</i>	3	0.11
<i>New York</i>	7	0.26
<i>Newcastle</i>	1	0.04
<i>Nice</i>	3	0.11
<i>Nîmes</i>	1	0.04
<i>Noisy</i>		
<i>Noisy-le-Grand</i>		
<i>Olivet</i>	63	2.35
<i>Orléans</i>	740	27.59
<i>Ormes</i>	10	0.37
<i>Paname</i>	1	0.04
<i>Paris</i>	737	27.48
<i>Perpignan</i>	1	0.04
<i>Perreux-sur-Marne</i>	2	0.07
<i>Pithiviers</i>	7	0.26
<i>Poitiers</i>	3	0.11
<i>Pornic</i>	1	0.04
<i>Prague</i>	1	0.04
<i>Puteaux</i>	6	0.22
<i>Quiberon</i>	1	0.04
<i>Rabat</i>	2	0.07
<i>Reims</i>	2	0.07
<i>Rennes</i>	4	0.15
<i>Reuilly</i>		

<i>Rohanne</i>	1	0.04
<i>Romainville</i>	7	0.26
<i>Rome</i>	5	0.19
<i>Romorantin</i>	3	0.11
<i>Rosny-sous-Bois</i>	16	0.60
<i>Roubaix</i>		
<i>Rouen</i>	4	0.15
<i>Rueil-Malmaison</i>	2	0.07
<i>Sables-d'Olonne</i>	3	0.11
<i>Saint-Amand-Montron</i>	1	0.04
<i>Saint-Antoine</i>		
<i>Saint-Ay</i>	7	0.26
<i>Saint-Benoît-sur-Loire</i>	1	0.04
<i>Saint-Bris-le-Vineux</i>	2	0.07
<i>Saint-Cloud</i>	7	0.26
<i>Saint-Cyr-en-Val</i>	5	0.19
<i>Saint-Denis</i>	8	0.30
<i>Saint-Denis-en-Val</i>	4	0.15
<i>Saint-Etienne</i>		
<i>Saint-Florent-sur-Cher</i>	1	0.04
<i>Saint-Georges-de-Didonne</i>	1	0.04
<i>Saint-Girons</i>	1	0.04
<i>Saint-Hilaire-Saint-Mesmin</i>	2	0.07
<i>Saint-Jean</i>	1	0.04
<i>Saint-Jean-de-Braye</i>	31	1.16
<i>Saint-Jean-de-la-Ruelle</i>	20	0.75
<i>Saint-Jean-de-Monts</i>	1	0.04
<i>Saint-Jean-le-Blanc</i>	18	0.67
<i>Saint-Malo</i>	3	0.11
<i>Saint-Mandé</i>		
<i>Saint-Mandrier</i>	1	0.04
<i>Saint-Maur</i>	2	0.07
<i>Saint-Nazaire</i>	1	0.04
<i>Saint-Ouen</i>	52	1.94
<i>Saint-Paul</i>		
<i>Saint-Pierre-le-Puellier</i>	1	0.04
<i>Saint-Pryvé-Saint-Mesmin</i>	1	0.04

<i>Saint-Quentin-en-Yvelines</i>	3	0.11
<i>Saint-Raphaël</i>	1	0.04
<i>Saint-Tropez</i>	1	0.04
<i>Sainte-Croix</i>	3	0.11
<i>Sainte-Maure</i>	1	0.04
<i>Sainte-Maure-de-Touraine</i>	1	0.04
<i>Santiago</i>	1	0.04
<i>Saran</i>	11	0.41
<i>Sarcelles</i>		
<i>Sceaux</i>		
<i>Semoy</i>	6	0.22
<i>Sens</i>	1	0.04
<i>Sevran</i>	9	0.34
<i>Sèvres</i>		
<i>Seyne-sur-Mer</i>	2	0.07
<i>Strasbourg</i>	3	0.11
<i>Sully-sur-Loire</i>	4	0.15
<i>Suresnes</i>	25	0.93
<i>Taiwan</i>	1	0.04
<i>Torcy</i>		
<i>Toulon</i>	5	0.19
<i>Toulouse</i>	6	0.22
<i>Tourcoing</i>		
<i>Tours</i>	73	2.72
<i>Vaucresson</i>	1	0.04
<i>Vélizy</i>	1	0.04
<i>Vendôme</i>	4	0.15
<i>Venise</i>	2	0.07
<i>Versailles</i>	4	0.15
<i>Vierzon</i>	5	0.19
<i>Villejuif</i>	2	0.07
<i>Villeneuve-la-Garenne</i>	2	0.07
<i>Villeneuve-le-Roi</i>	1	0.04
<i>Villeurbanne</i>		
<i>Vincennes</i>	9	0.34
<i>Vitry-aux-Loges</i>	8	0.30
Total	2682	

Appendix C

List of motion verbs

<i>Aller</i>	‘to go’
<i>Amener</i>	‘to bring’
<i>Arriver</i>	‘to arrive’
<i>Atterrir</i>	‘to land’
<i>Déménager</i>	‘to move’
<i>Descendre</i>	‘to go down’
<i>Entraîner</i>	‘to pull along’
<i>Entrer</i>	‘to enter’
<i>Etre muté.e</i>	‘to be transferred’
<i>Filer</i>	‘to flee’
<i>Finir</i>	‘to end up’
<i>Monter</i>	‘to go up’
<i>Partir</i>	‘to leave’
<i>Passer</i>	‘to go by’
<i>Rentrer</i>	‘to go back’
<i>Repartir</i>	‘to leave again’
<i>Retourner</i>	‘to go back’
<i>Revenir</i>	‘to come back’
<i>S'éloigner</i>	‘to move away’
<i>S'en aller</i>	‘to leave’
<i>Se balader</i>	‘to have a walk’
<i>Se déplacer</i>	‘to move’
<i>Se rapprocher</i>	‘to come closer’
<i>Se rendre</i>	‘to go’
<i>Sortir</i>	‘to go out’
<i>Venir</i>	‘to come’