

A question of change: Putting five complementary measures to the test with French polar interrogatives

Nathalie Dion

University of Ottawa, Canada
Email: ndio2@uottawa.ca

Abstract

This paper explores how five key complementary features of variable systems—overall rates, variant conditioning, productivity, contextual dispersion, and diffusion in the community—must be marshaled to provide a more comprehensive characterization of change in progress. We illustrate by revisiting a robustly variable sector of Canadian French morphosyntax whose variants are known to be in flux: the polar interrogative domain. Analyses extend the timeline of Elsig (2009)/Elsig and Poplack's (2006) diachronic analysis by an additional twenty-five years, bringing 2,000+ questions produced by 133 speakers to bear on developments occurring over a period of nearly a century and a half of spontaneous Québec French speech. Results underscore the need to consider more than rates and conditioning in the study of language change. Linguistic dispersion and diffusion in the community provide crucial insight into the mechanics of the transition period and contribute to identifying shifts in variant productivity at each point in time.

Keywords: (morpho)syntactic change; apparent time; real time; methods; transition period

Rationale

The relative frequencies of functionally equivalent forms when compared at various points in time (apparent or real) provide a basic means to identify a stable or shifting system. But rates can vary for a host of reasons unrelated to change and can only suggest a reorganization of the grammar. Conditioning of variant choice is far less susceptible to cross-corpora imbalances and is therefore able to cut through the “noise” of rates. But analysts are often required to contend with situations where substantial rate changes fail to disrupt the conditions that favor variants' use (e.g., Poplack & Dion, 2021:87–89), or, inversely, where shifts in conditioning emerge without any impact on their relative frequency (e.g., Poplack & Dion, 2021:96–102; Poplack & St-Amand, 2007:721–728). This complex and often unexpected relationship between rates and conditioning makes it difficult to characterize *what* has changed. Rates can further

obscure productivity, as when the rise of a variant results from increasing association with a narrow (but frequent) context (e.g., Poplack, Lealess, & Dion, 2013:165–167). Even stable conditioning may conceal shifts in productivity when variants are progressively restricted to environment(s) in which they are favored (e.g., Poplack & Dion, 2021:92–96). These facts underscore the need to go beyond consideration of variant rates and even conditioning when examining change over time.

This paper demonstrates the utility of a more holistic approach to assess the nature and, importantly, the locus of change by revisiting a sector of Canadian French morphosyntax especially well-suited for these purposes: the polar interrogative domain (e.g., Auger & Villeneuve, 2021; Comeau, King, & LeBlanc, 2022; Elsig, 2009; Fox, 1989). It features a robustly variable four-variant system, strong conditioning, and impressive rate shifts over time. Analyses will explore the complementary contributions of overall rates, conditioning, productivity, contextual dispersion and diffusion in the community by revisiting Elsig (2009)/Elsig and Poplack's (2006) Québec French data and extending the timeline of their diachronic analysis by twenty-five years.

Assessing and characterizing change

Measures of change

Because morphosyntactic change proceeds gradually, it must be assessed quantitatively. Understanding how competing forms are utilized by the speech community is key to tracking change. The following sections review how five key complementary features of variable systems may be characterized synchronically and compared diachronically to assess the nature and locus of change.

Overall rates of use. Taken at face value, *rates* (i.e., the relative proportion of variants realized within a variable context) constitute a general measure of a variant's prominence relative to its competitors. Comparison of rates at different points in time, or across age cohorts within a dataset, provides a means by which to track whether a variant is gaining ground or receding. Any shift in the relative frequencies of competing variants is suggestive of a potential reorganization within the system. But *monotonic* increases or decreases observed over three or more periods or age cohorts (e.g., Poplack & Malvar, 2007:144) are less likely to reflect fluctuation than slopes emerging from comparison of two datapoints and thus provide the most compelling evidence of change.

On the other hand, because rates reflect an amalgam of choices made by speakers of varying social profiles about which variant to use in each linguistic context and stylistic situation, they may vacillate across datasets for reasons independent of change (Poplack & Tagliamonte, 2001:92; Torres Cacoullós & Travis, 2021:290–291). Imbalances in the social characteristics of the individuals constituting the samples, dissimilar data collection conditions or topics discussed, and even the differing incidence of (dis)favorable linguistic contexts across datasets (*environmental changes in the textual habitat* [Szmrecsanyi, 2016:167–168]) may all artificially boost or suppress a variant's occurrence (e.g., Poplack, 1997:292–295; Robillard, 2021:231; Torres Cacoullós & Travis, 2021:291). Comparison of rates alone can therefore mislead the

analyst: apparent increases or decreases in frequency may be entirely spurious, and stability may conceal shifts in speakers' choice process.

Conditioning of variant choice. This is one of the reasons why variationists put a premium on comparing the *conditioning of the variability*: a set of rules (*constraints*) that capture which contexts favor a variant's realization relative to (an)other(s) (Poplack & Tagliamonte, 2001:91–95; Poplack & Torres Cacoullos, 2015:271). This conditioning, or *grammar of variant choice*, can be gleaned from the comparison of rates across speaker cohorts, speech styles, and linguistic environments: a variant may be likelier with monosyllabic than with polysyllabic verbs, more probable in informal than formal situations, produced more frequently by bilingual individuals than their monolingual counterparts, etc. The *direction* of these effects (i.e., *constraint hierarchies*) (Poplack & Tagliamonte, 2001:94) helps characterize the mechanisms of the choice process. When compared over time, they serve as tangible diagnostics to assess whether the grammar has changed (Poplack & Malvar, 2007:143,157–158; Travis & Torres Cacoullos, 2021:292). Emergence, neutralization, or reversal of effects all constitute change. These may be foreshadowed by strengthening or weakening over time of otherwise stable constraint hierarchies.

Productivity. Even when considered together, however, rates and conditioning may be silent as to a variant's *productivity*, that is, the extent of its breadth and versatility. If a variant is becoming increasingly favored in a narrow but frequent domain (e.g., a closed set of highly frequent verbs [Poplack & Dion, 2021:90–92; Poplack et al., 2013:165–167]), its rising rate and stable conditioning may misrepresent its waning vitality.¹ Productivity can be established by considering whether a variant's preferred contexts of use instantiate *specialized* (e.g., hyperformal speech styles), or *marked*² environments (e.g., structurally complex negative clauses; adverbially qualified conditions; lexically delimited contexts), or whether they signal greater range (as with canonical affirmative clauses). Rate changes over time and changes in the strength or direction of constraints should therefore be additionally characterized in terms of resulting gains or losses in productivity.

Linguistic dispersion. Breadth of use can also be gauged by considering a variant's *dispersion across environments* (Kastronic & Poplack, 2021:112; *contextual dispersion* in Travis & Torres Cacoullos [2021:1]), especially when compared with the distribution of contexts in the wider dataset (Poplack et al., 2013:165–170). For example, if an environment accounts for 30% of the data but 80% of a variant's tokens, this would signal a disproportionate association and restricted coverage across the variable context. Tracking dispersion diachronically will reveal whether a variant's productivity is on the rise or receding.

Diffusion across the speaker sample. Calculation of a variant's *diffusion across the speaker sample* (its attestation and frequency in individual community members' variable systems) provides yet another view on the spread and retreat of variants, both overall, and in key change-hosting contexts. For example, this measure establishes whether low overall rates stem from moderate use by all speakers or more copious

production from a smaller contingent. When examined diachronically (i.e., via comparison of age cohorts), it also clarifies whether emerging variants initially advance horizontally (i.e., through adoption by a greater number of individual speakers), via intrasystem gains in frequency (“incrementation” Labov, 2007:346), or both (e.g., Bailey, 1973:65–109). The measure likewise reveals whether a declining variant’s lower rate derives from reduced use by speakers in the younger cohort and/or from fewer of them using the variant altogether. Should changes in conditioning be detected, this measure can further elucidate how the variant comes to be adopted in a new environment or how it disappears from that context.

Quantitative analysis

To understand the relationship between rates and conditioning in the diachronic study of variable systems, we must establish each for each period and compare it to that operating at other points in time.

Statistics are a precious adjunct in assessing change, to confirm that “a non-chance explanation of the data is justified” (Paolillo, 2002:8). On the other hand, statistical significance alone is no guarantee that results are *linguistically* meaningful (Torres Cacoullos & Travis, 2021:291). In some cases, the specific constellation of features under study, data available, and research questions virtually preclude them. Here for example, while the vastness of the datasets provided more than two thousand tokens of polar questions (a prodigious number for sociolinguistic interviews), real-time comparisons require considering each period on its own. To obtain the crucial apparent-time adjunct to the analyses (e.g., Bailey, Wikle, Tillery, & Sand, 1991; Labov, 1963:291–294), intraperiod breakdowns by age cohort are necessary, subdividing the data even further. And to make matters worse, still more partitioning will be necessary to separate qualitatively different contexts, additionally reducing the set of tokens from which to extrapolate regular patterns of use. This type of situation is far from unusual in the diachronic study of morphosyntactic variability in speech, but it imposes unavoidable limits on the tools that can be used to safely calculate statistics (Paolillo, 2002:44). It is the analyst’s “responsibility to ensure that models are applied to appropriate data” (Paolillo, 2013:114).

Fortunately, conditioning may generally be apprehended by comparing frequencies across contexts.³ As will be shown in what follows, this and the other proposed measures, especially when supplemented by chi-square tests of statistical significance, offer ample means to characterize change using percentages alone.

Recognizing stability and change

Comparison of conditioning over time reveals whether a grammar is stable, changing, or changed. The following demonstrates how to recognize these trajectories in distributional results.

If a context favors a variant relative to another at various points in time, constraint hierarchies are maintained, and conditioning is considered intact. This can be discerned from the parallelism of the lines in displays such as [Figure 1](#), where the variant’s

use continues to be relatively more likely (i.e., favored) in one environment relative to another over time. As this figure crucially shows, stability may manifest even in scenarios where rates are changing.

Conditioning (and indeed the grammar) is deemed to have changed when erst-while influential contexts no longer differentiate between variants (*neutralization*; left pane of Figure 2), or in the inverse situation, when contexts begin to delineate variants (*emergence*; right pane). A *reversal of effect* (middle pane)—a more extreme instantiation of change—is a scenario in which a previously favorable context later *disfavors* the variant’s occurrence (and vice-versa).

Note that Period 2 in the neutralization scenario of Figure 2 serves as *transition* between Periods 1 and 3. Because environment A remains favorable relative to environment B, the direction of effect (and therefore conditioning) is the same, but the difference between the favorable and unfavorable environments is less pronounced. The *weakening* of the constraint is captured by the narrowing of the lines. In the emergence scenario, the effect that was incipient in Period 2 is maintained in Period 3 but has *strengthened* (increased distance between lines). These differ from the neutralization and emergence scenarios because in these cases, the conditioning is stable, but its strength has shifted.

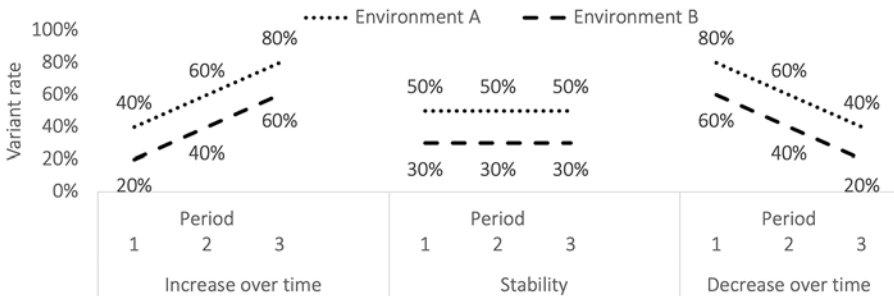


Figure 1. Stable conditioning in three rate change scenarios.

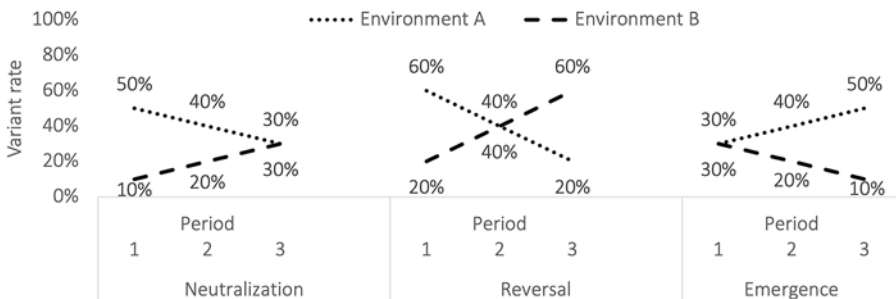


Figure 2. Change in conditioning.

A case study: The Québec French polar interrogative system

Putting these measures to the test requires both a variable displaying signs of change as well as data instantiating sufficient time depth, preferably offering a continuity of speaker dates of birth to cross-validate trajectories in real *and* apparent time. The following describes the suitability of the polar interrogative variable as well as the exceptional value of the Québec French datasets analyzed (Poplack, 1989, 2015; Poplack & St-Amand, 2007).

A system in flux

Québec French speakers have several variants at their disposal when forming polar (yes/no) questions. One of these, pronominal inversion (P-INV; [1]), features non-canonical word ordering, and the others maintain SV positioning but convey interrogation by either fronting the grammaticalized expression *est-ce que* (literally ‘is it that’) (ECQ; [2]), postposing the interrogative particle -TU⁴ (3), or suprasegmentally, via high-rising sentential intonation (INT; [4]).

- (1) *As-tu*_[P-INV] *rencontré des beaux hommes là-bas?* (21C.175.803)⁵
‘Did you meet good-looking men over there?’
- (2) *Est-ce-que*_[ECQ] *c’est l’année passée qu’on est allés en Europe?* (21C.170.300)
‘_[ECQ] it’s last year that we went to Europe?’
- (3) *Je peux-tu*_[TU] *me prendre une assiette de spaghetti?* (21C.110.1718)
‘I can_[TU] take a plate of spaghetti for myself?’
- (4) *Vous avez*_[INT] *couru?* (21C.150.90)
‘You ran?’

Elsig (2009)/Elsig and Poplack’s (2006) investigation focused on the comparison of the twentieth century (20C) variable system with its nineteenth century (19C) predecessor and their precolonization ancestor (15C-17C). They uncovered substantial differences in the relative frequency of variants over time but remarkable stability in the conditioning that motivated their choice. This is an ideal setup for assessing whether the changing rates, without disrupting the conditioning, may have come with shifts in productivity, dispersion, and diffusion. Also exciting is that the endpoint of the diachronic analysis presented a perfect storm for future change: the sustained decrease of P-INV and concomitant rise of -TU resulted in near equal distributions for these variants (27% and 33% respectively) and a third (INT; 34%) in 20C. While rates were changing monotonically, forthcoming developments were difficult to foresee: might the maximally robust variability in 20C act as a neutralizing force or destabilize these trajectories? If not, could the conditioning remain impervious to additional shifts in rate?

This four-variant scenario, which offers many possible paths for change, presents a much less predictable situation than those where a shift involving one variant necessarily affects the other in the opposite way (see Poplack & Malvar, 2007:159). This constellation of factors highlights the value of the interrogative domain, and access to data collected twenty-five years following Elsig (& Poplack)’s survey provides

a unique opportunity to consider the interplay of rates, conditioning, productivity, dispersion, and diffusion in a changing variable system.

Observing speech over the *longue durée*

This study homes in on the evolution of polar interrogatives in Québec French. The focus is on *speech*, precisely where morphosyntactic change originates. The recordings constituting the datasets referred to as 19C, 20C, and 21C⁶ described below were collected using similar methods that minimize self-monitoring and favor the use of the vernacular.

The 20C data come from the *Corpus du français parlé à Ottawa-Hull* (Poplack, 1989), a random and representative sample of francophones born, raised, and residing in the Ottawa/Gatineau region in the early 1980s. The 120 speakers constituting the full corpus were born between 1893 and 1964, but as explained in Elsig (2009:38–41), analyses focused on the forty-eight individuals sampled in the two Québec neighborhoods (Vieux-Hull and Mont Bleu; aged 17–89) to ensure maximal comparability to the 19C data that was entirely comprised of Québec-born speakers.

The 19C data are those of the *Récits du français d'autrefois* corpus (Poplack & St-Amand, 2007), a curated collection of folklorists' recordings of the spontaneous speech of rural Quebecers born between 1846 and 1895. The forty-one speakers who produced polar questions in these data were all at least fifty-five years old when recorded in the 1940s and 1950s. While the focus on the elderly during corpus constitution cannot provide a view of all age cohorts at that time, the perfect continuity in the dates of birth of these speakers (1846–1895) and those of the 20C corpus (1893–1964) offers a precious opportunity to extend the analysis of change in apparent time by an additional forty-seven years. And since the recordings predate those of the subsequent corpora, they simultaneously instantiate a real-time benchmark for comparison.

The 21C data examined in this study to supplement those previously analyzed are drawn from the *Corpus du français en contexte* (Poplack, 2015), collected twenty-five years after the 20C corpus was constituted, in the same geographic locale (the Québec side of Canada's capital region). A subsample of the wider corpus (the sociolinguistic interviews carried out with forty-four teenagers [15–18 years old; born 1988–1991]) serves as the comparison point for the 19C and 20C data, further extending the temporal span of the study and providing the chance to learn what transpired a quarter of a century after the 20C data was collected.

These carefully assembled large-scale datasets provide an unparalleled opportunity to track the development of the polar interrogatives in the vernacular. The 133 speakers retained for study (Table 1) were recorded over the course of roughly sixty-five years and offer a view on the spoken French of Quebecers born over a span of nearly a century and a half. Where data permit, intraperiod breakdowns of speakers by age will provide an apparent-time complement to the real-time comparisons of the three datasets.

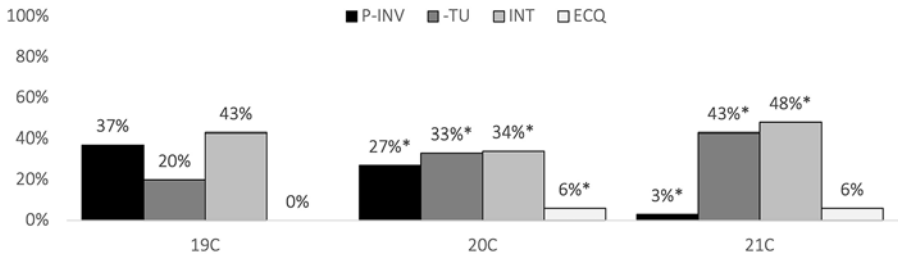
Variant rates and cross-speaker diffusion

Relative frequency of variants over time

Overall rates provide a basic indication of whether change has occurred. Figure 3 captures the considerable shifts that took place in the interrogative system in the

Table 1. Characteristics of the polar interrogative vernacular subsample⁷

	19C	20C	21C	Overall
Data collection	1940s and 1950s	1982	2005–2007	1940s–2007
<i>N</i> speakers	41	48	44	133
Year of birth	1846–1895	1893–1965	1988–1991	1846–1991
Age	55–104	17–89	15–18	15–104
<i>N</i> polar questions	789	771	524	2,084

**Figure 3.** Variant rates over time (*N*s and *p*-values in Table A1 of the appendix).⁸

twenty-five years that followed 20C corpus constitution. P-INV has nearly been ousted, not only at the expense of previously rising -TU, but unexpectedly, also INT, which was previously trending downward. The peripheral ECQ variant is holding steady in 21C at the low rate observed in the 20C data.

These substantial rate differences depict major changes: transformation of a robustly variable system involving three variants in 19C and especially 20C to what appears to be a two-variant system only a few decades later. Is this reflected in the individual speakers' grammars in each period?

Diffusion of variants in the speaker samples

Despite the low incidence of polar questions overall (an average of sixteen tokens/speaker), examination of individuals' variant repertoires⁹ independently corroborates a restructuring of the system along the lines of that suggested by the rate distributions. Variability was most robust in 20C, with 77% of the speakers ($n = 37/48$) making use of all three of the main variants at least once. Prior to this and afterwards, a healthy proportion of the sample (34% in 19C [14/41]; 59% in 21C [26/44]) only alternated between two variants (INT and P-INV for the 19C speakers, but INT and -TU for all but one of these 21C speakers).

That said, -TU's rate of 20% in 19C does not capture that more than half of the speakers' grammars (56%, $n = 23/41$) featured this variant along with P-INV and INT. Even more surprisingly, given P-INV's 3% rate in 21C, nearly a third of the individuals in this sample (30%, $n = 13/44$) make use of the three main variants. In fact, Figure 4 shows that rates of occurrence (solid lines), for each variant in each period, grossly underrepresent the diffusion of variants in the sample (dotted lines).

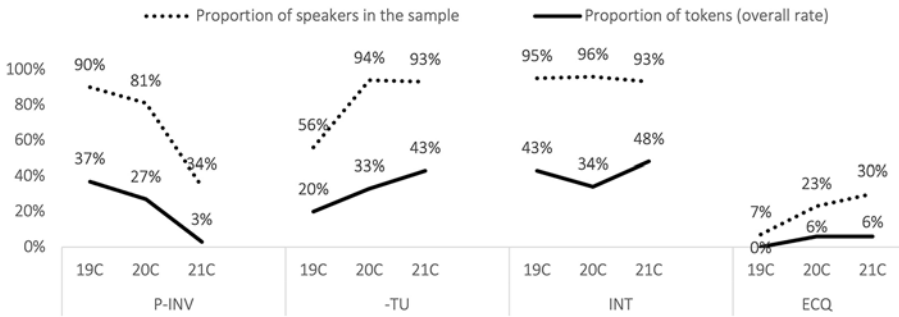


Figure 4. Variant attestation in the speaker sample relative to their frequency in the data over time (*N*s and *p*-values for attestation in Table A2 of the Appendix).

The figure also demonstrates the important insight that may be gained from examining shifts in rate as a function of diffusion. -TU's rise in frequency from 19C to 20C, for example, undervalues the extent of the inroads it made over this period, going from being used by only half of the speakers in 19C to virtually all of them thereafter. This comparison also uncovers that its rate increases from 19C to 20C and 20C to 21C came from different sources. While the former is in part due to -TU's (horizontal) spread to a greater number of speakers, the upsurge that followed this expansion must be attributed to (vertical) gains *within* the linguistic system (i.e., incrementation).¹⁰ Results likewise shed light on P-INV's trajectory. Its decline from 19C to 20C results mainly from intrarepertoire losses, since the number of sample members who use it only drops slightly. But the rate decrease that follows is partly due to a substantial proportion of individuals not using the variant at all. The inverse situation applies to ECQ. While its rate holds steady at 6% from 20C to 21C, diffusion calculations show that it may have gained some ground: whereas only 23% of the 20C speakers used the variant, it is attested in 30% of the 21C teenagers' repertoires, a result made more compelling by the fact that this cohort produced fewer polar interrogatives on average ($n = 12$ versus 16 for 20C) and thus had fewer opportunities to realize the variant.

The major changes in the frequency of variants over time and especially the differences in the variant repertoires in each corpus raise the question of whether (and how) each variant's preferred contexts of use may also have shifted in the process. Rates provide a general measure of their relative importance within the system, but the conditioning is the primary way to address what role each of them plays. Determining which factors drive the choice process will reveal whether some constraints were neutralized or lost, whether some have emerged, strengthened, or reversed course. These will measure how resilient the variable grammar is to the shifts in rates that were observed. In what follows, we review the major factors held to influence the variability: sentential polarity and grammatical person.

The role of sentential polarity

Conditioning

Interrogative variants are quite divided as a function of sentential polarity, with negative contexts strongly disfavoring (or altogether excluding) all variants other

than INT¹¹ (Auger & Villeneuve, 2021:67; Comeau, et al., 2022:642; Elsig, 2009:46; Robillard, 2021:217). Any weakening of this crucial constraint since 20C certainly could have impacted the overall rates of all variants in the system. But as can be seen in Figure 5, polarity remains strong in 21C;¹² INT continues to monopolize the negative domain (5), having even recovered the small losses it had incurred at the expense of -TU.

- (5) *Il aime*_[INT] *pas lire?* (21C.176.385)
 ‘He doesn’t like reading?’

The polarity constraint clearly outweighs all others. Regardless of stylistic context or any other element of the linguistic environment, negative structures block all interrogative processes that are syntactic in nature, whether inversion or the insertion of a fronted or postverbal particle. Polarity also manifestly overrides any effect relating to speakers’ sociodemographic profiles.

Linguistic dispersion and productivity

The breakdown of variants by polarity exposed limitations on the productivity of P-INV, -TU, and ECQ. INT’s monopoly of the negative sector would suggest greater breadth, providing it is not totally entrenched in this environment. Dispersion calculations indeed attest to its range of use, revealing that only 49% → 31% → 23% of its tokens (in 19C, 20C, and 21C respectively) are realized in negative questions. The decrease over time suggests that INT is less and less disproportionately concentrated in this specialized, narrower, and more marked context (i.e., that it is increasingly productive). But because environments themselves may not occur with the same frequency across datasets due to differences in topic, lexical items, and genres (Szmrecsanyi 2016:160–164), such values can be misleading. Indeed, in this case, the figure for 19C is amplified by a greater incidence of negative questions in this period relative to the others (22% → 12% → 12%). To control for imbalances when assessing changes in productivity over time, what must be compared is the extent to which the variant is *disproportionately* associated to the marked environment, that is, the *difference* between the proportion of tokens realized in the environment and the frequency of the context itself (e.g., 49%–22% for 19C). The resulting values (27% → 19% → 11%) do confirm

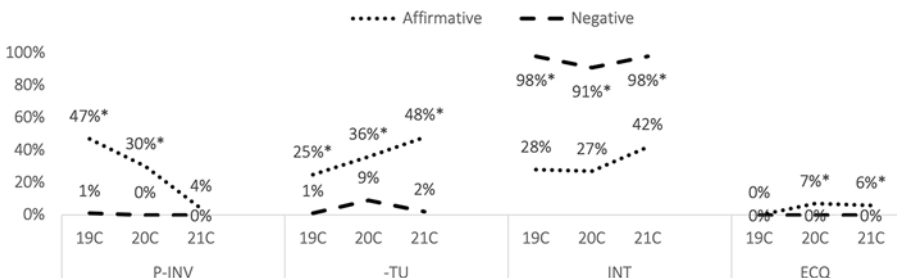


Figure 5. Variant rates as a function of polarity over time (asterisks identify the favorable environment of statistically significant differences between polarity contexts; *N*s and *p*-values in Table A3 of the Appendix).

that INT's vitality has increased. Overall, then, dispersion measures demonstrate that INT is highly productive. Not only are essentially all negative questions formed with the variant, but it also gained ground elsewhere.

The impact of dispersion on overall rates

Because negative questions are clearly the domain of INT, their relative frequency with respect to affirmative counterparts across datasets has an impact on overall rates. The greater incidence of negative tokens in 19C relative to 20C and 21C (22% → 12% → 12%) artificially boosts INT's rate in this earlier period (Figure 3), suggesting that its use decreased in 20C when in fact it is the favorable negative context that was less frequent in that dataset. In actuality, as revealed in the previous section, the variant became *more* productive over this period. On the other hand, INT's rate increase from 20C to 21C can be taken at face value because the corresponding datasets fortuitously contained the same proportion of negative questions. The rise is most accurately attributable to its advance in the *affirmative* context.

Affirmative polar questions

The near-categorical association of INT and negative questions warrants removing the nonaffirmative tokens from the further quantitative analyses of linguistic conditioning (Labov, 1969:728–729; Paolillo, 2002:63) as was done in the previous analyses of 19C and 20C (Elsig, 2009:34) and many studies thereafter (e.g., Auger & Villeneuve, 2021:67; Robillard, 2021:212; Villeneuve, 2020:119). Including them would necessarily obscure any effect that other potentially influential factors might have on tokens for which variant choice remains *variable*. Moreover, since -TU's rise and P-INV's dramatic decline are necessarily occurring outside the negative context (from which they are absent), focusing on affirmative polar questions also ensures that we are homing in on the *actual* locus of the changes. Due to their paucity, tokens of ECQ (primarily used to mark hyperformal style [Elsig, 2009:103]) are also set aside.

Contextualizing the three main variants participating in the variability in the affirmative sector with respect to one another (Figure 6)¹³ reveals that whereas -TU's rise has been continuous, INT's sixteen-point increase since 20C is sudden and recent. P-INV's drop in 21C, while foreshadowed by a progressive fall in the earlier periods, was also notably abrupt. Were these substantial rate changes (all $p < .0001$) accompanied with shifts in conditioning?

The role of grammatical person

Conditioning

In previous analyses of the 19C and 20C data and in most other studies of Laurentian and Acadian French (e.g., Auger & Villeneuve, 2021:68; Comeau, 2016:196; Fox, 1989:324; Robillard, 2021:228), polar questions addressed to second-person subjects are generally found to contrast with other persons: the former disfavor or exclude -TU/-TI, but P-INV's use is all but confined to this context, rarely occurring with

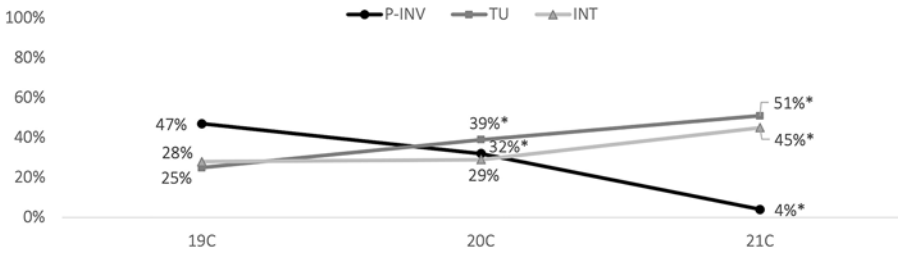


Figure 6. Distribution of the three main variants in affirmative polar questions over time.

first- and third-person subjects, if at all. Since this factor plays a central role in variant choice, it is important to assess whether its influence has changed since 20C.

Direct comparison with earlier results requires contrasting second and other persons, but conflation of *tu* and *vous*¹⁴ would obscure substantial differences between them and conceal the shakeup that occurred *within* the second-person context (Figure 7).

Grammatical person has influenced the choice of P-INV and -TU in very different ways over time. In 19C, both types of second-person subjects strongly favored P-INV and strongly disfavored -TU,¹⁶ which was instead favored by other subjects. But the homogeneity of the second-person contexts dissipated thereafter; it became more and more common to find particle -TU in questions whose subject was pronominal *tu* (i.e., *tu* VERB-TU; [6]). P-INV concomitantly lost the narrow association it had previously enjoyed with this subject. And while it initially maintained its association with *vous*, this too has seemingly faded since.¹⁷

- (6) **Tu peux-tu**_[-TU] *m'appeler un taxi?* (20C.088.1239)
 'You can_[-TU] call me a taxi?'

Both changes progressed essentially to the point of completion in 21C (i.e., full neutralization of the distinction between *tu* and other subjects for -TU [n.s. $p = .4979$], and a small [but statistically significant] rate difference for P-INV [7% versus 0%; $p < .0001$]). The grammatical person effect for both variants now consists of a statistically significant contrast between *vous* and all other subjects (including *tu*) rather than

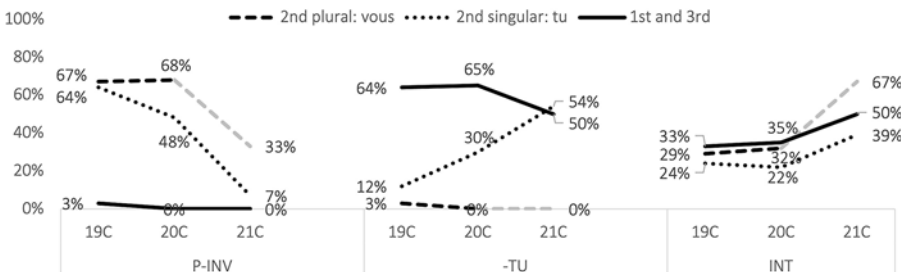


Figure 7. Variant rates by grammatical person over time (Ns and p -values in Table A4 of the Appendix).¹⁵

the 19C situation where both types of second-person subjects together aligned against the others. INT's use has not been strongly affected by the subject of the question. Its initial stability and subsequent increase occurred indiscriminately.

Variable context redefined

Although the variants alternate with one another in the broader sense (i.e., in affirmative polar questions), Figure 8 shows that they only co-vary when the subject of the question is second-person singular *tu*. *Vous* questions and those with first- and third-person subjects each exclude one of the variants, leaving only two (different pairs) to alternate (see also Auger & Villeneuve [2021:59] and Comeau [2016:196]).

The fact that the different grammatical persons host distinct sets of variants indicates that they are operating under different rules. This is supported by the fact that change is not proceeding to the same extent with each of them: only *tu* questions underwent change from 19C to 20C, and shifts overall have been more substantial with these subjects. -TU's opposing trajectories across subject types—a rising rate with *tu*, stability with *vous*, and a drop with other persons—further reinforces the independence of these subdomains of the interrogative sector. Accordingly, ensuing analyses control for the strong influence of person by partitioning the dataset (Paolillo, 2002:143; see also Comeau [2016:196] and Auger & Villeneuve [2019:226–227]). The three grammatical person contexts are considered in turn. In each case, variant rates and diffusion in the sample are considered by period and age cohort, and the resulting implications for their relative productivity are discussed.

Change by subject type

Second person vous questions. Although overall rates suggest that P-INV has all but dropped out of usage (Figure 6), breakdowns by grammatical person (Figure 8) clarify that it is hanging on when the subject of the question is *vous*. In this environment, where -TU is essentially barred, INT is its only rival (7).

As was shown in Figure 8, P-INV's strong hold on this context in 19C (67%), did not waver in 20C (68%; differences n.s.). A sharp decline (down to 33%) followed, but the 21C percentage is calculated on the basis of only nine tokens and does not register

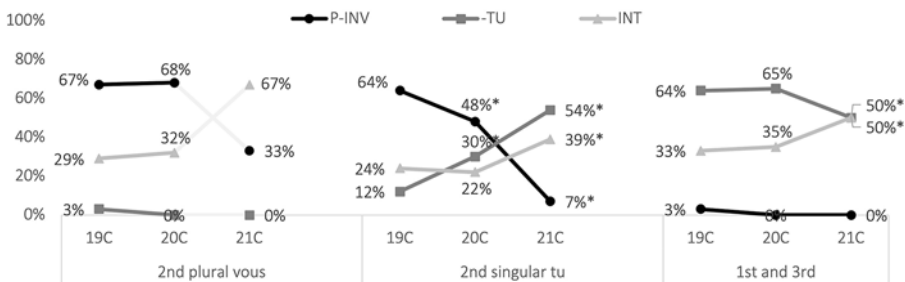


Figure 8. Variant rates by grammatical person of the subject over time (See the Appendix for *Ns* [Table A4] and *p*-values [Table A5]).

- (7) a. *Elle dit “voulez-vous_[P-INV] une soupe au poulet?”*. (21C.110.1160)
 ‘She says “would you like a bowl of chicken soup?”’
- b. *Vous voulez_[INT] faire de quoi?* (21C.159.604)
 ‘You want to do something?’

as significant. As it turns out, *vous* questions are less and less frequent in the data from one period to the next (25% → 17% → 2%). Because this environment favors P-INV, the decrease artificially amplifies the extent of its regression. The converse increase of nonsecond-person subjects over time (29% → 40% → 47%)—a context where P-INV does *not* occur—further conspires to exaggerate its fall overall.

The diminishing place of *vous* questions is discernible in real time and in apparent time, not only in the number of raw tokens produced by cohort (eighty-nine to nine), but also the proportion of speakers who realized them (85%-18%) and the average number of tokens for each of them (five to one). While some of this decline likely results from the waning of *vouvoisement* (the use of plural *vous* as a formal second singular pronoun of address) as a societal practice, *vous* remains the sole plural second-person pronoun, and as such, seems unlikely to disappear altogether. Instead, the dwindling incidence of *vous* in these corpora likely stems from differing opportunities to use it in the three datasets. Regardless of reason, the unfortunate consequence is that conclusions about what is going on in this sector of the interrogative domain become increasingly tenuous the younger the cohort.

With this caveat in mind, apparent-time distributions (Figure 9) suggest that P-INV’s use was relatively stable for the cohorts who produced the most *vous* questions: P-INV shows no signs of regressing until its rate drops for the 18-24-year-olds of the 20C corpus ($p = .0008$).

P-INV’s vitality in 20C is independently discernable in Figure 10. The variant is used by a large proportion of the speakers who use *vous* in each cohort—including the aforementioned 18-24-year-olds who display lower rates. It is only in the most recent dataset that we observe a drop in the proportion of the sample that uses the variant.

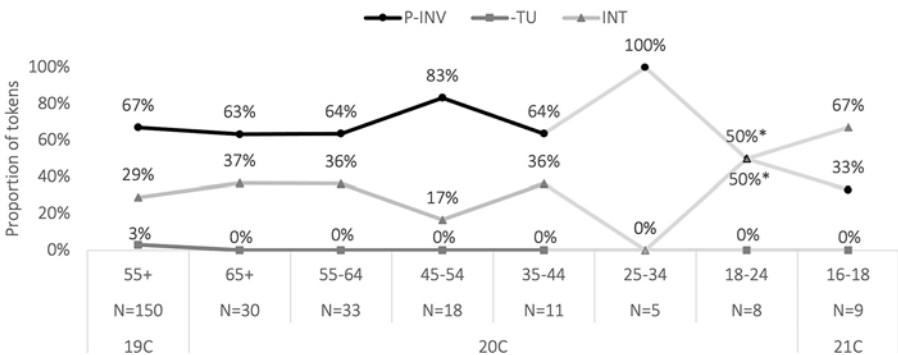


Figure 9. Variant rates with subject *vous* in real and apparent time.

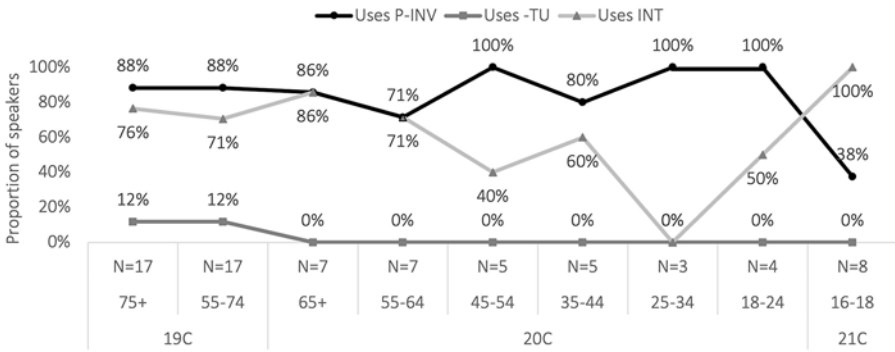


Figure 10. Attestation of variants for speakers of each cohort who produce *vous* questions.

All told then, the rarity of *vous* questions may be undermining P-INV’s vitality to a certain extent. Apparent-time analyses of diffusion provide evidence that it has not entirely retreated from this context. In the grand scheme of things, however, restriction to such a narrow sector of the interrogative domain speaks to very limited productivity overall.

Second person tu questions. P-INV’s decline in its only other context of use (with subject *tu*; [8]), was quite dramatic (64% → 48% → 7%; Figure 8). Because *tu* questions are so frequent in the datasets (46% → 43% → 51%), this translates to a huge drop in its overall rate of use (Figure 6). In 19C, P-INV mostly alternated with INT, at rates quite similar to those they displayed in the only other environment in which they co-vary (i.e., with *vous* questions), and -TU was rarely selected (12%). But recall that this is the context that hosted the most substantial change: a significant decrease of P-INV and concurrent increase of -TU (both $p < .0001$).¹⁸

- (8) a. *Mais **vas-tu**_[P-INV] à l’église tous les dimanches?* (21C.163.796)
‘But do you go to church every Sunday?’
- b. ***Tu manges-tu**_[-TU] des protéines?* (21C.118.663)
‘You eat_[-TU] protein?’
- c. ***Tu faisais**_[INT] des mauvais coups?* (21C.104.100)
‘You were doing bad stuff?’

But the apparent-time breakdown in Figure 11 reveals a more nuanced story. P-INV’s decline was not as progressive as we may have expected; the 19C speakers and most of the 20C cohorts showed no signs of moving away from the variant. It was only when the 25-34-year-olds started to use more -TU ($p = .0066$) that P-INV decreased ($p = .0013$). Both changes progressed monotonically such that for the youngest 20C cohort -TU was the clear majority variant ($p = .0181$) and P-INV only used 20% of the time (n.s. $p = .0539$). The switchover between P-INV and -TU did not affect INT, which held steady at a fairly low frequency (n.s., $p = .6095$). 21C saw P-INV’s rate

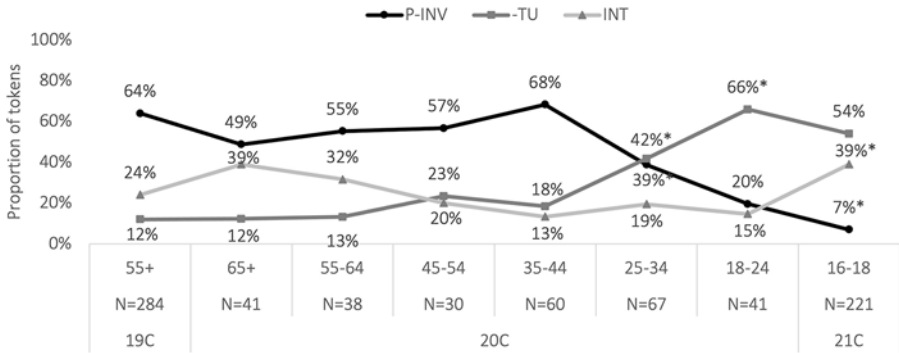


Figure 11. Variant rates with subject *tu* in real and apparent time.

dwindle further ($p = .0147$), and while *-TU* remains the majority variant, it backtracked somewhat (n.s. $p = .1736$), because *INT*'s use increased, as it did with *vous* subjects.

The diffusion breakdowns in Figure 12 provide context for these seemingly abrupt changes. They demonstrate that in the earlier periods, when the rates were relatively stable, change was actually brewing: from one age group to the next, more and more speakers were accepting the postverbal postposition of particle *-TU* with *tu* questions. It was only once a large contingent of speakers accepted the use of these [*tu* + VERB-TU] constructions that *P-INV* dropped appreciably in rate (Figure 11) and diffusion (Figure 12).

This calculation captures the huge headway that *-TU* made in this context: from being used by only 21% of the speakers to (nearly) all of them over the course of only a few generations. The homophony¹⁹ of particle *-TU* and inverted pronoun *tu* likely factors into this situation (Auger & Villeneuve, 2021:62; Elsig, 2009:182; Picard, 1992:69–70). It is quite possible, as suggested by Vinet (2000:386), that some speakers mistake *-TU* for inverted *tu*.

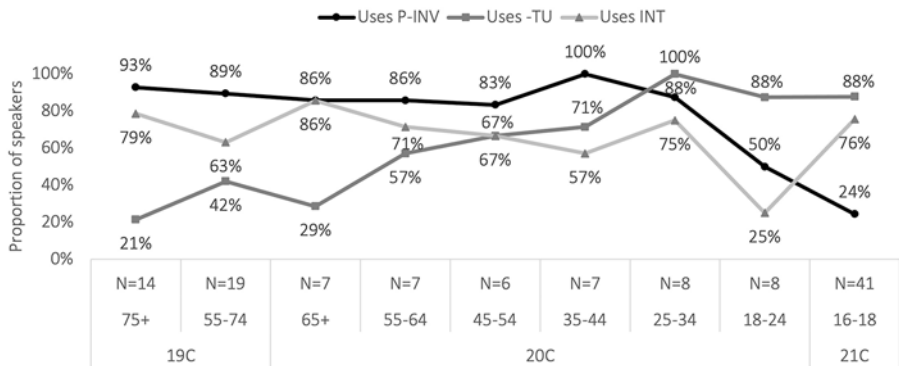


Figure 12. Attestation of variants for speakers of each cohort who produce *tu* questions.

Questions addressing first- and third-person subjects. The final sector of the polar interrogative domain involves alternation between -TU and INT (9). While P-INV's exclusion from this context could theoretically have shielded the environment from change, Figure 8 showed that after initial stability from 19C to 20C, -TU dropped ($p = .0022$) and INT rose ($p = .0016$), resulting in an equal division of the labor in 21C.

- (9) a. **Ça valait-tu**_[-TU] *la peine?* (21C.101.591)
 'It was_[-TU] worth it?'
 b. **Ça te dérange**_[INT] *si je joue avec ta casquette?* (21C.108.60)
 'It bothers you if I play with your ball cap?'
 c. **Ils en font-tu**_[-TU] *un autre bientôt?* (21C.119.445)
 'They're making_[-TU] another one soon?'
 d. **Ils en vendent**_[INT] *là-bas?* (21C.151.521)
 'They sell some over there?'

But breakdowns of rates in apparent time fail to support a “stability followed by change” scenario. Instead, they turn up only a slight (n.s.) ebb and flow over time. Diffusion calculations likewise feature continued stability. Because the “other persons” label subsumes disparate subject types (first person *je*, third person *on* [in both third indefinite and first plural meaning], as well as a variety of other third-person subjects [NPs, personal pronouns *il(s)/elle(s)*, impersonal *il*, indefinite demonstratives *ce/ça*]), it is worth considering whether these trajectories reflect the behavior of all members of the group.

It is immediately apparent from the breakdown in Figure 13 that while these subjects behaved homogeneously in 19C, -TU globally preferred (65%–72%) over INT (28%–35%), they begin to show inclinations toward one variant or the other thereafter.²⁰ This emergence of a subject effect in 20C was not apparent from the amalgamated percentages in Figure 8, because the increases with some subjects were offset by decreases with others. On the other hand, the overall rise of INT and fall of -TU from 20C to 21C in

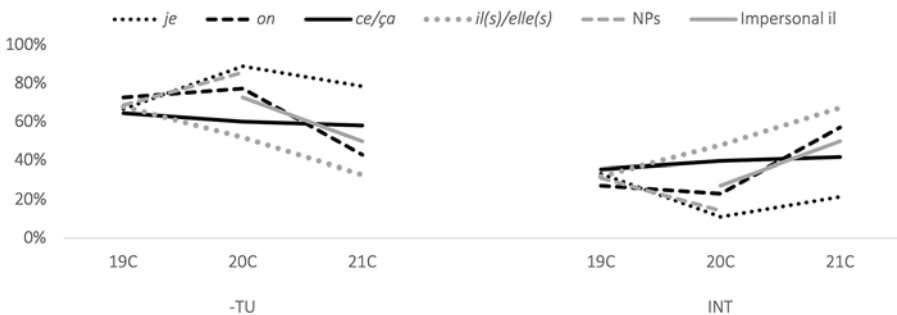


Figure 13. Rate of -TU (left) and INT (right) by subject over time (*N*s and *p*-values in Table A6 of the Appendix).

Figure 8 does accurately reflect a pansubject trend. Note that -TU's decline would be even more palpable were it not for the fact that its most moderate drop occurred with the frequent *ce/ça* subjects that account for more than half of the first- and third-person data in each corpus.

These results show that -TU's decline has been more than numerical. The variant has also become less productive, going from being the preferred option with all nonsecond-person subjects to disproportionately associated to only *je* and *ça/ce*, its grasp on even these contexts dwindling. INT appears poised to gain further ground with first- and third-person subjects.

A holistic view of change

This paper set out to explore how competing measures must be marshaled to accurately depict the process of language change. The complementary measures of rate, conditioning, productivity, dispersion, and diffusion have helped uncover details of the variable system's evolution that would otherwise have escaped notice.

P-INV's high rate in 19C—entirely due to the frequency of the environments in which it was favored—obscured the fact that it was already highly contextually restricted at that time, all but excluded from negative contexts and affirmative questions involving first- and third-person subjects. Its dramatic numerical drop in 20C and especially 21C turned out to be exaggerated by disproportions in the incidence of various subject types in the three datasets. Analyses of dispersion uncovered that both the waning of *vous* questions and the increase in first- and third- person subjects from one period to the next conspired to amplify its decline. But beyond these misleading rate changes, P-INV also lost considerable ground due to a change in *conditioning*—its association with highly frequent *tu* questions weakened over time. Apparent-time distributions of rate and diffusion showed that P-INV's decline only occurred once a large enough contingent of speakers admitted -TU with these subjects. P-INV's low overall rate in 21C (4%) obscures the fact that a third of the speakers in this period use the variant and that it remains a contender with *vous* questions. Nonetheless, the narrowness of this context and decreasing attestation in individual speaker repertoires reflect limited productivity and bode ill for its long-term survival as a serviceable variant of polar interrogation.

With respect to -TU, it would seem from increasing overall rates and progressive diffusion that this is an expanding variant. But the dispersion of subject types contributes to this impression. The relatively lower incidence of first- and third-person contexts in 19C (29% versus 40% and 47%), when these subjects favored the variant, downplays its frequency in that period. Likewise, the progressive dwindling of *vous* questions (which exclude -TU) exaggerate its increase in 20C and 21C. While the variant nevertheless did experience considerable growth, both numerically and in terms of productivity, the expansion that resulted from the neutralization of the constraint against its selection with the highly frequent *tu* questions occurred as it lost its hold on questions formed with first- and especially third-person subjects. As a result, while the variant is present in speaker repertoires and varies robustly with INT in 21C, its use has become increasingly context-dependent: it is barred from occurring in negative questions and with *vous*, is on its way out with most third-person subjects, is losing

ground in its strongholds of *je* and *ce/ça*, and is losing momentum with *tu*. While its gains with *tu* and maintenance with *ça/ce* have translated to healthy (and rising) overall rates of use, its productivity has diminished and its increase slowed as a result of the more recent expansion of INT.

INT's trajectory may be the most surprising. Its apparent drop in overall frequency from 19C to 20C was the product of the fewer negative questions in the latter. Dispersion analyses revealed that its productivity in fact increased over this period. Its stability with first- and third-person subjects from 19C to 20C likewise concealed its incipient encroachment upon -TU's territory in this context. INT's increase since 20C appears to be occurring wholesale, with *tu*, *vous*, and most other subjects. It now constitutes a strong rival to -TU, enjoying greater diffusion, not only across speakers, but also within the linguistic system. Its rate is robust not only in contexts where -TU is admitted, but beyond, since it is not blocked from being used with *vous*, nor is it excluded with negative questions. Its use (and indeed monopoly) in the latter context applies to all subjects, all verbs, all stylistic contexts, and speakers of all sociodemographic profiles. Rates of use conceal that INT is without doubt the most productive polar interrogative variant.

Discussion

The previously uncharted results of this study emerged entirely from variant distributions, showing that so long as rates are properly contextualized, they can provide considerable insight on the process of the language change. Results validate simultaneously the utility of the apparent-time construct and of comparing spontaneous speech data collected at different points in real time in different sociocultural contexts.

Overall rates enable the identification of substantial shifts but emerge as the most misleading measure of change. One reason is that they are extremely sensitive to contextual dispersion. When highly influential environments are disproportionately frequent in the data, cross-cohort or cross-dataset imbalances artificially impact frequencies of use. In this study, such disproportions caused impressions of stability where change had occurred and suggested change where none had transpired. Rates also overstated productivity when variants were preferred in narrow but highly recurrent contexts, and, conversely, concealed high productivity when preferred environments were infrequent but linguistically unmarked and diverse. Finally, rates were found to grossly underrepresent diffusion of variants in speaker repertoires.

Conditioning is a much less volatile measure of change. Since it is apprehended by comparing the rate *within* an environment to that in another, the relative frequency of said environments—whether corpus-internally or across datasets—does not normally complicate this assessment. And since constraint hierarchies emerge from the *relative* ordering of rates, conditioning can be readily gauged even as variant frequencies rise or fall over time. In the case at hand, much conditioning proved to be stable, testimony to the resilience of the variable grammar to shifts in rate. The near-qualitative exclusion of variants from certain contexts detected in the speech of fishermen and loggers recorded in the 1940s was evident not only among speakers of all ages in the 20C corpus but also among the contemporary urban millennial teenagers. That said, this measure also

identified emerging, strengthening, weakening, and neutralized effects—key stages in the transition from one grammar to another.

But because it is silent as to the markedness and frequency of the favorable environments, conditioning cannot speak to whether a variant's preferences instantiate productivity or contextual restriction. Likewise, change in conditioning over time can involve expansion or constriction and thus has important implications for assessing the role of a variant in the system. The measure of productivity proved especially useful in this study, because it helped clarify that neutralization of one influential factor broadened the vitality of a variant, while the emergence of another constraint reduced its hold on a different environment. Identifying and describing changes in conditioning is thus insufficient. These must be further contextualized in terms of growth or decay.

Dispersion contributes to assessing the legitimacy of apparent rate changes over time and to establishing each variant's productivity both synchronically and as the system develops. In this case, such analyses identified which changes derived from imbalances in the proportion of contexts across datasets and played a key role in evaluating a variant's breadth at a given point in time and from one period to another.

Diffusion in the community (via attestation in speaker samples) provided important insight on how variants advance and recede overall. Some rate changes come from adoption or abandonment of variants across generations, others by shifts in extent of usage. Diffusion analyses also afforded a granular view of how a variant expanded into a new context. It was not accepted by all speakers at the same time but increased cohort to cohort, as a greater proportion of speakers adopted its use in the novel environment. We also learned that overall rates underestimate diffusion, since variants may enjoy reasonable cross-speaker attestation even when exceedingly rare.

This exercise has demonstrated that we cannot fully apprehend language change without going beyond the usual measures of rates of use and variant conditioning. Linguistic dispersion and diffusion in the community provide key insights into the mechanics of the transition period and contribute to identifying shifts in variant productivity at each point in time.

Acknowledgments. This research was supported by a doctoral fellowship from the Social Sciences and Humanities Research Council of Canada and the uOttawa Office of the Vice-President, International and *Francophonie*. I am deeply grateful to Martin Elsig for carrying out his monumental study and generously sharing his 19C and 20C token files with me. All data were extracted from uniquely rich datasets housed at the uOttawa Sociolinguistics Lab. I thank Shana Poplack for granting me access to the corpora and lab materials, and extend my appreciation to her and Stephen Levey as well as *Language Variation and Change* reviewers for providing comments that substantially enhanced the presentation of these results.

Competing interests. The author declares none

Notes

1. The idea adopted here that “conventionalization” (i.e., “an increasing association between structural context and variant and concomitant reduction or loss of ‘free’ [...] choice amongst them”) embodies lessened productivity comes from Poplack, Torres Cacoullos, Dion, de Andrade Berlinck, Digesto, Lacasse, and Steuck (2018:222). It is based on processes invoked in the grammaticalization literature (“ritualization [Haiman, 1994], obligatorification [Lehmann, 2015] and loss of ‘freedom’ [Haspelmath, 1998:318]” [Poplack et al. 2018:222]).

2. See Haspelmath (2006:26–40) for a review of twelve senses of markedness in linguistics, roughly grouped under the rubrics of complexity, difficulty, and abnormality.
3. Although see Sankoff (1988:986) for a discussion about how the nonindependence of factor groups may yield misleading rates.
4. This variant is generally referred to as -TI when realized with an unrounded [i] in Europe and Acadian varieties of French (e.g., Comeau et al., 2022), but as -TU in Québec French, where the vowel came to be realized (for “obscure” reasons [Léard, 1996:113]) as rounded [y] (Morin, 1985:794; Picard, 1992:69). Elsig notes the occurrence of the unrounded pronunciation in the oldest data, but concludes that “the patterns of conditioning do not necessitate a different grammatical or syntactic treatment of the one as opposed to the other” (2009:46). Following his convention (2009:2), these two historically related forms (Foulet, 1921; Léard, 1996:112; A. Morin, 2017:81–85; Picard, 1992:65–66) have been subsumed under the same label (-TU).
5. Examples are reproduced verbatim from the audio recordings housed at the uOttawa Sociolinguistics Laboratory constituting the corpora described in the next section. Codes in parentheses refer to corpus, speaker, and line number.
6. This characterization roughly reflects the century during which speakers finished acquiring their vernacular: late nineteenth (to very early twentieth) for those in the earlier corpus, twentieth for the second, and late 1990s/early 2000s for the third.
7. Disparities between token counts presented here and those found in Elsig (2009) and/or Elsig and Poplack (2006) are the result of efforts to standardize the variable context (Elsig, 2009:42–46) and coding schema for the diachronic analysis. The discrepancy with Elsig (2009:38) in the number of speakers included in the 19C study (forty-one versus forty) is due to the inclusion here of a speaker who only produced negative polar questions.
8. Unless indicated otherwise, asterisks in figures identify statistically significant shifts with respect to the previous period.
9. Given the low incidence of tokens by speaker, diffusion assessments will reflect the proportion of speakers who make use of the variant at least once (i.e., attestation in speakers’ repertoire).
10. Change across the lifespan is not implied here. Rather, individual speakers in the 21C corpus used the variant more often than their 20C counterparts.
11. Explanations for this effect revolve around the problematic nature (syntactic or pragmatic) of the negative environment for the other variants (Martineau & Vinet, 2005:202; A. Morin, 2017:154–163; Vinet, 2000).
12. The neutralization of the polarity constraint for P-INV is a result of its reduced use in affirmative contexts, which becomes too infrequent in 21C to differentiate itself in a statistically significant way from the consistently inhospitable negative environment.
13. Raw token counts are found in the affirmative rows of Table A3 in the Appendix. Percentages reflect the removal of the ECQ tokens from the denominators.
14. This category combines the semantic plural as well as the formal second-person singular form of address. Elsig (2009:184) reports no distributional differences between them.
15. In this and subsequent figures, paler sections of lines signal values based on especially low token counts.
16. Elsig invokes both a feature (i.e., number) mismatch (2009:184–186) and the stylistic incompatibility of -TU with formal *vous* (2009:78; see also Villeneuve, 2020:120) to explain this result. Auger and Villeneuve (2021:59) propose that it is -TU’s phonetic form (homophony with subject *tu*) that makes it incompatible with *vous* questions, noting that early attestations of nonhomophonous -TI occurred with subject *vous*. This result is marred by the low occurrence of *vous* in the 21C data ($n = 9$ tokens).
18. Auger and Villeneuve (2021:168) also note the unexpected increase of -TU with second-singular subjects (and exclusion from *vous*) in Montreal French (in televised series and a novel, both dating from 2016).
19. This is a byproduct of the replacement of -TI, attested in the earlier materials (Elsig, 2009:46), by -TU (Comeau et al., 2022: 626–27; Léard, 1996:113; Morin, 1985:794).
20. None of the shifts for individual subjects are statistically significant, likely due to low token counts (Table A6). Multivariate analyses (not displayed here) show that the factor group of grammatical person is selected as significant in 20C (and 21C) but not 19C.

References

- Auger, Julie, & Villeneuve, Anne-José. (2019). Building on an old feature in langue d’Oïl: Interrogatives in Vimeu Picard. *Journal of French Language Studies* 29(2): 209–233.

- Auger, Julie, & Villeneuve, Anne-José. (2021). Étude comparative des particules interrogatives en picard et dans deux variétés de français parlées au Canada. *Langue française* 212(4): 57–74.
- Bailey, Charles-James. (1973). *Variation and linguistic theory*. Arlington, VA: Center for Applied Linguistics.
- Bailey, Guy, Wikle, Thomas, Tillery, Jan, & Sand, Lori. (1991). The apparent time construct. *Language Variation and Change* 3(3): 241–264.
- Comeau, Philip. (2016). An extension of the comparative sociolinguistics approach for sociosyntax: Comparing a single linguistic constraint across multiple sociolinguistic variables. *Linguistic Variation* 16:183–220.
- Comeau, Philip, King, Ruth, & LeBlanc, Carmen L. (2022). Continuity and change in the evolution of French yes-no questions: A cross-variety perspective. *Diachronica* 39(5): 616–657.
- Elsig, Martin. (2009). *Grammatical variation across space and time: The French interrogative system*. Amsterdam/Philadelphia: John Benjamins Publishing.
- Elsig, Martin, & Poplack, Shana. (2006). Transplanted dialects and language change: Question formation in Québec. *University of Pennsylvania Working Papers in Linguistics* 12(2): 8.
- Foulet, Lucien. (1921). Comment ont évolué les formes de l'interrogation. *Romania* 47(186/187): 243–348.
- Fox, Cynthia. A. (1989). *Syntactic variation and interrogative structures in Québécois*. Doctoral dissertation, Indiana University.
- Haiman, John. (1994). Ritualization and the development of language. In W. Pagliuca (ed.), *Perspectives on grammaticalization*. Amsterdam: Benjamins. 3–28.
- Haspelmath, Martin. (1998). Does grammaticalization need reanalysis? *Studies in Language* 22:315–351.
- Haspelmath, Martin. (2006). Against markedness (and what to replace it with). *Journal of Linguistics* 42(1): 25–70.
- Kastronic, Laura, & Poplack, Shana. (2021). *Be that as it may*: The unremarkable trajectory of the English subjunctive in North American speech. *Language Variation and Change* 33(1): 107–134.
- Labov, William. (1963). The social motivation of a sound change. *Word* 19:273–309.
- Labov, William. (1969). Contraction, deletion, and inherent variability of the English copula. *Language* 45(4): 715–762.
- Labov, William. (2007). Transmission and diffusion. *Language* 83:344–387.
- Léard, Jean-Marcel. (1996). *Ti/-tu, est-ce que, qu'est-ce que, ce que, hé que, don*: des particules de modalisation en français? *Revue québécoise de linguistique* 24(2): 107–124.
- Lehmann, Christian. (2015). *Thoughts on grammaticalization*, 3rd edition. Berlin: Language Science Press.
- Martineau, France, & Vinet, Marie-Thérèse. (2005). Microvariation in French negation markers: A historical perspective. In M. Batllori, M.-L. Hernanz, C. Picallo, and F. Roca (eds.), *Grammaticalization and parametric variation*. Oxford: Oxford University Press. 194–205.
- Morin, Yves-Charles. (1985). On the two French subjectless verbs *voici* and *voilà*. *Language* 61(4): 777–820.
- Morin, Annick. (2017). *Questioning particles: A cross-linguistic approach to Quebec French polar interrogatives*. Doctoral dissertation, University of Toronto.
- Paolillo, John C. (2002). *Analyzing linguistic variation: Statistical models and methods*. Stanford: Center for the Study of Language and Information.
- Paolillo, John C. (2013). Individual effects in variation analysis: Model, software, and research design. *Language Variation and Change* 25(1): 89–118.
- Picard, Marc. (1992). Aspects synchroniques et diachroniques du tu interrogatif en québécois. *Revue québécoise de linguistique* 21(2): 65–74.
- Poplack, Shana. (1989). The care and handling of a mega-corpus: The Ottawa-Hull French project. In R. Fasold & D. Schiffrin (eds.), *Language change and variation*. Amsterdam: Benjamins. 411–451.
- Poplack, Shana. (1997). The sociolinguistic dynamics of apparent convergence. In G. Guy, C. Feagin, J. Baugh & D. Schiffrin (eds.), *Toward a social science of language: Papers in honor of William Labov*. Amsterdam: Benjamins. 285–310.
- Poplack, Shana. (2015). Norme prescriptive, norme communautaire et variation diaphasique. In K. Kragh & J. Lindschouw (eds.), *Variations diastématiques et leurs interdépendances*. Strasbourg: Éditions de linguistique et de philologie. 293–319.
- Poplack, Shana, & Dion, Nathalie. (2021). Cartographie de la variation et du changement morphosyntaxique en français: leçons à retenir. *Cahiers internationaux de sociolinguistique* 18(1): 83–115.
- Poplack, Shana, & Malvar, Elisabete. (2007). Elucidating the transition period in linguistic change: The expression of the future in Brazilian Portuguese. *Probus* 19(1): 121–169.

- Poplack, Shana, & St-Amand, Anne. (2007). A real-time window on 19th-century vernacular French: The *Récits du français québécois d'autrefois*. *Language in Society* 36(5): 707–734.
- Poplack, Shana, & Tagliamonte, Sali A. (2001). *African American English in the diaspora*. Malden: Wiley-Blackwell.
- Poplack, Shana, & Torres Cacoulios, Rena. (2015). Linguistic emergence on the ground: A variationist paradigm. In B. MacWhinney, and W. O'Grady, W. (eds.), *The handbook of language emergence*. Malden: Wiley Blackwell. 267–291.
- Poplack, Shana, Lealess, Allison V., & Dion, Nathalie. (2013). The evolving grammar of the French subjunctive. *Probus* 25:139–95.
- Poplack, Shana, Torres Cacoulios, Rena, Dion, Nathalie, de Andrade Berlinck, Rosane, Digesto, Salvio, Lacasse, Dora, & Steuck, Jonathan. (2018). Trajectories of change in Romance sociolinguistics. In W. Ayres-Bennett, and J. Carruthers (eds.), *Manuals in linguistics: Romance sociolinguistics*. Berlin: De Gruyter. 217–252.
- Robillard, Suzanne. (2021). *Implicit norms and "School French" forms: Linguistic cohesion of second-generation francophones in Victoria, BC*. Doctoral dissertation, University of Ottawa.
- Sankoff, David. (1988). Variable rules. In U. Ammon, N. Dittmar & K. J. Mattheier (eds.), *Sociolinguistics: An international handbook of the science of language and society*. Berlin: Walter de Gruyter. 984–997.
- Szmrecsanyi, Benedikt. (2016). About text frequencies in historical linguistics: Disentangling environmental and grammatical change. *Corpus Linguistics and Linguistic Theory* 12(1): 153–171.
- Torres Cacoulios, Rena, & Travis, Catherine E. (2021). Alternating or mixing languages. In D. Pérez, M. Hundt, J. Kabatek, and D. Schreier (eds.), *English and Spanish: World languages in interaction*. Cambridge: Cambridge University Press. 287–311.
- Travis, Catherine E., & Torres Cacoulios, Rena. (2021). Categories and frequency: Cognition verbs in Spanish subject expression. *Languages* 6(3): 126.
- Villeneuve, Anne-José. (2020). Variation stylistique et accommodation langagière: l'interrogation totale en français québécois soutenu. In D. Bigot (ed.), *Les français d'ici en perspective*. Québec: Presses de l'Université Laval. 109–129.
- Vinet, Marie-Thérèse. (2000). Feature representation and *-tu (pas)* in Quebec French. *Studia Linguistica* 54(3): 381–411.

Appendix

Table A1. Variant rates over time

	19C		19C → 20C	20C		20C → 21C	21C		Overall
	%	<i>N</i>	<i>p</i> -value	%	<i>N</i>	<i>p</i> -value	%	<i>N</i>	<i>N</i>
P-INV	37%	291	* <i>p</i> < .001	27%	205	* <i>p</i> < .001	3%	18	514
-TU	20%	158	* <i>p</i> < .001	33%	255	* <i>p</i> < .001	43%	223	635
INT	43%	337	* <i>p</i> < .001	34%	263	* <i>p</i> < .001	48%	253	852
ECQ	0%	3	* <i>p</i> < .001	6%	50	<i>p</i> = .639	6%	30	83
Total	100%	789		100%	773		100%	524	2084

Table A2. Variant attestation in the speaker sample

	19C		19C → 20C		20C		20C → 21C		21C		19C → 21C	
	% spkrs	N	p-value	% spkrs	N	p-value	% spkrs	N	p-value	% spkrs	N	p-value
P-INV	90%	37	$p = .377$	81%	39	$*p < .001$	34%	15	$*p < .001$			
-TU	56%	23	$*p < .001$	94%	45	$p = 1$	93%	41	$*p < .001$			
INT	95%	39	$p = 1$	96%	46	$p = .667$	93%	41	$p = 1$			
ECQ	7%	3	$p = .077$	23%	11	$p = .487$	30%	13	$*p = .012$			
Total		41			48			44				

Table A3. Variant rates as a function of polarity over time (asterisks identify the favorable context of statistically significant hierarchies)

Variant	Polarity context	19C			20C			21C		
		%	N	Total N	%	N	Total N	%	N	Total N
P-INV	Negative	1%	2	169	0%	0	88	0%	0	60
	Affirmative	47%*	289	620	30%*	205	683	4%	18	464
	p-value	$p < .001$			$p < .001$			$p < .247$		
-TU	Negative	1%	1	169	9%	8	88	2%	1	60
	Affirmative	25%*	157	620	36%*	246	683	48%*	222	464
	p-value	$p < .001$			$p < .001$			$p < .001$		
INT	Negative	98%*	166	169	91%*	80	88	98%*	59	60
	Affirmative	28%	171	620	27%	182	683	42%	194	464
	p-value	$p < .001$			$p < .001$			$p < .001$		
ECQ	Negative	0%	0	169	0%	0	88	0%	0	60
	Affirmative	0%	3	620	7%*	50	683	6%*	30	464
	p-value	$p = 1$			$p < .004$			$p < .037$		

Table A4. Variant rates by grammatical person over time

Variant	Grammatical person	19C			20C			21C		
		%	N	Total N	%	N	Total N	%	N	Total N
P-INV	<i>tu</i>	64%	183	284	48%	133	277	7%	15	221
	<i>vous</i>	67%	101	150	68%	71	105	33%	3	9
	other	3%	5	183	0%	1	251	0%	0	204
	<i>tu</i> versus <i>vous</i>	$p = .596$			$*p < .001$			$*p = .025$		
	<i>tu</i> versus other	$*p < .001$			$*p < .001$			$*p < .001$		
	<i>vous</i> versus other	$*p < .001$			$*p < .001$			$*p < .001$		

Variant	Grammatical person	19C			20C			21C		
		%	N	Total N	%	N	Total N	%	N	Total N
-TU	<i>tu</i>	12%	34	284	30%	83	277	54%	119	221
	<i>vous</i>	3%	5	150	0%	0	105	0%	0	9
	other	64%	118	183	65%	163	251	50%	103	204
	<i>tu</i> versus <i>vous</i>	* <i>p</i> = .002			* <i>p</i> < .001			* <i>p</i> = .001		
	<i>tu</i> versus other	* <i>p</i> < .001			* <i>p</i> < .001			<i>p</i> = .498		
	<i>vous</i> versus other	* <i>p</i> < .001			* <i>p</i> < .001			* <i>p</i> = .003		
INT	<i>tu</i>	24%	67	284	22%	61	277	39%	87	221
	<i>vous</i>	29%	44	150	32%	34	105	67%	6	9
	other	33%	60	183	35%	87	251	50%	101	204
	<i>tu</i> versus <i>vous</i>	<i>p</i> = .204			* <i>p</i> = .046			<i>p</i> = .163		
	<i>tu</i> versus other	* <i>p</i> = .033			* <i>p</i> = .001			* <i>p</i> = .040		
	<i>vous</i> versus other	<i>p</i> = .053			<i>p</i> = .714			<i>p</i> = .498		

Table A5. *P*-values of differences in variant rate by grammatical person over time (Ns in Table A4)

Variant	<i>tu</i>		<i>vous</i>		Other persons	
	19C → 20C	20C → 21C	19C → 20C	20C → 21C	19C → 20C	20C → 21C
P-INV	* <i>p</i> < .001	* <i>p</i> < .001	<i>p</i> = 1	<i>p</i> = .064	<i>p</i> = .087	<i>p</i> = 1
-TU	* <i>p</i> < .001	* <i>p</i> < .001	<i>p</i> = .080	<i>p</i> = 1	<i>p</i> = .919	* <i>p</i> = .002
INT	<i>p</i> = .688	* <i>p</i> < .001	<i>p</i> = .679	<i>p</i> = .064	<i>p</i> = .758	* <i>p</i> = .002

Table A6. Rate of -TU (relative to INT only) by subject over time

	19C		19C → 20C	20C		20C → 21C	21C	
	%	N	<i>p</i> -value	%	N	<i>p</i> -value	%	N
<i>ce/ça</i>	65%	60/90	<i>p</i> = .575	60%	77/128	<i>p</i> = .792	58%	64/110
<i>il(s)/elle(s)</i>	68%	27/40	<i>p</i> = .186	52%	23/44	<i>p</i> = .065	33%	18/55
<i>je</i>	67%	12/18	<i>p</i> = .126	89%	24/27	<i>p</i> = .394	79%	11/14
Nouns	69%	11/16	<i>p</i> = .621	86%	6/7	<i>p</i> = .183	33%	1/3
<i>on</i>	73%	8/11	<i>p</i> = 1	77%	17/22	<i>p</i> = .158	43%	3/7
Impersonal <i>il</i>	-	0/0	<i>p</i> = 1	73%	16/22	<i>p</i> = .266	50%	6/12
Total	66%	118/178		65%	163/250		50%	103/201

Cite this article: Dion N (2023). A question of change: Putting five complementary measures to the test with French polar interrogatives. *Language Variation and Change* 35, 247–271. <https://doi.org/10.1017/S0954394523000170>