



Language Choice, Mixed Inner Speech, and Code-Switching in Putonghua-Dialect Speakers

Ran Qiao

Southeast University, China

Corresponding Author Email: qr15632952527@163.com

DOI: <https://doi.org/10.36892/ijlls.v8i5.2771>

APA Citation: Qiao, R. (2026). Language Choice, Mixed Inner Speech, and Code-Switching in Putonghua-Dialect Speakers. *International Journal of Language and Literary Studies*, 8(5).319-340. <https://doi.org/10.36892/ijlls.v8i5.2771>

Received:

20/06/2026

Accepted:

20/08/2026

Keywords:

inner speech;
language
choice; code-
switching;
Putonghua;
dialect identity

Abstract

Inner speech is usually treated as private cognition, yet its language may reflect socially organized repertoires. This mixed-methods study examined language choice, mixed inner speech, and within-episode code-switching among Putonghua-dialect users in China. Phase 1 involved 360 adults completing a fully crossed $2 \times 3 \times 2$ vignette task manipulating formality, imagined interlocutor, and emotional intensity; 90 entered a seven-day experience-sampling phase, 79 met the confirmatory compliance threshold, and 30 completed interviews. Formal contexts reduced dialect dominance ($OR = 0.50$), whereas hometown-intimate interlocutors strongly increased it relative to institutional authorities ($OR = 25.82$). High emotional intensity ($OR = 1.78$), productive dialect access, habitual dialect use, and dialect identity also increased dialect availability. Hometown-intimate contexts, emotion, and productive access predicted mixing, while identifiable switching was more likely in episodes containing multiple remembered or imagined voices. Cross-method correspondence was very strong for dominance ($r = .943$), weaker for mixing, and absent for switching. Interviews showed that dialect often appeared as remembered voices, recurrent phrases, or lexical fragments. The findings support a contextualized account in which social cues interact with speaker-specific repertoire conditions and show that dominance, mixing, and switching are distinct outcomes.

1. Introduction

Language choice is usually examined as observable social practice, but linguistic selection also occurs without overt speech. In mainland China, Putonghua is associated with education, institutions, mobility, and cross-regional communication, whereas regional Sinitic varieties (fangyan, hereafter dialects) may index family, local belonging, humor, and intimacy. Here, “dialect” is used as a socially situated gloss for fangyan, not as a claim of mutual intelligibility or linguistic hierarchy; many regional Sinitic varieties are mutually unintelligible and may be treated as distinct languages in other classificatory traditions. These associations are socially learned and vary across speakers and situations.

Inner speech is verbally mediated thought without audible articulation and supports planning, self-regulation, memory, problem solving, emotion, and interactional rehearsal (Alderson-Day & Fernyhough, 2015a; de Guerrero, 2018; Perrone-Bertolotti et al., 2014). It can be condensed or expanded, monologic or dialogic, and may include remembered or imagined voices (Alderson-Day et al., 2018; McCarthy-Jones & Fernyhough, 2011). Its private location therefore does not make it socially unstructured.

The Putonghua–dialect ecology does not map neatly onto a simple L1–L2 distinction. A locally acquired variety may weaken in active production after schooling or migration while Putonghua becomes dominant in education, work, media, and peer interaction. Speakers with similar proficiency may also differ in dialect attachment and maintenance, making internal language selection a product of both immediate context and speaker-specific repertoire conditions.

Three outcomes are distinguished. Inner-speech language choice concerns the relative dominance of Putonghua or dialect within a verbal episode. Mixed inner speech means that both varieties are present, whether or not they can be segmented. Within-episode code-switching is narrower: an identifiable transition from one variety to the other. A mixed episode may therefore contain no discrete switch.

Multilingual inner-speech research links internal language use to proficiency, acquisition history, habitual use, emotional salience, and autobiographical experience (Dewaele, 2015; Pavlenko, 2014; Resnik, 2021). Experience-sampling research also documents spontaneous inner-speech code-switching shaped by recent context and longer-term language history (Dallashah Khatib et al., 2026). Less is known about socially stratified standard–regional repertoires, imagined interlocutors, productive access, and dialect identity.

To address this gap, the study develops a Contextualized Internal Language Selection Framework. Formality, imagined interlocutor, and emotional intensity are treated as proximal cues whose effects may vary with productive access, language dominance, habitual use, identity, mobility, maintenance, and perceived Putonghua–dialect distance. Audience design motivates the role of imagined addressees (Bell, 1984), while indexical approaches explain how institutional competence, local affiliation, intimacy, and distance become contextually activated (Eckert, 2008; Ochs, 1992).

The study combines a fully crossed $2 \times 3 \times 2$ factorial vignette task, a seven-day experience-sampling phase, and semi-structured interviews, allowing experimentally elicited reports to be compared with momentary daily-life reports.

It contributes by extending multilingual inner-speech research to a Putonghua–dialect ecology, separating dominance, mixing, and identifiable switching, and testing how contextual cues interact with repertoire structure and identity across controlled, momentary, and qualitative evidence.

2. Literature Review

2.1 Inner speech as socially situated verbal thought and a measurement problem

In the Vygotskian tradition, inner speech develops through the internalization and transformation of socially organized speech and supports planning and self-regulation (Vygotsky, 1987). Contemporary accounts treat it as heterogeneous—dialogic or monologic, condensed or expanded, and sometimes associated with other people’s voices (Alderson-Day et al., 2018; McCarthy-Jones & Fernyhough, 2011). An imagined parent, teacher, supervisor, or friend can therefore function as an internalized participant.

Measurement is difficult because questionnaires, vignette tasks, and experience sampling capture related but non-identical objects (Morin et al., 2018; Sabag et al., 2023). Generalized self-reports may diverge from momentary reports and may overestimate some forms of inner experience (Alderson-Day & Fernyhough, 2015b; Heavey et al., 2019). Accordingly, vignettes are treated as contextually elicited reports rather than direct recordings of cognition; ESM increases temporal proximity, and interviews clarify how participants reconstruct episodes. The study therefore examines correspondence across methods rather than assuming equivalence.

2.2 Internal language choice, mixed inner speech, and code-switching

Multilingual speakers may draw on different languages for planning, emotion, autobiographical memory, and self-communication. Reported inner-speech language use varies with acquisition history, proficiency, current use, and emotional attachment (Dewaele, 2015;

Pavlenko, 2014; Resnik, 2021), supporting a dynamic rather than first-language-default account.

Conceptual precision is essential. Language choice concerns relative dominance; mixed inner speech indicates the co-presence of both resources; and code-switching is treated as the subset of mixed episodes containing an identifiable transition (Poplack, 1980). This follows interactional work that analyzes switching as patterned alternation rather than as any evidence of multiple available languages (Auer, 1998; Gardner-Chloros, 2009; Green & Abutalebi, 2013; Gumperz, 1982).

Recent ESM research documents spontaneous multilingual switching in inner speech (Dallaseh Khatib et al., 2026). Putonghua–dialect repertoires add a further complication because varieties may be structurally related yet socially differentiated (Zhong et al., 2024), while bidialectal switching varies with dominance and use (Declerck & Kirk, 2021; Zeng et al., 2024). The present study therefore models dominance, productive access, and perceived distance and analyzes outcomes hierarchically: verbal presence, relative dominance, mixed composition, and—among mixed episodes—identifiable switching.

2.3 Imagined interlocutors, audience design, identity, and indexicality

The dialogicality of inner speech links it to audience orientation. Bell's (1984) audience-design model proposes stylistic adjustment in relation to addressees and audience roles. Extended cautiously to internal rehearsal, an imagined supervisor, non-local peer, parent, or hometown friend may activate different expectations about authority, intimacy, appropriateness, and shared linguistic knowledge; multilingual online practice likewise shows orientation to imagined audiences (Liu, 2021).

Indexical theory explains why such audience effects need not be uniform: linguistic forms acquire associations with stances, activities, personae, and identities through social practice (Eckert, 2008; Ochs, 1992). Identity is therefore treated as relational and situated rather than a fixed demographic possession (Bucholtz & Hall, 2005). Dialect identity here captures local belonging, relational-affective attachment, and maintenance commitment, while productive access is kept distinct because receptive speakers may recognize familiar dialect voices or fragments without sustaining self-generated dialect inner speech.

2.4 Dialect identity, language dominance, mobility, and maintenance in contemporary China

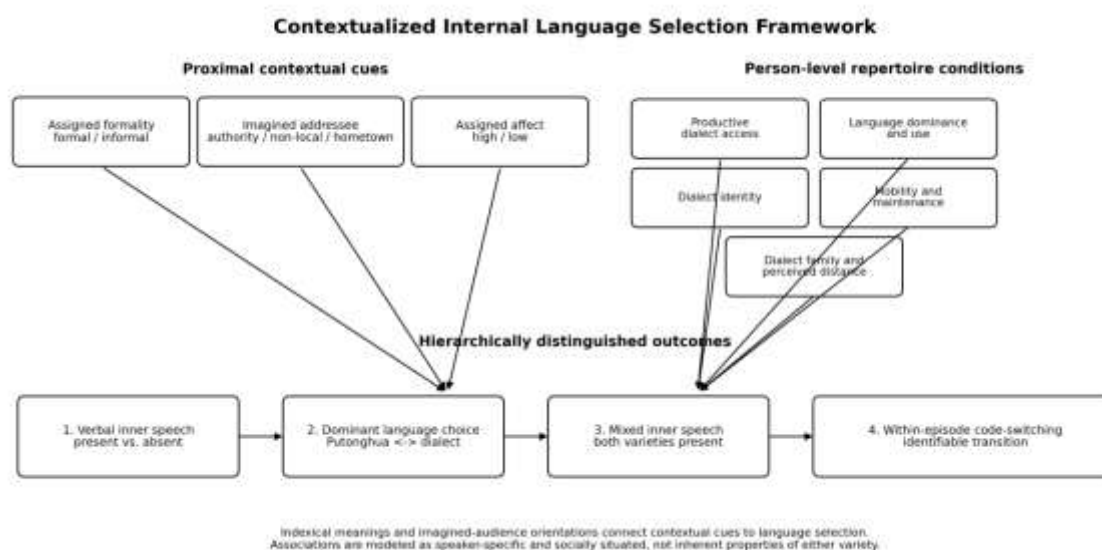
Putonghua and regional Sinitic varieties occupy unequal but changing positions. Putonghua is institutionally supported in education, administration, national media, and cross-regional mobility, whereas regional varieties remain resources for place, family, solidarity, humor, and cultural memory; attitudes vary across provinces (Garrett, 2010; Li et al., 2023), and young speakers may reconstruct dialect identity through digital participation (Mei, 2024).

Language dominance complements identity because it captures history, current use, and proficiency rather than symbolic attachment alone; the Bilingual Language Profile provides a multidimensional basis for this distinction (Birdsong et al., 2012; Olson, 2023). Mobility can reorganize language–place relations rather than simply reduce dialect use (Johnstone et al., 2006), and internal migrants in China negotiate Putonghua and hometown varieties through capital, integration, and heritage maintenance (Yang & Shang, 2026). Accordingly, years away, maintenance, current use, perceived distance, and broad dialect family are modeled separately, with family-level analyses treated as sensitivity tests rather than essential differences.

2.5 Conceptual framework, research questions, and hypotheses

Figure 1 summarizes the framework. It distinguishes proximal contextual cues from person-level repertoire conditions and predicts four analytically separate outcomes: verbal inner-speech presence, dominant language choice, mixed inner speech, and within-episode code-switching. The model predicts context-dependent tendencies and interactions rather than intrinsic emotional or cognitive properties of either variety.

Figure 1
Contextualized Internal Language Selection Framework



RQ1. How should verbal inner-speech episodes among Putonghua-dialect speakers be distributed across dominant language choice, mixed inner speech, and identifiable within-episode code-switching?

RQ2. How do assigned formality, imagined interlocutor, and emotional intensity affect these outcomes in a fully crossed factorial vignette design?

RQ3. How are contextual effects moderated by productive dialect access, language dominance, habitual use, dialect identity, mobility, dialect maintenance, dialect family, and perceived Putonghua-dialect distance?

RQ4. To what extent do vignette-elicited patterns correspond to momentarily sampled inner speech in daily life, and how do participants explain convergent, divergent, and counter-pattern cases?

H1. Assigned formal contexts will increase the probability of Putonghua-dominant inner speech relative to informal contexts, controlling for participant and vignette variation.

H2. Imagining a hometown intimate will increase dialect-dominant language choice and mixed inner speech relative to imagining an institutional authority or non-local peer; imagining an institutional authority will show the strongest Putonghua orientation.

H3. High assigned emotional intensity will increase dialect-dominant and mixed inner speech, but this effect will be stronger among participants with higher dialect identity, higher habitual dialect use, and productive dialect access.

H4. Productive dialect users will report more dialect-dominant inner speech, mixed inner speech, and identifiable switching than receptive-dominant users after continuous proficiency, dominance, use, and dialect distance are controlled.

H5. Greater Putonghua dominance and longer residence away from the hometown will be associated with lower dialect availability, whereas dialect maintenance practices and identity will attenuate these associations.

H6. The direction of the principal contextual effects will replicate in experience-sampling data, although effect sizes may differ because momentary reports are not constrained by experimentally standardized scenarios.

3. Methodology

3.1 Research design

A multiphase explanatory mixed-methods design combined a factorial vignette experiment with an embedded experience-sampling (ESM) phase and follow-up interviews. Phase 1 included a linguistic-background questionnaire, the Dialect Identity Scale, and a fully crossed vignette task. Phase 2 followed a stratified subsample for seven days of ESM and then selected interviewees from quantitative and momentary profiles. Qualitative analysis was used to explain, qualify, and challenge the statistical patterns.

3.2 Participants and sampling strategy

Phase 1 yielded a final analysis sample of 360 adults after prespecified exclusions. Eligibility required age 18 or older, everyday Putonghua use, and sustained childhood, family, hometown, or community exposure to a regional Sinitic variety. Local Mandarin and non-Mandarin Sinitic varieties were included, whereas cases involving only a non-Sinitic minority language or foreign language as the second resource were excluded. Recruitment drew on university mailing lists ($n = 111$, 30.8%), workplace networks ($n = 79$, 21.9%), social media ($n = 89$, 24.7%), hometown associations ($n = 36$, 10.0%), and community referrals ($n = 45$, 12.5%).

Productive dialect users had listening and speaking ratings of at least 4 on seven-point scales and could produce spontaneous multi-sentence speech; receptive-dominant users had listening ratings of at least 4 but speaking ratings of 3 or below. Intermediate cases remained in continuous analyses but were excluded from the categorical contrast. Ninety participants entered ESM with balancing across access group, dominance, identity, mobility, and broad dialect family; 30 were selected for interviews through maximum-variation and negative-case sampling.

3.3 Conceptual and operational distinctions

Table 1 summarizes the three principal language outcomes. Definitions and examples in participant instructions avoided implying that any response was more bilingual, authentic, or desirable than another.

Table 1

Distinguishing Language Choice, Mixed Inner Speech, and Within-Episode Code-Switching

Construct	Conceptual definition	Primary operational measure	Primary model
Inner-speech language choice	The relative dominance of Putonghua or dialect within a verbal inner-speech episode. It does not require both varieties to occur.	Seven-point ordered dominance rating from exclusively Putonghua to exclusively dialect, analyzed only when verbal inner speech is reported.	Cumulative-link mixed-effects model.
Mixed inner speech	The presence of both Putonghua and dialect within the same episode, whether sequentially, through quotation, through different imagined voices, or in an unsegmentable mixture.	Binary both-varieties-present response plus optional subtype and example.	Mixed-effects binary logistic model; multinomial sensitivity analysis for composition type.
Within-episode code-switching	An identifiable transition from one variety to the other within the same episode. This is a subset of mixed inner speech.	Among mixed episodes: switch present/absent/uncertain, direction, approximate number of switches, and perceived trigger.	Mixed-effects binary logistic model among classifiable mixed episodes.

3.4 Measures

3.4.1 Linguistic background, dominance, and mobility

The background questionnaire recorded demographics, hometown and current residence, named variety, acquisition history, family language, current use, and language use with major relationship categories. Putonghua and dialect listening and speaking proficiency were rated separately; reading and writing were descriptive because many regional varieties lack standardized literacy practices.

Language dominance adapted the history, current-use, and self-rated-proficiency modules of the Bilingual Language Profile (Birdsong et al., 2012; Olson, 2023). Positive scores indicated greater Putonghua dominance and negative scores greater dialect dominance; attitude items were excluded because identity was measured separately. Mobility and maintenance covered years away, return visits, family contact, online dialect use, hometown networks, and deliberate maintenance or relearning.

3.4.2 Dialect classification and heterogeneity controls

Participants named and located their variety and rated its mutual intelligibility and phonological, lexical, and grammatical distance from Putonghua. Two trained coders mapped varieties to regional Mandarin, Wu, Yue, Min, Hakka, Gan, Xiang, Jin, or other classifiable Sinitic groups, with specialist consultation for unresolved cases. Broad family and perceived distance were retained as heterogeneity controls or sensitivity variables.

3.4.3 Dialect Identity Scale

The 18-item Dialect Identity Scale measured local belonging and self-definition, relational-affective attachment, and maintenance commitment, with six items per dimension rated from 1 (strongly disagree) to 7 (strongly agree). Items that directly equated dialect with anger, comfort, or authentic emotional expression were excluded to avoid criterion contamination. Item development followed expert review and cognitive interviewing principles (Boateng et al., 2018).

3.4.4 Fully crossed factorial vignette task

The central instrument was a $2 \times 3 \times 2$ within-participant factorial vignette task (Su & Steiner, 2020): formality (formal/informal) $\times$ imagined interlocutor (institutional authority/non-local peer/hometown intimate) $\times$ emotional intensity (high/low), yielding 12 orthogonal conditions. Each condition had two parallel versions, creating a 24-item pool; each participant received one randomly assigned version from every cell in randomized order. Parallel versions reduced topic-specific confounding, and pilot ratings were used to balance non-target dimensions.

After each vignette, participants reported whether identifiable verbal inner speech was present. Verbal episodes were rated on a seven-point scale from exclusively Putonghua to exclusively dialect and classified as Putonghua only, dialect only, both, another language or variety, or uncertain. Mixed episodes were followed by questions about identifiable switching, first-switch direction, approximate switch count, represented voices, and an optional non-sensitive example. Assigned factors were confirmatory predictors; perceived formality, felt intensity, closeness, power distance, vividness, realism, and confidence were manipulation checks or exploratory perceptions.

Table 2 shows the factorial structure. Each participant sees one A or B version in every cell.

Table 2
Fully Crossed $2 \times 3 \times 2$ Factorial Vignette Design

Formality	Imagined interlocutor	Emotion	Versions in pool	Items per participant
Formal	Institutional authority	Low	2 parallel versions	1 randomly assigned

Formality	Imagined interlocutor	Emotion	Versions in pool	Items per participant
Formal	Institutional authority	High	2 parallel versions	1 randomly assigned
Formal	Non-local peer	Low	2 parallel versions	1 randomly assigned
Formal	Non-local peer	High	2 parallel versions	1 randomly assigned
Formal	Hometown intimate	Low	2 parallel versions	1 randomly assigned
Formal	Hometown intimate	High	2 parallel versions	1 randomly assigned
Informal	Institutional authority	Low	2 parallel versions	1 randomly assigned
Informal	Institutional authority	High	2 parallel versions	1 randomly assigned
Informal	Non-local peer	Low	2 parallel versions	1 randomly assigned
Informal	Non-local peer	High	2 parallel versions	1 randomly assigned
Informal	Hometown intimate	Low	2 parallel versions	1 randomly assigned
Informal	Hometown intimate	High	2 parallel versions	1 randomly assigned

3.4.5 Experience-sampling substudy

For seven consecutive days, the ESM application issued four semi-random prompts per day between 09:00 and 22:00, separated by at least 90 minutes. Participants reported the experience immediately preceding the alert after training that distinguished verbal inner speech from images, feelings, unsymbolized thought, and responses constructed only after the prompt.

Each prompt repeated the vignette measures of verbal presence, language composition, dominance, and switching and recorded activity, location, social company, imagined interlocutor, emotional valence and arousal, recent language exposure and production, and thought function. The confirmatory compliance threshold was at least 17 of 28 prompts; lower-compliance participants were retained in attrition summaries but excluded from confirmatory ESM models.

3.4.6 Semi-structured interviews

Semi-structured interviews were individualized from each participant's vignette and ESM profile. They probed Putonghua-dominant, dialect-dominant, mixed, switching, and non-verbal episodes; distinguished participants' own voices from remembered or imagined voices; and tested counterfactual changes in interlocutor, emotional stakes, or setting. Examples were treated as retrospective reconstructions, and dialect expressions were translated when analytically relevant.

3.5 Pilot study and manipulation checks

Pilot cognitive interviews refined the distinctions among verbal inner speech, mixed inner speech, and switching. Pilot ratings assessed intended formality, emotion, power, and closeness contrasts; unbalanced or weakly discriminating vignette versions were revised before the main study.

3.6 Data collection procedure

After informed consent, Phase 1 participants completed the background questionnaire, Dialect Identity Scale, attention checks, and randomized vignette task. Emotion–language questions were placed after the vignette task to reduce priming. The ESM subsample then

completed standardized training and the seven-day protocol, after which interviewees were selected.

3.7 Quantitative analysis

3.7.1 Data screening, missing data, and scale validation

Exclusions followed prespecified criteria for ineligibility, duplicate participation, substantial missingness, multiple failed attention checks, implausibly short completion times, and mechanically invariant responding combined with failed comprehension checks. Item-level missingness, ESM nonresponse, and interview attrition were tracked separately; the Dialect Identity Scale was summarized with internal-consistency estimates and subscale scores.

3.7.2 Factorial vignette models

The primary vignette outcome was the ordered seven-point Putonghua–dialect dominance rating among verbal episodes. A cumulative-link mixed-effects model estimated assigned formality, imagined interlocutor, emotional intensity, and prespecified interactions, with crossed participant and vignette-item random intercepts and by-participant random slopes where supported by design and convergence diagnostics (Baguley et al., 2022; Barr et al., 2013; Hox et al., 1991).

Person-level predictors included productive-access group, continuous dominance, dialect speaking and listening proficiency, dialect-use frequency, identity, age of acquisition, mobility, maintenance, perceived distance, and broad dialect family. Key cross-level tests involved emotion $\times$ identity, hometown interlocutor $\times$ identity, emotion $\times$ productive access, and formality $\times$ Putonghua dominance; continuous predictors were centered and scaled.

No-verbal-inner-speech, mixed-inner-speech, and identifiable-switching outcomes were modeled separately with mixed-effects logistic models. Switching analyses were restricted to classifiable mixed episodes, with uncertain cases described separately. Productive-versus-receptive effects were examined using both the categorical contrast and continuous proficiency measures.

3.7.3 Experience-sampling models and cross-method correspondence

ESM outcomes were analyzed with ordinal or generalized mixed models for observations nested within participants. Momentary predictors included activity, imagined interlocutor, emotional valence and arousal, recent language exposure and production, location, and social company; identity, dominance, productive access, mobility, and dialect family were considered person-level moderators.

Cross-method correspondence was assessed by comparing participant-level vignette estimates with aggregated ESM proportions and by testing whether analogous contextual tendencies appeared in daily-life episodes. Correlations were interpreted as correspondence rather than evidence that the methods measured an identical construct.

3.7.4 Simulation-based power analysis

A priori power was evaluated by Monte Carlo simulation using pilot-informed outcome probabilities, random-effect variation, missingness, and the repeated 12-vignette structure (Green & MacLeod, 2016; Kumle et al., 2021). A target of 360 provided at least 80% simulated power for the weakest primary interaction under the prespecified assumptions.

3.8 Qualitative analysis and explicit treatment of counter-pattern cases

Interview transcripts and open-ended responses were analyzed with the Framework Method (Gale et al., 2013), combining deductive domains with inductive categories. A case-by-theme matrix compared access group, dialect family, dominance, identity, mobility, and quantitative profile. Negative-case analysis explicitly examined counter-patterns, and a second trained coder independently coded a purposively varied subset; disagreements refined category definitions and alternative interpretations.

3.9 Mixed-methods integration

Mixed-method integration followed connecting, building, and merging procedures (Fetters et al., 2013): quantitative profiles guided ESM and interview sampling, ESM discrepancies informed interview prompts, and joint displays aligned statistical effects with momentary evidence, qualitative themes, and counter-cases.

3.10 Preregistration, transparency, and open science

Confirmatory hypotheses, outcomes, exclusions, access-group thresholds, model strategy, interaction tests, and sensitivity analyses were specified before the main analyses. De-identified materials, codebooks, instruments, analysis scripts, and reproducible outputs were prepared for public release where consent and re-identification risk permitted; raw audio and highly identifying dialect material were excluded. The full vignette pool, Dialect Identity Scale, and core ESM prompt are provided as supplementary materials.

3.11 Ethical considerations

Formal ethical approval was not required under the applicable institutional requirements. Participants provided informed consent, could skip sensitive items, and were not required to provide verbatim private inner-speech content. ESM did not access private messages, microphones, or other device content and minimized precise location data. Interview audio and identifying information were stored separately from analytic data, and potentially identifying combinations of hometown, rare dialect background, occupation, and migration history were generalized.

4. Results and Discussion

4.1 Data screening, participant characteristics, and measurement quality

A total of 372 participants completed Phase 1. Twelve cases were excluded in accordance with the prespecified quality-control criteria: three likely duplicate participations, three cases with multiple failed attention checks, two cases with excessive missingness, two cases combining implausibly short completion times with invariant response patterns, and two participants who did not meet the sustained dialect-exposure criterion. The final Phase 1 analysis sample therefore consisted of 360 participants.

Participants ranged in age from 18 to 58 years ($M = 27.21$, $SD = 8.62$). The sample included 196 women (54.4%), 156 men (43.3%), three participants who selected a non-binary or another gender category (0.8%), and five who preferred not to report gender (1.4%). Based on the prespecified access criteria, 163 participants were classified as productive dialect users (45.3%), 145 as receptive-dominant users (40.3%), and 52 as intermediate users (14.4%). The sample covered nine broad Sinitic groupings, with regional Mandarin varieties forming the largest group ($n = 110$), followed by Min ($n = 52$), Wu ($n = 51$), Yue ($n = 43$), Xiang ($n = 36$), Gan ($n = 24$), Hakka ($n = 23$), Jin ($n = 11$), and other classifiable Sinitic varieties ($n = 10$).

The mean adapted Bilingual Language Profile score was 29.62 ($SD = 18.56$), with positive values indicating greater Putonghua dominance. Mean overall dialect use was .31 ($SD = .19$), and participants had lived away from their hometown for an average of 2.59 years ($SD = 2.10$). The 18-item Dialect Identity Scale showed high internal consistency ($\alpha = .94$); the belonging, relational-affective attachment, and maintenance subscales each showed $\alpha = .87$. The mean total identity score was 4.49 ($SD = 0.74$).

The factorial scenario manipulations operated as intended. Formal scenarios received higher perceived-formality ratings ($M = 5.70$, $SD = 0.29$) than informal scenarios ($M = 2.42$, $SD = 0.30$), $t(359) = 144.14$, $p < .001$. High-emotion scenarios received higher felt-intensity ratings ($M = 5.80$, $SD = 0.30$) than low-emotion scenarios ($M = 2.21$, $SD = 0.30$), $t(359) = 170.46$, $p < .001$. Closeness was highest for hometown intimates, followed by non-local peers and institutional authorities, $\chi^2(2) = 705.85$, $p < .001$; power distance showed the reverse ordering, $\chi^2(2) = 716.04$, $p < .001$.

Table 3
Sample, Completion, and Measurement Summary

Domain	Key result	Value
Phase 1 sample	Recruited / included / excluded	372 / 360 / 12
Access groups	Productive / receptive-dominant / intermediate	163 / 145 / 52
Scenario task	Completed trials / scheduled trials	4,269 / 4,320 (98.8%)
ESM	Completed prompts / scheduled prompts	1,947 / 2,520 (77.3%)
Interviews	Participants / coded segments	30 / 120
Dialect Identity Scale	Total-scale internal consistency	$\alpha = .94$

Note. The ESM confirmatory analyses included 79 participants who completed at least 17 of 28 prompts.

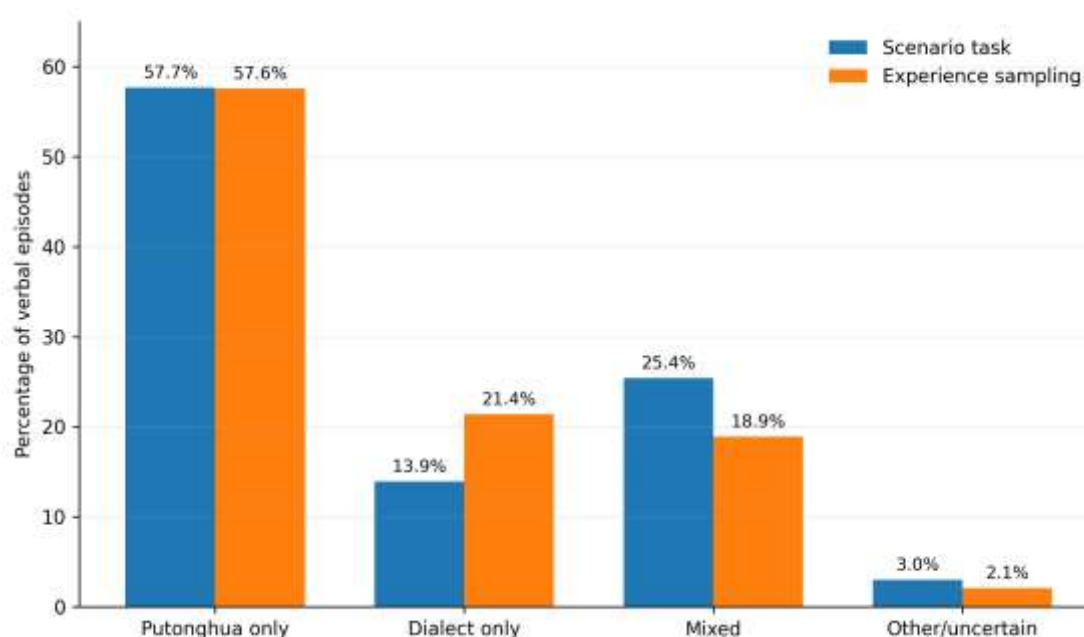
4.2 Distribution of inner-speech outcomes

The 360 included participants contributed 4,320 scheduled scenario trials, of which 4,269 were completed. Identifiable verbal inner speech was reported in 3,643 completed trials (85.3%). No identifiable verbal language was reported in 314 trials (7.4%), and 312 responses were classified as uncertain (7.3%).

Among the 3,643 verbal episodes, 2,101 were Putonghua-only (57.7%), 506 were dialect-only (13.9%), and 927 contained both varieties (25.4%). Sixty-five episodes were compositionally uncertain and 44 involved another language or variety. The mean dominance rating was 2.62 ($SD = 1.86$) on the scale from 1 (exclusively Putonghua) to 7 (exclusively dialect).

Of the 927 mixed episodes, 530 contained an identifiable transition between varieties (57.2%), 289 contained both varieties without an identifiable switch (31.2%), and 108 were uncertain (11.7%). Among switch-present episodes, 265 shifted from dialect to Putonghua (50.0%), 210 shifted from Putonghua to dialect (39.6%), and 55 followed another or more complex pattern (10.4%). The mean reported number of switches was 1.52 ($SD = 0.88$). Thus, the presence of both varieties did not by itself establish that a sequential switch had occurred.

Figure 2
Composition of Identifiable Verbal Inner-Speech Episodes in the Factorial Scenario Task and Experience-Sampling Phase



Note. Percentages are calculated within verbal episodes; ESM percentages are based on participants meeting the confirmatory compliance threshold.

4.3 Contextual and participant-level predictors in the factorial scenario task

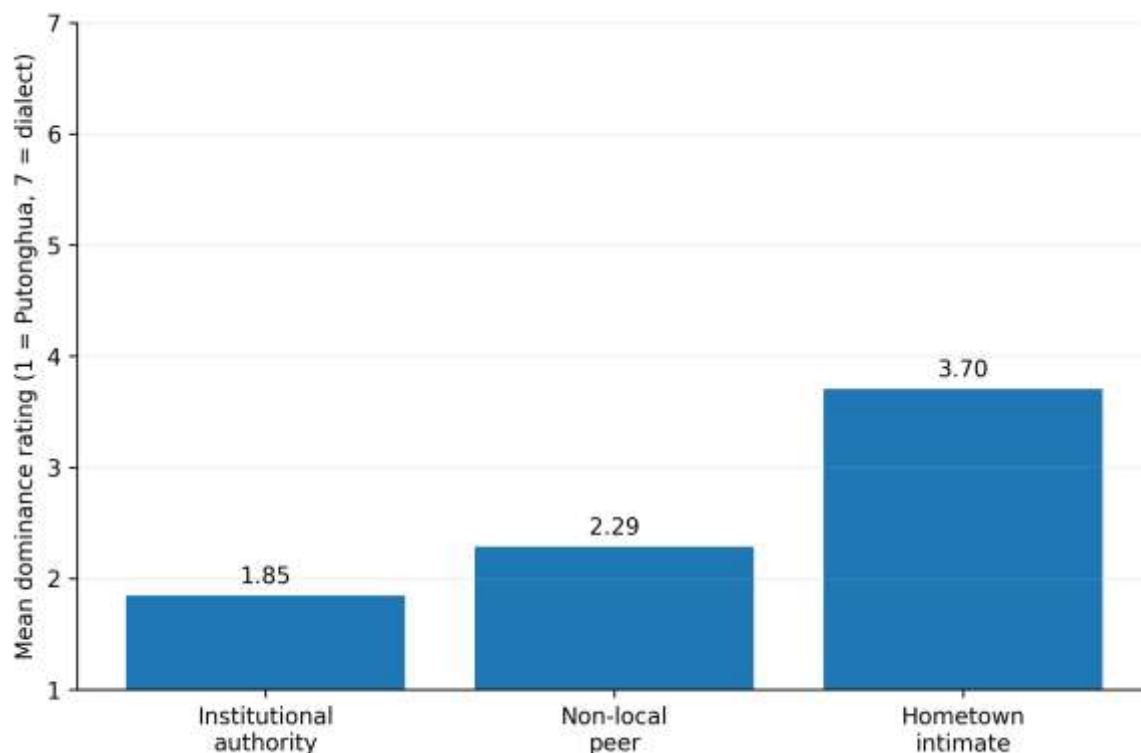
4.3.1 Dominant language choice

Higher values represented greater dialect dominance. Formal contexts reduced the odds of responding in a more dialect-dominant category relative to informal contexts, $OR = 0.50$, 95% CI [0.45, 0.57], $p < .001$. The mean dominance rating was 2.43 in formal scenarios and 2.82 in informal scenarios.

Imagined interlocutor produced the largest contextual difference. With institutional authorities as the reference category, non-local peers increased the odds of greater dialect dominance, $OR = 2.43$, 95% CI [2.09, 2.83], $p < .001$, whereas hometown intimates produced a much larger increase, $OR = 25.82$, 95% CI [20.98, 31.77], $p < .001$. A direct contrast further showed that hometown-intimate episodes were more dialect-oriented than non-local-peer episodes, $OR = 10.63$, 95% CI [8.95, 12.64], $p < .001$. The corresponding mean dominance ratings were 1.85 for institutional authorities, 2.29 for non-local peers, and 3.70 for hometown intimates.

Figure 3

Mean Putonghua–Dialect Dominance by Imagined-Interlocutor Condition in the Factorial Scenario Task



Note. Higher scores indicate greater dialect dominance.

High assigned emotional intensity increased dialect dominance, $OR = 1.78$, 95% CI [1.49, 2.13], $p < .001$. The mean rating was 2.77 in high-emotion conditions and 2.48 in low-emotion conditions. Productive users were more dialect-oriented than intermediate users, $OR = 2.27$, 95% CI [1.46, 3.53], $p < .001$, whereas receptive-dominant users were less dialect-oriented than intermediate users, $OR = 0.42$, 95% CI [0.27, 0.65], $p < .001$. The direct productive-versus-receptive contrast was significant, $OR = 5.45$, 95% CI [3.23, 9.21], $p < .001$.

A one-standard-deviation increase in Putonghua dominance reduced dialect availability, $OR = 0.57$, 95% CI [0.42, 0.79], $p < .001$. Habitual dialect use, $OR = 1.75$, 95% CI [1.20, 2.54], $p = .004$, dialect identity, $OR = 1.59$, 95% CI [1.23, 2.05], $p < .001$, and perceived Putonghua–

dialect distance, OR = 1.18, 95% CI [1.03, 1.34], $p = .016$, were positively associated with dialect dominance. Years away from the hometown was not significant, OR = 0.94, 95% CI [0.79, 1.12], $p = .465$; maintenance practices showed a positive but non-significant association, OR = 1.23, 95% CI [0.98, 1.53], $p = .074$.

The formality-by-Putonghua-dominance interaction was significant, OR = 0.88, 95% CI [0.79, 0.99], $p = .039$. The remaining prespecified interactions involving emotion, identity, productive access, mobility, and maintenance were not significant, all $ps > .12$.

4.3.2 Mixed inner speech and identifiable switching

Formality did not independently predict the presence of both varieties, OR = 0.94, 95% CI [0.82, 1.09], $p = .427$. Relative to institutional authorities, hometown intimates increased the probability of mixed inner speech, OR = 1.59, 95% CI [1.33, 1.90], $p < .001$; hometown-intimate episodes were also more likely to be mixed than non-local-peer episodes, OR = 1.63, 95% CI [1.37, 1.93], $p < .001$. High emotional intensity increased mixing, OR = 1.33, 95% CI [1.07, 1.65], $p = .010$. Productive users were more likely than receptive-dominant users to report mixed episodes, OR = 1.77, 95% CI [1.25, 2.49], $p = .001$. Perceived dialect distance also predicted mixing, OR = 1.12, 95% CI [1.03, 1.22], $p = .006$.

The switching model included 819 mixed episodes classified as either switch-present or switch-absent. Episodes involving more than one remembered or imagined voice were more likely to contain an identifiable switch than single-voice episodes, OR = 1.81, 95% CI [1.27, 2.59], $p = .001$. Productive users were also somewhat more likely than receptive-dominant users to identify a switch, OR = 1.62, 95% CI [1.00, 2.62], $p = .049$. Formality, emotion, interlocutor condition, identity, dominance, habitual use, mobility, maintenance, and perceived distance did not independently predict switching once an episode had already been classified as mixed.

Table 4
Selected Primary Estimates From the Factorial Scenario Models

Outcome	Predictor or contrast	OR	95% CI	p
Dominance	Formal vs. informal	0.50	[0.45, 0.57]	< .001
Dominance	Non-local peer vs. authority	2.43	[2.09, 2.83]	< .001
Dominance	Hometown intimate vs. authority	25.82	[20.98, 31.77]	< .001
Dominance	High vs. low emotion	1.78	[1.49, 2.13]	< .001
Dominance	Productive vs. receptive-dominant	5.45	[3.23, 9.21]	< .001
Dominance	Putonghua dominance (1 SD)	0.57	[0.42, 0.79]	< .001
Dominance	Dialect identity (1 SD)	1.59	[1.23, 2.05]	< .001
Mixed	Hometown intimate vs. authority	1.59	[1.33, 1.90]	< .001
Mixed	High vs. low emotion	1.33	[1.07, 1.65]	.010
Mixed	Productive vs. receptive-dominant	1.77	[1.25, 2.49]	.001
Switch	Multiple voices vs. one voice	1.81	[1.27, 2.59]	.001

Note. For the ordinal outcome, ORs above 1 indicate greater odds of a more dialect-dominant response. Mixed and switch models are binary logistic models.

4.4 Experience-sampling results and cross-method correspondence

Ninety participants entered the seven-day experience-sampling phase. They received 2,520 prompts and completed 1,947, yielding an overall response rate of 77.3%. Seventy-nine participants met the confirmatory compliance criterion of at least 17 completed prompts and contributed 1,817 responses to the confirmatory analyses.

Identifiable verbal inner speech was reported in 1,251 responses (68.9%). Among these verbal episodes, 721 were Putonghua-only (57.6%), 268 were dialect-only (21.4%), 236 were mixed (18.9%), and 26 involved another language or variety (2.1%). The mean momentary dominance rating was 3.03 ($SD = 2.00$). Among the 236 mixed episodes, 127 contained an identifiable switch (53.8%), 87 did not contain a switch (36.9%), and 22 were uncertain (9.3%).

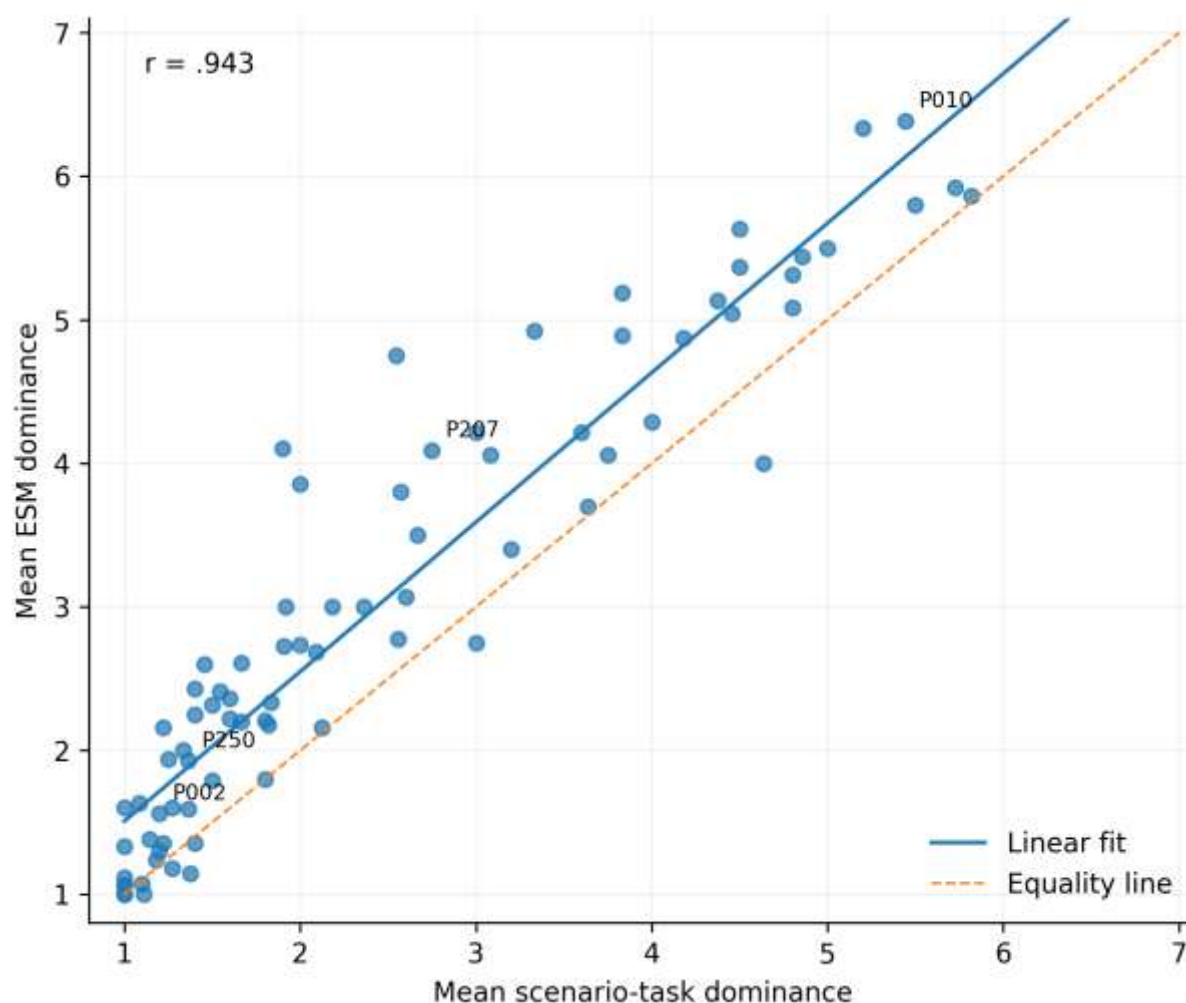
Work- or study-related episodes were less dialect-dominant than other activities, $OR = 0.61$, 95% CI [0.50, 0.75], $p < .001$. Imagining an institutional authority reduced dialect dominance relative to having no identifiable imagined interlocutor, $OR = 0.56$, 95% CI [0.39, 0.80], $p = .002$, whereas imagining a hometown intimate increased dialect dominance, $OR = 5.66$, 95% CI [4.10, 7.82], $p < .001$. The non-local-peer contrast was not significant. Recent dialect or mixed-language exposure during the preceding ten minutes increased momentary dialect dominance, $OR = 1.89$, 95% CI [1.63, 2.19], $p < .001$. Dialect identity also predicted greater momentary dialect dominance, $OR = 1.85$, 95% CI [1.26, 2.72], $p = .002$. Momentary arousal and absolute emotional valence were not significant.

In the ESM mixed-inner-speech model, non-local peers, $OR = 1.71$, 95% CI [1.15, 2.55], $p = .008$, and hometown intimates, $OR = 1.94$, 95% CI [1.38, 2.74], $p < .001$, increased the probability of mixing relative to episodes without an identifiable imagined interlocutor. Productive users were more likely than intermediate users to report mixing, $OR = 2.19$, 95% CI [1.18, 4.05], $p = .012$. Recent dialect exposure had a smaller positive association, $OR = 1.17$, 95% CI [1.01, 1.36], $p = .033$.

Participant-level dominance estimates showed strong cross-method correspondence, $r = .943$, $p < .001$. Correspondence was weaker for mixed-inner-speech rates, $r = .382$, $p < .001$, and verbal-inner-speech rates, $r = .282$, $p = .012$. No participant-level correspondence was found for the proportion of classifiable mixed episodes containing an identifiable switch, $r = -.028$, $p = .827$. Rates of identifiable verbal inner speech and mixing were also higher in the scenario task than in ESM. Figure 2 shows this difference.

Figure 4

Participant-Level Correspondence Between Mean Language-Dominance Ratings in the Factorial Scenario Task and the Experience-Sampling Phase



Note. Higher values indicate greater dialect dominance.

4.5 Qualitative findings and integration with the quantitative results

Thirty participants were selected for interviews through maximum-variation and negative-case sampling. The qualitative dataset contained 120 coded segments. The two coders initially agreed on 96 segments (80.0%), and all disagreements were subsequently reviewed and resolved. Six themes were retained: authority and formality (22 segments), switching triggers (22), mobility and maintenance (22), measurement and method differences (22), intimacy and remembered family voices (20), and productive constraints (12). Ten interview profiles were classified as cross-method convergent, three as complementary, and 17 as divergent. Because divergent and counter-pattern profiles were deliberately oversampled, these counts were not interpreted as population frequencies.

4.5.1 Authority, formality, and internal response preparation

The quantitative models showed that formal scenarios and institutional authorities shifted language choice toward Putonghua. Interview accounts located this shift at the stage of internal preparation: participants described silently arranging reasons, choosing respectful wording, and rehearsing a coherent sequence before speaking.

“When speaking to a teacher or supervisor, I first arrange the Putonghua sentence in my mind.” (P007, intermediate Xiang speaker)

The wording “first arrange” identified Putonghua as the medium of pre-formulation rather than merely the language eventually spoken aloud. P007’s average dominance was 1.40 in the scenario task and 2.25 in ESM, indicating a generally Putonghua-oriented profile despite somewhat greater dialect presence in daily life.

A contrasting pattern appeared when authority and local affiliation overlapped.

“Formal does not always mean Putonghua; if the other person is from my hometown, dialect can feel more forceful.” (P010, productive Yue speaker)

P010 reported high current dialect use (.764), a dialect-favoring dominance profile (BLP = -16.72), and mean inner-speech dominance ratings of 5.44 in the scenario task and 6.38 in ESM. Here, “more forceful” showed that a locally shared variety could construct seriousness or authority rather than only intimacy. By comparison, P126, a receptive-dominant Gan speaker with no reported current dialect use, had mean dominance ratings of 1.00 and 1.12 and described Putonghua preparation even in emotionally significant interactions. The contrast showed that the aggregate formality effect contained distinct speaker-level pathways.

4.5.2 Intimacy, remembered family voices, and productive constraints

Hometown-intimate scenarios produced the strongest dialect dominance and the highest mixed-inner-speech rate. The interviews showed, however, that this effect did not always take the form of a sustained self-generated dialect monologue; dialect could instead appear as a remembered family voice, a recurrent expression, or a short lexical fragment.

“When I imagine my mother’s tone, it is not a full dialect monologue but one or two phrases she often says.” (P072, productive Gan speaker)

Although P072 met the productive-access threshold and reported strong dialect identity (5.83), the participant remained Putonghua-oriented in both methods (scenario $M = 1.55$; ESM $M = 2.41$). The remembered voice supplied dialectal material without determining the language of the whole thought.

“For matters involving family, dialect seems to appear first, and then I organize it in Putonghua.” (P002, receptive-dominant regional Mandarin speaker)

P002 had low dialect use (.161), strong Putonghua dominance (BLP = 51.01), and low mean dominance ratings (1.20 in the scenario task; 1.56 in ESM), yet mixed episodes occurred in 30.0% of verbal scenario responses and 43.8% of verbal ESM reports. P340 showed a similar structure: despite no current overt dialect use and a Putonghua-dominant BLP score of 55.70, mixed episodes occurred in 27.3% of scenario responses and 23.8% of ESM responses. These cases showed why a low mean dominance score did not imply complete absence of dialect.

“I understand it and feel close to it, but producing many consecutive sentences in dialect, even internally, is difficult.” (P126, receptive-dominant Gan speaker)

The contrast between understanding and producing “many consecutive sentences” matched the productive-versus-receptive distinction: receptive access could support recognition, affective affiliation, and short fragments while sustained self-generated dialect speech remained constrained.

4.5.3 Identifiable switching as a change in function or voice

The quantitative switching model showed that multiple remembered or imagined voices increased the probability of an identifiable switch. Interviews distinguished two recurrent structures: a functional transition within one developing thought and a transition between socially differentiated voices.

“When emotion rises, one dialect phrase comes first; once I start analyzing what to do, I switch back to Putonghua.” (P002, receptive-dominant regional Mandarin speaker)

Here the switch boundary coincided with a change from immediate reaction to deliberate problem solving. The varieties were sequentially ordered, and the participant could identify both direction and functional change. This pattern corresponded to the dialect-to-Putonghua direction, which accounted for half of switch-present scenario episodes.

“The other person’s voice is in dialect while my own answer is in Putonghua, so both varieties appear.” (P010, productive Yue speaker)

This case involved role differentiation rather than a single undifferentiated internal speaker. The episode was mixed because two varieties were present, but it counted as switching only if the voices were experienced as a segmentable transition. The contrast therefore supported the article’s central distinction between mixed inner speech and identifiable within-episode code-switching.

4.5.4 Similar identity scores, different inner-language profiles

Dialect identity predicted dialect dominance at the sample level, but the cross-case matrix showed that similar identity scores could coexist with sharply different inner-language profiles. P072 and P337 both had total identity scores of 5.83. P072, a productive Gan speaker, reported overall dialect use of .301, Putonghua dominance (BLP = 35.32), and mean dominance scores of 1.55 and 2.41. P337, an intermediate regional Mandarin speaker, reported dialect use of .564, a near-balanced dominance score (BLP = 2.77), and mean dominance scores of 4.80 and 5.08. Thus, equally strong reported identity co-occurred with different levels of use, dominance, and productive automaticity.

A second contrast involved P207 and P070. Both were productive users with strong identity, but P207’s ESM dominance (4.09) was considerably higher than the scenario estimate (2.75), whereas P070 showed similar low-to-mid dialect orientation across methods (2.56 and 2.78). Identity therefore aligned with dialect availability in some profiles but did not determine its magnitude or cross-method stability.

4.5.5 Mobility, maintenance, and non-linear trajectories

Years away from the hometown did not show a significant independent effect in the primary quantitative model. Interview accounts indicated that mobility could reduce routine production while increasing awareness or emotional salience.

“After leaving home I used dialect less, but hearing it made homesickness more salient.” (P207, productive other-Sinitic speaker)

P207 had lived away for six years, reported substantial maintenance practices (4.90), and shifted from a scenario-task mean of 2.75 to an ESM mean of 4.09. Reduced use and increased salience therefore appeared simultaneously rather than as mutually exclusive outcomes.

“I deliberately use dialect on calls with family; otherwise it feels increasingly unfamiliar.” (P070, productive Yue speaker)

P070 had lived away for seven years and reported the highest maintenance score among the focal cases (5.35), identifying maintenance as an intentional practice rather than a direct consequence of residence duration. P362 provided a further contrast: despite five years away, very high current dialect use (.860) and a dialect-favoring BLP score were accompanied by stable dialect-dominant inner speech across both methods (5.82 and 5.86). Together, these cases explained why years away alone did not capture the direction of change.

4.5.6 Measurement-related divergence between the scenario task and ESM

The scenario task produced higher verbal-inner-speech and mixing rates than ESM. Participants explicitly distinguished a reconstructed likely response from a report of the immediately preceding experience.

“With vignettes I consider what response seems reasonable; phone prompts feel more like capturing what just happened.” (P250, receptive-dominant regional Mandarin speaker)

P250's mixed-inner-speech rate increased from 9.1% in the scenario task to 35.7% in ESM. The extract identified different reporting operations: the scenario task encouraged construction of a plausible response, whereas the phone prompt required attention to an already occurring episode. This provided a participant-level explanation for cross-method divergence rather than treating it as random inconsistency.

“Sometimes the prompt makes me realize there were no words, only images and feelings.” (P010, productive Yue speaker)

This report corresponded to the lower ESM verbal-inner-speech rate and clarified that absence of identifiable verbal language was itself a reportable outcome rather than a missing response. Together, the two extracts distinguished method-related reconstruction from momentary capture and explained why dominance converged more strongly across methods than mixing and identifiable switching.

Table 5

Mixed-Method Integration of the Principal Quantitative Patterns and Qualitative Themes

Theme	Quantitative pattern	Illustrative cases	Qualitative qualification
Authority and formality	Formal and authority conditions favored Putonghua.	P007 vs. P010/P126	Putonghua often organized institutional responses, but locally shared authority could activate dialect.
Intimacy and family voice	Hometown intimates increased dialect dominance and mixing.	P072, P002, P340	Dialect often appeared as a remembered voice, recurrent phrase, or lexical fragment rather than a full monologue.
Productive access	Productive users were more dialect-dominant and mixed more often.	P002/P340 vs. P010	Receptive speakers could recognize or quote dialect without sustaining self-generated dialect speech.
Identifiable switching	Multiple voices predicted identifiable switches.	P002 and P010	Identifiable switching could mark a functional transition or a change between imagined voices; some mixed episodes had no discrete switch.
Mobility and maintenance	Years away was not independently significant.	P207, P070, P362	Mobility reduced routine use for some speakers but increased salience or prompted deliberate maintenance.
Measurement mode	Dominance converged strongly; mixing and identifiable switching converged less.	P250 and P010	Scenarios elicited plausible reconstruction; ESM captured recent verbal and non-verbal experiences.

Note. Interview excerpts were anonymized and translated into English. Participant labels are pseudonymous.

4.6 Summary in relation to the hypotheses

H1 was supported: formal contexts increased Putonghua orientation. H2 was supported: institutional authorities produced the strongest Putonghua orientation, whereas hometown intimates produced the strongest dialect orientation and the highest probability of mixing. H3

received partial support: high-emotion scenarios increased dialect dominance and mixing, but the predicted interactions were not significant and momentary arousal and valence did not reproduce the scenario effect. H4 was supported for dominance and mixing and received limited support for identifiable switching. H5 received partial support: Putonghua dominance reduced dialect availability, while habitual dialect use and identity increased it; years away and maintenance did not show the full predicted pattern. H6 was partially supported: authority, hometown-interlocutor, productive-access, and identity patterns were directionally reproduced in ESM, but emotional effects and identifiable switching showed weaker cross-method correspondence.

4.7 Interpretation and Implications

Across the vignette task, ESM, and interviews, Putonghua–dialect inner speech varied with both immediate context and speaker-specific repertoire conditions. Institutional authorities favored Putonghua, hometown intimates increased dialect dominance and mixing, and productive users were more dialect-oriented, while receptive speakers still reported dialect fragments and remembered voices. These patterns reinforce the distinction among dominance, mixing, and identifiable switching.

4.7.1 Audience design and indexical social meaning

The formality and imagined-interlocutor effects extend audience design into silent rehearsal: participants oriented internally to socially differentiated addressees as they anticipated evaluation or reproduced familiar voices (Bell, 1984). P010's counter-case, in which local affiliation activated dialect in an authority context, further shows that Putonghua and dialect carry context-dependent indexical meanings rather than fixed associations with formality or intimacy (Eckert, 2008; Ochs, 1992).

4.7.2 Emotional intensity as a socially mediated cue

High-emotion scenarios increased dialect dominance and mixing, but momentary arousal and valence did not predict ESM dominance. Emotional language choice therefore cannot be reduced to an automatic preference for an earlier-acquired or locally meaningful variety; it depends on the relationship and activity in which emotion is embedded (Dewaele, 2015; Pavlenko, 2014).

4.7.3 Productive access, dominance, and identity

Productive users reported more dialect-dominant and mixed inner speech, whereas receptive knowledge often supported recognition, quotation, or affective resonance without sustained self-generated dialect speech. Continuous dominance, habitual use, and identity explained additional variation, supporting the treatment of productive access, language dominance, use, and identity as related but distinct repertoire dimensions (Birdsong et al., 2012; Olson, 2023).

4.7.4 Mixed inner speech is not synonymous with code-switching

Mixed inner speech was not synonymous with code-switching: only 57.2% of mixed scenario episodes contained an identifiable transition, while nearly one third contained both varieties without a clear switch. Here and throughout the discussion, switching refers only to a segmentable transition within a mixed episode. Interview reports showed remembered phrases, lexical fragments, functional transitions, and different imagined voices. These structures support analyzing co-presence and identifiable switching separately rather than treating all multilingual inner speech as sequential alternation (Auer, 1998; Poplack, 1980).

4.7.5 Mobility and maintenance as heterogeneous processes

Years away from the hometown was not independently significant once current dominance, use, identity, access, maintenance, and distance were considered. The interviews suggest why: mobility can simultaneously reduce routine production, heighten dialect salience, and prompt deliberate maintenance, consistent with a reorganization rather than uniform attrition of language–place relations (Johnstone et al., 2006).

4.7.6 What the scenario–ESM comparison reveals about measurement

Dominance converged strongly across vignette and ESM data, but mixing and identifiable switching did not. Scenarios encouraged plausible reconstruction, whereas ESM captured immediately preceding verbal and non-verbal experience. The additional ESM effect of recent language exposure further distinguishes longer-term repertoire organization from short-term linguistic activation (Alderson-Day & Fernyhough, 2015b; Dallasheh Khatib et al., 2026).

4.7.7 Theoretical and methodological contributions

The study extends multilingual inner-speech research to a standard-language–regional-variety ecology that is poorly represented by a simple L1–L2 contrast. Methodologically, the combination of factorial vignettes, ESM, and interviews separates controlled contextual effects, daily-life ecology, and participants' explanations. The hierarchy of verbal presence, dominance, mixed composition, and identifiable switching also offers a reusable measurement strategy.

4.7.8 Limitations and future directions

The study relies on self-reports of private experience, and ESM improves temporal proximity without providing direct access to cognition. The sample was relatively young and educated, dialect-family cells were uneven, and several repertoire variables were self-rated. Future work could add spontaneous production, independent distance measures, event-contingent sampling, and immediate reconstruction interviews, especially for rare identifiable-switching episodes or cases with difficult-to-segment transition boundaries.

5. Conclusion

Putonghua–dialect inner speech was structured jointly by immediate social context and longer-term repertoire conditions. Formal and institutional contexts generally favored Putonghua, whereas hometown-intimate contexts increased dialect dominance and mixed inner speech. Productive access, habitual dialect use, identity, perceived distance, and recent dialect exposure increased dialect availability; by contrast, years away from the hometown was not independently significant and maintenance showed only a positive, nonsignificant adjusted association. Counter-pattern cases showed that neither variety had a fixed social meaning.

Most importantly, dialect-dominant inner speech, mixed inner speech, and identifiable code-switching were empirically distinct. Speakers could remain Putonghua-dominant while experiencing dialectal fragments or remembered voices, and both varieties could be present without a clearly segmentable switch. Preserving these distinctions provides a more precise account of how socially meaningful linguistic repertoires enter silent verbal thought.

References

- Alderson-Day, B., & Fernyhough, C. (2015a). Inner speech: Development, cognitive functions, phenomenology, and neurobiology. *Psychological Bulletin*, 141(5), 931–965. <https://doi.org/10.1037/bul0000021>
- Alderson-Day, B., & Fernyhough, C. (2015b). Relations among questionnaire and experience sampling measures of inner speech: A smartphone app study. *Frontiers in Psychology*, 6, Article 517. <https://doi.org/10.3389/fpsyg.2015.00517>
- Alderson-Day, B., Mitrenga, K., Wilkinson, S., McCarthy-Jones, S., & Fernyhough, C. (2018). The Varieties of Inner Speech Questionnaire-Revised (VISQ-R): Replicating and refining links between inner speech and psychopathology. *Consciousness and Cognition*, 65, 48–58. <https://doi.org/10.1016/j.concog.2018.07.001>
- Auer, P. (Ed.). (1998). *Code-switching in conversation: Language, interaction and identity*. Routledge.

- Baguley, T., Dunham, G., & Steer, O. (2022). Statistical modelling of vignette data in psychology. *British Journal of Psychology*, 113(4), 1143–1163. <https://doi.org/10.1111/bjop.12577>
- Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, 68(3), 255–278. <https://doi.org/10.1016/j.jml.2012.11.001>
- Bell, A. (1984). Language style as audience design. *Language in Society*, 13(2), 145–204. <https://doi.org/10.1017/S004740450001037X>
- Birdsong, D., Gertken, L. M., & Amengual, M. (2012). *Bilingual Language Profile: An easy-to-use instrument to assess bilingualism*. Center for Open Educational Resources and Language Learning, University of Texas at Austin. <https://blp.coerll.utexas.edu>
- Boateng, G. O., Neilands, T. B., Frongillo, E. A., Melgar-Quinonez, H. R., & Young, S. L. (2018). Best practices for developing and validating scales for health, social, and behavioral research: A primer. *Frontiers in Public Health*, 6, Article 149. <https://doi.org/10.3389/fpubh.2018.00149>
- Bucholtz, M., & Hall, K. (2005). Identity and interaction: A sociocultural linguistic approach. *Discourse Studies*, 7(4–5), 585–614. <https://doi.org/10.1177/1461445605054407>
- Dallasheh Khatib, R., Eviatar, Z., & Kreiner, H. (2026). Spontaneous inner speech in multilinguals: Short-term and long-term effects on code-switching. *Journal of Cultural Cognitive Science*, 10(1), 49–63. <https://doi.org/10.1007/s41809-025-00191-9>
- Declerck, M., & Kirk, N. W. (2021). Is it easier to use one language variety at a time, or mix them? An investigation of voluntary language switching with bidialectals. *PLOS ONE*, 16(9), Article e0256554. <https://doi.org/10.1371/journal.pone.0256554>
- de Guerrero, M. C. M. (2018). Going covert: Inner and private speech in language learning. *Language Teaching*, 51(1), 1–35. <https://doi.org/10.1017/S0261444817000295>
- Dewaele, J.-M. (2015). From obscure echo to language of the heart: Multilinguals' language choices for (emotional) inner speech. *Journal of Pragmatics*, 87, 1–17. <https://doi.org/10.1016/j.pragma.2015.06.014>
- Eckert, P. (2008). Variation and the indexical field. *Journal of Sociolinguistics*, 12(4), 453–476. <https://doi.org/10.1111/j.1467-9841.2008.00374.x>
- Fetters, M. D., Curry, L. A., & Creswell, J. W. (2013). Achieving integration in mixed methods designs: Principles and practices. *Health Services Research*, 48(6 Pt 2), 2134–2156. <https://doi.org/10.1111/1475-6773.12117>
- Gale, N. K., Heath, G., Cameron, E., Rashid, S., & Redwood, S. (2013). Using the Framework Method for the analysis of qualitative data in multi-disciplinary health research. *BMC Medical Research Methodology*, 13, Article 117. <https://doi.org/10.1186/1471-2288-13-117>
- Gardner-Chloros, P. (2009). *Code-switching*. Cambridge University Press.
- Garrett, P. (2010). *Attitudes to language*. Cambridge University Press.
- Green, D. W., & Abutalebi, J. (2013). Language control in bilinguals: The adaptive control hypothesis. *Journal of Cognitive Psychology*, 25(5), 515–530. <https://doi.org/10.1080/20445911.2013.796377>
- Green, P., & MacLeod, C. J. (2016). SIMR: An R package for power analysis of generalized linear mixed models by simulation. *Methods in Ecology and Evolution*, 7(4), 493–498. <https://doi.org/10.1111/2041-210X.12504>

- Gumperz, J. J. (1982). *Discourse strategies*. Cambridge University Press.
- Heavey, C. L., Moynihan, S. A., Brouwers, V. P., Lapping-Carr, L., Krumm, A. E., Kelsey, J. M., Turner, D. K., II, & Hurlburt, R. T. (2019). Measuring the frequency of inner-experience characteristics by self-report: The Nevada Inner Experience Questionnaire. *Frontiers in Psychology*, 9, Article 2615. <https://doi.org/10.3389/fpsyg.2018.02615>
- Hox, J. J., Kreft, I. G. G., & Hermkens, P. L. J. (1991). The analysis of factorial surveys. *Sociological Methods & Research*, 19(4), 493–510. <https://doi.org/10.1177/0049124191019004003>
- Johnstone, B., Andrus, J., & Danielson, A. E. (2006). Mobility, indexicality, and the enregisterment of 'Pittsburghese.' *Journal of English Linguistics*, 34(2), 77–104. <https://doi.org/10.1177/0075424206290692>
- Kumle, L., Vo, M. L.-H., & Draschkow, D. (2021). Estimating power in (generalized) linear mixed models: An open introduction and tutorial in R. *Behavior Research Methods*, 53(6), 2528–2543. <https://doi.org/10.3758/s13428-021-01546-0>
- Li, T., Ke, X., & Li, J. (2023). Public attitudes towards dialects: Evidence from 31 Chinese provinces. *PLOS ONE*, 18(10), Article e0292852. <https://doi.org/10.1371/journal.pone.0292852>
- Liu, K. (2021). Language choices as audience design strategies in Chinese multilingual speakers' WeChat posts. *Global Media and China*, 6(4), 391–415. <https://doi.org/10.1177/20594364211035201>
- McCarthy-Jones, S., & Fernyhough, C. (2011). The varieties of inner speech: Links between quality of inner speech and psychopathological variables in a sample of young adults. *Consciousness and Cognition*, 20(4), 1586–1593. <https://doi.org/10.1016/j.concog.2011.08.005>
- Mei, Z. (2024). Dialect inheritance and youth identity: An online ethnography study on dialect enthusiasts community on the new media platforms. *Frontiers in Communication*, 9, Article 1464284. <https://doi.org/10.3389/fcomm.2024.1464284>
- Morin, A., Duhnych, C., & Racy, F. (2018). Self-reported inner speech use in university students. *Applied Cognitive Psychology*, 32(3), 376–382. <https://doi.org/10.1002/acp.3404>
- Ochs, E. (1992). Indexing gender. In A. Duranti & C. Goodwin (Eds.), *Rethinking context: Language as an interactive phenomenon* (pp. 335–358). Cambridge University Press.
- Olson, D. J. (2023). Measuring bilingual language dominance: An examination of the reliability of the Bilingual Language Profile. *Language Testing*, 40(3), 521–547. <https://doi.org/10.1177/02655322221139162>
- Pavlenko, A. (2014). *The bilingual mind: And what it tells us about language and thought*. Cambridge University Press.
- Perrone-Bertolotti, M., Rapin, L., Lachaux, J.-P., Baciú, M., & Loevenbruck, H. (2014). What is that little voice inside my head? Inner speech phenomenology, its role in cognitive performance, and its relation to self-monitoring. *Behavioural Brain Research*, 261, 220–239. <https://doi.org/10.1016/j.bbr.2013.12.034>
- Poplack, S. (1980). Sometimes I'll start a sentence in Spanish y termino en español: Toward a typology of code-switching. *Linguistics*, 18(7–8), 581–618. <https://doi.org/10.1515/ling.1980.18.7-8.581>

- Resnik, P. (2021). Multilinguals' use of L1 and L2 in inner speech. *International Journal of Bilingual Education and Bilingualism*, 24(1), 72–90. <https://doi.org/10.1080/13670050.2018.1445195>
- Sabag, T., Zohar, A. H., Kreiner, H., Lev-Ari, L., & Rabinowitz, D. (2023). The psychometric properties of the Varieties of Inner Speech Questionnaire-Revised in Hebrew. *Frontiers in Psychology*, 13, Article 1092223. <https://doi.org/10.3389/fpsyg.2022.1092223>
- Su, D., & Steiner, P. M. (2020). An evaluation of experimental designs for constructing vignette sets in factorial surveys. *Sociological Methods & Research*, 49(2), 455–497. <https://doi.org/10.1177/0049124117746427>
- Vygotsky, L. S. (1987). Thinking and speech. In R. W. Rieber & A. S. Carton (Eds.), *The collected works of L. S. Vygotsky: Vol. 1. Problems of general psychology* (N. Minick, Trans., pp. 39–285). Plenum Press.
- Yang, L., & Shang, G. (2026). Plurilingual family language policy of internal immigrant families in Suzhou, China: Capital, identity and agency. *International Journal of Multilingualism*, 23(1), 384–403. <https://doi.org/10.1080/14790718.2025.2542481>
- Zeng, W., Shea, C., Beckman, J., & Kutlu, E. (2024). Bidialectal variety switching: The effects of language use and social contexts. *Frontiers in Language Sciences*, 2, Article 1302027. <https://doi.org/10.3389/flang.2023.1302027>
- Zhong, X., Ang, L. H., & Sharmini, S. (2024). Why do Mandarin speakers code-switch? A case study of conversational code-switching in China. *Humanities and Social Sciences Communications*, 11, Article 999. <https://doi.org/10.1057/s41599-024-03534-z>

Short bio

Ran Qiao is affiliated with Southeast University and conducts research on sociolinguistics, multilingualism, inner speech, code-switching, language choice, and language education.