

Article

Lexical Stratification and Semantic Reallocation Under Long-Term Language Contact: Evidence from Northern Shaanxi Jin

Xuemei He 

School of Chinese Language and Literature, Xi'an International Studies University, Xi'an 710128, China;
107242018100029@snnu.edu.cn

Abstract

Research on dialect contact often frames lexical shifts under pressure from dominant standard or prestige varieties as unidirectional replacement: supra-local forms spread and inherited local vocabulary recedes. This article re-examines that model through Northern Shaanxi Jin, a group of Jin varieties in long-term vertical contact with Standard Mandarin. The analysis draws on structured wordlists, contextual elicitation, an 810 min corpus of spontaneous speech, and a supplementary apparent-time response task with 38 younger speakers. Four domains are examined in detail: weather expressions, the hou (猴) diminutive network, the leng (冷)/liang (凉) temperature contrast, and the generalized classifier ge (个). The corpus and elicited data show that Mandarin-aligned forms have expanded, especially in reportive, institutional, cross-dialectal, and younger-peer settings. These forms, however, do not displace inherited lexical resources along a single trajectory. Local forms remain active in fixed weather and customary expressions, tactile distinctions, address and idiomatic uses, and routine nominal quantification. This pattern is analyzed as lexical stratification supported by semantic reallocation: sustained contact redistributes overlapping forms across semantic ranges, constructional environments, and interactional settings rather than producing uniform lexical replacement.

Keywords: lexical stratification; semantic reallocation; language contact; Jin Chinese; Northern Shaanxi

1. Introduction

In Kuanzhou (宽州) Township, Qingjian County, an autumn-harvest interview provided a clear example of context-sensitive lexical choice. Asked whether recent rain allowed villagers to pick prickly ash, a consultant replied: “No—wait until tian (天) opens. When tian is good, the prickly ash turns red.” Here tian refers to weather; tian-hao (天好) refers to fine weather, and tian-kai (天开) refers to the clearing of the weather. The local pattern tian + descriptive element is the ordinary way of speaking about weather in this setting. When asked about weather forecasts, however, the same consultant shifted toward a more Mandarin-aligned formulation: “The forecast says qing-tian (晴天) ‘sunny’, yin-tian (阴天) ‘cloudy’, yu-tian (雨天) ‘rainy’... But for farming, we always look at tian (天).” The contrast is not a simple alternation between an old word and a new word. It reflects different ways of expressing the same concept in different interactional contexts, and it reveals different lexical networks.



Academic Editor: Umberto Ansaldi

Received: 20 January 2026

Revised: 10 July 2026

Accepted: 15 July 2026

Published: 30 July 2026

Copyright: © 2026 by the author.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

The standard replacement account would predict gradual convergence toward Standard Mandarin under schooling, media exposure, and mobility (Aikhenvald, 2007; Backus, 2004; Matras, 2009; Mufwene, 2001; Thomason & Kaufman, 1988). Northern Shaanxi Jin appears to be a likely setting for such convergence. The field data, however, show that convergence is uneven: in daily domains such as weather judgment, bodily sensation, quantification, and affective stance, inherited lexical items often remain the unmarked choices in local interaction.

A central difficulty in Sinitic lexical contact is that graphic identity can conceal functional divergence. The same Chinese character may invite speakers and researchers to assume lexical equivalence where the local system gives the form a different function. For example, *tian* (天) may be glossed in Standard Mandarin as “sky/day”, but in Northern Shaanxi Jin weather expressions it often functions as “weather”. The taste word *tian* (甜) not only refers to sweetness but may also refer to “not salty/plain” or “without additives”; in the local taste field it occupies part of the semantic space covered in Standard Mandarin by both *tian* (甜) and *dan* (淡). *Mi* (米) is a general written form for rice or grain in Standard Mandarin, but in Northern Shaanxi Jin it can refer specifically to millet (小米), preserving a narrower local sense. *Shan* (山), a noun in the common language, may also function locally as an adjective meaning “uncultivated, ill-mannered, or lacking proper conduct”. These examples show why dialectal and Standard Mandarin forms written with the same graph should not be treated automatically as the same functional unit.

Xing (2007) argues that semantic differences are among the most complex problems in Chinese dialect vocabulary, and identifies several recurrent types: the same written form may differ in the number of senses; it may have different non-overlapping meanings; its meanings may partially overlap across dialects; form and meaning may be cross-cutting rather than one-to-one; and a graphically identical form may occupy different phonological layers, lexical strata, or positions within the local semantic field. This point is directly relevant here. Literacy and Standard Mandarin education can encourage speakers and researchers to perceive *tian* (天), *leng* (冷), *hou* (猴), and *ge* (个) as “the same words” as their Standard Mandarin counterparts, especially when elicitation proceeds through written prompts or translation equivalents. The field data show why this inference must be tested rather than assumed. The same graph may correspond to different readings, semantic ranges, distributions, and productive functions. Shared written forms are therefore used here as points of comparison, not as criteria for classification. A form is assigned to a local inherited layer or to a Mandarin-aligned layer only after considering phonological form, semantic range, constructional behavior, collocational network, and interactional frame together.

The field materials therefore require a distributional account. Rather than asking only whether a local form survives or disappears, the analysis asks where each form is used, with whom, and in what kind of discourse. The resulting distribution is treated here as lexical stratification supported by semantic reallocation, with three diagnostics: semantic scope in unmarked use, constructional participation, and network density.

The paper first defines these diagnostics and the coding procedure used to identify interactional frames. It then focuses on four sets of alternative expressions across naturalistic speech, contextual elicitation, and age groups: *tian/tiānqì* in weather talk, *hou/xiao* in diminutive and stance-related uses, *leng/liang* in tactile and temperature description, and *ge/specialized classifiers* in quantification. The material pair *mu-tou/mu-shi* and the sun-related pair *ri-tour/tai-yang* are treated as secondary layered pairs because they are closely tied to everyday life and customary expression, but they are not used as the main statistical evidence.

2. Analytical Framework: Stratification, Scope, and Reallocation

This article uses scope in two related but distinct senses. Semantic scope concerns whether a form can cover several referents, sensory domains, event types, or pragmatic situations without sounding marked. Constructional scope concerns whether it enters the constructions that speakers actually use, including modifier-noun combinations, predicative uses, reduplications, compounds, and fixed expressions. Both types of scope are tested through contextual elicitation, substitution judgments, acceptability judgments, and corpus distribution.

Semantic reallocation is defined here as the redistribution of partially overlapping forms across meaning, construction, and interactional settings. In a layered pair, an inherited form may continue to be the ordinary local choice, with broad scope and strong combinatory productivity. A Mandarin-aligned form may become natural only in more specific contexts, such as reportive, institutional, cross-dialectal, younger peer, or other marked settings. This usage-based separation is the object of analysis here.

Britain and Trudgill (1999) use reallocation for dialect-contact situations in which competing variants survive by acquiring new social, structural, or functional distributions. In the Northern Shaanxi Jin case, the process takes a different shape. The contact relation is vertical rather than only interdialectal: local Jin varieties coexist with Standard Mandarin as a supra-local norm. The shared Chinese writing system can also create the appearance of lexical identity where functional distribution differs. Finally, monosyllabic morphemes, disyllabic standard forms, and productive local word families create different degrees of network centrality. The present use of semantic reallocation is therefore a vertical, standard-oriented variant of Britain and Trudgill's notion, operating within a dialect/standard constellation in Auer's (2005) sense. The focus on recurrent use, productivity, and entrenched distribution follows a usage-based view of lexical organization (Bybee, 2010).

3. Background: Northern Shaanxi Jin as a Contact Zone

Northern Shaanxi lies in northwestern China and has long been characterized by population movement and layered language contact. The region has experienced different contact types across historical periods. This article focuses only on vertical contact between Northern Shaanxi Jin and Standard Mandarin, in order to examine lexical stratification under contemporary Mandarin-Jin contact.

Northern Shaanxi Jin is relatively conservative and retains several phonological features that help identify inherited lexical strata. These include glottal-stop reflexes of Middle Chinese entering-tone syllables, local aspirated readings after the devoicing of historically voiced obstruents, velar readings for some Jian-series open Grade II items, vernacular non-nasal rhyme readings for some Geng-she open Grade II items, and vernacular [m]-initial readings for some Middle Chinese wei (微) items. Such features help distinguish competing expressions and even different readings written with the same character. For example, the graph ya (鸭) keeps a checked-syllable local reading in ya-zi (鸭子) "duck", mu-ya-zi (母鸭子) "female duck", and ya-dan (鸭蛋) "duck egg", but appears with an open-syllable Mandarin-aligned reading in compounds such as ya-bo (鸭脖) "duck neck", ya-jia (鸭架) "duck frame", and ya-rou (鸭肉) "duck meat". The contrast is phonological rather than graphic: the same graph 鸭 participates in a checked local reading in inherited forms and in an open-syllable Mandarin-aligned reading in newer compounds.

These phonological facts are not included as general background only. They matter because locally entrenched lexical items often preserve vernacular readings, whereas Mandarin-aligned items are more likely to appear with standardized readings or in contexts of schooling, media, administration, or cross-dialectal accommodation. The mechanism is comparable to lexical diffusion in phonology: replacement does not normally

occur as a whole-system switch. Instead, both phonological and lexical forms shift gradually across different interactional settings. The analysis below therefore uses phonological layer, semantic scope, constructional productivity, and interactional frame together when identifying inherited and incoming lexical strata.

4. Materials and Methods

4.1. Survey Sites

This study relies on field-based lexical data gathered from 19 county-level localities in Northern Shaanxi. The 19 sites run from Shenmu and Fugu in the north, within the Yulin corridor, to Yanchuan and Qingjian in the south, within the Yan'an corridor; Qingjian's Kuanzhou Township provides the focal interview example used in the introduction.

4.2. Data Collection

Structured wordlists. Targeted elicitation of core semantic domains was conducted using a 1732-item instrument, refined from an initial set of 1735 prompts after removing duplicates and consolidating near-identical items. The main wordlist elicited inherited lexical material, Mandarin-aligned alternatives, collocations, and speaker judgments about the contexts in which alternative forms were natural. The sampling was designed first to recover the inherited lexical layer and then to test layered contrasts across interactional frames; it was not designed as a balanced sociolinguistic model of education, occupation, or individual style. At least one consultant aged sixty or above was interviewed in each locality to ensure recovery of the inherited layer. The broader elicitation records also include younger speakers from the same localities, but the present article uses younger-speaker data only in the supplementary apparent-time task described below. All consultants cited here were long-term local residents. Younger speakers had not left the locality for long-term residence before university age, and the older local consultants had not lived outside the locality for extended periods. The field records include locality, birth year, education, occupation, and residence history. Educational backgrounds ranged from primary school to doctoral study, and occupations ranged from farming to teaching, local administration, manual work, and rural medicine. These metadata served solely to verify local residence and speech background, not to model social class.

Contextual elicitation. Contextual elicitation was used to avoid treating translation equivalents as functional equivalents. When speakers provided more than one form for a concept, the interviewer asked for concrete usage contexts, typical collocations, and negative evidence. For example, prompts for “cold water” were followed by questions about “cold wind”, “a cool night”, “cold hands”, and “cold to the touch” in order to test whether *leng* (冷) and *liang* (凉) differed by sensory modality, register, or interactional frame. This procedure is comparable to the treatment of literary and vernacular readings in Chinese dialect phonology: alternative forms must be located in concrete usage conditions rather than simply listed as synonyms.

Naturalistic observation. To corroborate frequency and productivity patterns in situ, the study recorded 810 min of spontaneous speech from thirty-six long-term local consultants, including at least one speaker aged sixty or above in each locality. The recordings contain all identifiable tokens reported in the quantitative overview below; the counts are therefore not based on a random sample. Sessions combined family interaction with guided topical discussion on weather, agricultural planning, and household matters. The conversations often began with weather or agricultural planning because these topics naturally lead into seasonal work, household arrangements, and local judgments. When speakers shifted from weather to other topics, the recording was not interrupted. This procedure

preserved topic development and allowed register shifts to be observed without forcing repeated target words.

Coding of interactional frames. Everyday interaction was defined as local spoken exchange among dialect-area speakers, including family conversation, local dialect-area talk, agricultural planning, bodily experience, speakers' own weather judgments, affectionate address, teasing, and informal evaluation. Reportive/institutional frames were defined as references to weather forecasts in media or documentary sources, school, government, media, administrative categories, official documents, Mandarin quotation, or speech directed to non-local interlocutors. Ambiguous cases, such as a dialect speaker quoting an official document while speaking to another dialect speaker, were coded according to the form that speakers identified as the most natural choice in local conversation; if the surrounding context still did not allow a stable decision, the token was excluded from the 2×2 statistical test and retained only for qualitative discussion. Coding was carried out by the author using these criteria and checked against the speaker's own contextual judgments and the surrounding discourse, rather than being assigned from the English gloss alone. Because the analysis concerns field-based lexical diagnostics rather than independent annotator classification, coding reliability was strengthened through repeated contextual questioning, substitution judgments, and comparison with the surrounding discourse.

Apparent-time supplement. To address the possibility that the observed stratification reflects only the speech of older consultants, a supplementary apparent-time response task was conducted with 38 speakers aged 18–30 from the same 19 Northern Shaanxi Jin localities. The supplement included two speakers from each locality: one university student or graduate student, and one young speaker living locally. All were raised in the relevant locality, had not left the locality for long-term residence before university age, and had regular exposure to Standard Mandarin through schooling and media. No clear difference was observed between the two subgroups in the targeted lexical choices. The materials were elicited by adapting the contexts in the older-speaker interviews for the layered pairs *tian/tiānqì*, *hou/xiao*, *leng/liang*, and *ge/specialized classifiers*. Young speakers were asked to restate the same situations in the form that sounded most natural to them. This supplement was not designed as a second naturalistic corpus parallel to the 810 min recordings. It was a contextual response task based on matched prompts adapted from the older-speaker interviews. The resulting response counts show production and acceptance under matched contexts and are not pooled with the spontaneous-speech token frequencies reported in the quantitative overview. The two datasets serve distinct evidential roles: the spontaneous-speech counts report observed token frequencies, whereas the younger-speaker task reports elicited response availability under matched prompts. The two counts are not pooled and are not directly comparable as frequencies. Multiple responses were allowed; when a speaker accepted or produced two alternative forms in the same context, both responses were counted. This counting procedure records the range of forms available to speakers; it is not intended to estimate spontaneous usage probability.

4.3. Analytical Diagnostics

Network behavior is a central diagnostic because it shows whether a form is part of a living local pattern rather than an isolated survival. A form is treated as a structural or lexical anchor only when it supports related expressions that speakers actually use or extend. Substitution tests then ask what changes when the local form is replaced by a Mandarin-aligned form: register, local naturalness, pragmatic force, or, in some cases, truth conditions.

Layer assignment did not depend on a single feature. Historical phonological layering and present-day interactional layering are analytically distinct: the former helps identify older local strata, while the latter shows how those strata are redistributed under contem-

porary contact with Standard Mandarin. Phonological layer, including vernacular readings, checked-tone reflexes, and local segmental reflexes, was used as supporting evidence. In practice, forms were classified only when several pieces of evidence pointed in the same direction: broad semantic range, participation in local constructions or idioms, productive related expressions, everyday local use, and markedness under Mandarin replacement.

5. Lexical Composition and Quantitative Overview

The lexicon of Northern Shaanxi Jin (Xing, 2006) comprises interacting components: inherited items traceable to earlier stages of Chinese; common-language items that have entered through schooling and supra-local contact; dialect-specific vocabulary supported by productive local morphology; and regionally shared northern forms that are neither strictly local nor fully standardized. This componential view builds on local dialectological work on Northern Shaanxi and riverine Jin materials (Xing, 2006; Xing et al., 2012), but the present article operationalizes layering through interactional frame, semantic scope, and constructional scope. This article focuses on Mandarin-Jin vertical contact.

A practical challenge in dialect lexicology is distinguishing inherited items that are merely old from items that remain structurally central. The present analysis draws a line between archaic survivals and structural anchors by emphasizing network behavior. Archaic survivals tend to be isolated, limited in scope, and restricted to special registers. Structural anchors, by contrast, are productive, frequent, and tightly integrated into everyday constructions.

Table 1 summarizes the distribution of selected inherited and Mandarin-aligned lexical pairs across interactional contexts in the 810 min corpus. The counts are token frequencies from all identifiable occurrences of the selected lexical pairs in the corpus, rather than a random sample. Fisher's exact test was used for the 2×2 comparison because some cells contain small counts. The result is consistent with a strong association between lexical layer and interactional frame ($p < 0.001$; Cramer's $V = 0.78$). This quantitative pattern should be read together with contextual elicitation and speaker judgments: the low token counts of some Mandarin-aligned forms do not by themselves prove a stable niche, but they are consistent with the qualitative finding that these forms cluster in reportive, institutional, or cross-dialectal contexts.

The younger-speaker supplement shows redistribution by interlocutor and setting. In local peer interaction, Mandarin-aligned or Mandarin-compatible forms such as xiao (小), liang (凉), and tiānqì (天气) have expanded beyond strictly institutional frames. In the response task, this expansion was reflected in expressions such as tiānqì liáng (天气凉) rather than tian-leng (天冷), xiao-feng (小风) rather than hou-feng (猴风), xiao-gezi (小个子) rather than hou-ge-zi (猴个子), and hao tiānqì (好天气) rather than hao-tian (好天). At the same time, inherited forms remain available where they encode tactile distinctions, fixed expressions, local address terms, idioms, or productive morphosyntactic patterns. Leng (冷) remains preferred in tactile contexts such as shou-leng (手冷) "cold hands" and leng-shui (冷水) "cold/unheated water"; tian (天) remains natural in bian-tian (变天) "the weather changes" and tian-kai (天开) "the weather clears"; hou (猴) remains active in address terms and idioms such as hou-da (猴大) "youngest uncle", hou-r (猴儿) "youngest son", and gege laoren qin hou-r (个个老人亲猴儿) "every parent is especially fond of the youngest child"; and ge (个) remains a generalized classifier in everyday speech, as in yi ge mao (一个猫) "one cat" and yi ge niu (一个牛) "one cow".

Table 1. Token frequencies of selected layered pairs across interactional contexts in the 810 min corpus.

Layered Pair	Form	Tokens (N)	Everyday Interaction	Reportive/ Institutional	Cross-Dialect (Quoted/ Adopted)	Notes
leng/liang	leng (冷) [lən ³¹²]	135	✓ e.g., leng-bu-leng (冷不冷) “cold or not”; tian-leng (天冷) “cold weather”; shou-leng (手冷) “cold hands”; leng-shui (冷水) “cold/unheated water”; mo-zhe-leng (摸着冷) “cold to the touch”	✗	✗ (rare with non-Jin interlocutors)	Default for tactile/bodily coldness; also extends to weather and environmental contexts.
leng/liang	liang (凉) [liɑŋ ²⁴]	9	✗	✓	✓ (preferred with non-Jin interlocutors)	Mandarin-aligned in the 810 min corpus; appears in fixed collocations such as liang-mian (凉面) and liang-fen (凉粉). Younger use widens (Table 2).
tian/tiānqì	tian (天) [ti̯e ²¹³]	216	✓ e.g., tian-hao (天好) “good weather”; tian-bu-hao (天不好) “bad weather”; bian-tian (变天) “weather changes”; hao-tian (好天) “fine weather”	✗	✗ (rare with non-Jin interlocutors)	Core everyday weather item; supports a productive tian (天) family, e.g., tian-kai (天开).
tian/tiānqì	tiānqì (天气) [tʰi̯e ²¹ tɕʰi̯ ⁵¹]	13	✗	✓	✓ (used for cross-dialect alignment)	Institutional/official or topic-noun form; matches Standard Mandarin tiānqì (天气).
hou/xiao	hou (猴) [xəu ²⁴]	387	✓ e.g., hou-feng (猴风) “gentle breeze”; hou-ren (猴人) “kiddo”; hou-ge-zi (猴个子) “small fellow”; hou-wa-wa (猴娃娃) “little child”	✗	✗ (restricted to Jin-internal use)	Retained in address terms, idioms, and forms without direct Standard Mandarin equivalents, such as hou-da (猴大), hou-r (猴儿), and hou-hou-r (猴猴儿).
hou/xiao	xiao (小) [ɕiau ³¹²]	9	✗	✓	✓ (neutral in cross-dialect talk)	Mandarin-aligned neutral smallness marker in the 810 min corpus; wider in younger speech (Table 2).

Note. The counts represent all identifiable occurrences of the selected lexical pairs in the 810 min spontaneous-speech corpus, not a random sample. Cross-dialect (quoted/adopted) refers to tokens produced when accommodating non-Jin interlocutors and/or adopting Mandarin-aligned forms in interaction. Statistical tests were computed on a 2×2 table (everyday vs. reportive/institutional $\times$ inherited vs. incoming), with cross-dialect tokens excluded. Ambiguous cases were excluded from the statistical test and retained only for qualitative discussion. The letters in the phonetic transcriptions are International Phonetic Alphabet (IPA) symbols, and the superscript numerals indicate tone values. In Table 1, ✓ indicates that the form is used in the relevant frame, whereas ✗ indicates that it is not used.

The symbols used in Table 2 are: ✓ = regularly used; △ = accepted or context-dependent; ✕ = rarely used or rejected in the relevant frame. In Table 2, “older local” refers to prompts in which younger speakers were asked how they would speak to older local dialect-area interlocutors.

Table 2. Apparent-time supplement: elicited response counts for selected layered pairs among younger speakers aged 18–30.

Layered Pair	Form	Response Counts N	Frame Pattern	Interpretation
leng/liang	leng (冷) [lən ³¹²]	135	Older local ✓; peer △; reportive ✕; cross-dialect ✕	Used with older local interlocutors and tactile/bodily contexts; indicates a stronger degree of coldness.
leng/liang	liang (凉) [liɑŋ ²⁴]	96	Older local △; peer ✓; reportive ✓; cross-dialect ✓	Expanded among younger speakers and no longer limited to fixed collocations; remains distinct from leng in leng-shui (冷水) vs. liang-shui (凉水).
tian/tiānqì	tian (天) [tʰiɛ ²¹³]	209	Older local ✓; peer △; reportive ✕; cross-dialect ✕	Less frequent in younger peer talk, but retained in fixed local expressions such as bian-tian (变天) and tian-kai (天开).
tian/tiānqì	tiānqì (天气) [tʰiɛ ²¹ tɕʰi ⁵¹]	54	Older local △; peer ✓; reportive ✓; cross-dialect ✓	Used more often as a topic noun in younger peer interaction and in forecast, school, media, or cross-dialectal frames.
hou/xiao	hou (猴) [xəu ²⁴]	321	Older local ✓; peer △; reportive ✕; cross-dialect ✕	Retained in address terms, idioms, and forms without direct Standard Mandarin equivalents, such as hou-da (猴大), hou-r (猴儿), and hou-hou-r (猴猴儿).
hou/xiao	xiao (小) [ɕiau ³¹²]	289	Older local △; peer ✓; reportive ✓; cross-dialect ✓	Expanded among younger speakers for neutral smallness, especially in peer and Mandarin-aligned contexts.
ge/specialized classifiers	ge (个) [kuə ²³]	489	Older local ✓; peer ✓; reportive ✓; cross-dialect ✓	The generalized classifier remains available across all frames and has the widest distribution in ordinary speech; specialized classifiers become more acceptable in school-based, careful, Mandarin-oriented, or cross-dialectal speech.
ge/specialized classifiers	specialized classifiers (条/只/头, etc.)	15	Older local △; peer △; reportive ✓; cross-dialect ✓	Rare or context-dependent with older local interlocutors; used occasionally, especially in Mandarin-oriented, school-based, careful, or cross-dialectal speech.

Note. N values in Table 2 are not comparable to Table 1, which reports spontaneous-speech tokens from the 810 min corpus; Table 2 reports elicited response counts from younger speakers under matched prompts. Counts in Table 2 refer to elicited response tokens across contextual prompts, not to the number of individual speakers. Multiple responses were allowed; when a speaker accepted or produced two alternative forms in the same context, both forms were counted. Symbols indicate usage tendency: ✓ = regularly used; △ = accepted or context-dependent; ✕ = rarely used or rejected in the relevant frame. These counts indicate response availability under matched prompts, not corpus frequency or probability of use in spontaneous speech. The letters in the phonetic transcriptions are International Phonetic Alphabet (IPA) symbols, and the superscript numerals indicate tone values.

6. The Tian (天) Weather Network

Weather vocabulary provides one of the clearest cases of lexical centrality. In Northern Shaanxi Jin, the monosyllabic base tian (天) [tʰiɛ²¹³] is not simply a label for weather; it anchors everyday agricultural planning, causal explanation, and local evaluation of seasonal conditions.

The alternation between tian (天) and Mandarin tiānqì (天气) is structurally significant. The tian network supports productive local expressions such as tian-zao (天旱) “weather-caused crop loss”, tian-hong (天红) “dry air”, tian-shai (天晒) “scorching sun”, and tian-kai

(天开) “clearing up”. These forms encode local causal judgments and are used in everyday planning. By contrast, *tiānqì* (天气) is concentrated in forecast, school, administrative, or cross-dialectal talk. This distribution is consistent with treating *tian* as a local lexical anchor, while the Mandarin-aligned form occupies a narrower reportive frame.

In weather-change contexts, speakers accepted *bian-tian* (变天) but rejected or marked *tiānqì bian-le* (天气变了) as a Mandarin-style paraphrase rather than a natural local expression. Likewise, *tian-kai* (天开) is the expected form for weather clearing, whereas *tiānqì kai-le* (天气开了) is unnatural. These contrasts show that *tian* is not merely a shorter form of *tiānqì*: it participates in a local set of eventive and evaluative expressions.

Among younger speakers, *tiānqì* has expanded as a topic noun in peer interaction and in reportive frames, especially in expressions such as *tiānqì liáng* (天气凉) and *hao tiānqì* (好天气). Yet *tian* remains the preferred form in fixed local expressions such as *bian-tian* (变天) “weather changes” and *tian-kai* (天开) “the weather clears”. The younger data thus separate two functions: *tiānqì* is becoming a common topic noun for “weather” among some young speakers, while *tian* remains embedded in local eventive expressions.

7. The Hou (猴) Diminutive and Stance Network

The *hou* (猴) network illustrates how a form can remain structurally important even when a Mandarin-aligned alternative expands among younger speakers. In older-speaker data, *hou*-based forms occur in address, teasing, diminutive description, and idiomatic expressions. Examples include *hou-wa-wa* (猴娃娃) “small child”, *hou-ge-zi* (猴个子) “small-sized person”, *hou-da* (猴大) “youngest uncle”, *hou-r* (猴儿) “youngest son”, *hou-hou-r* (猴猴儿) “the youngest child or animal offspring”, and *hou-feng* (猴风) “gentle breeze”. Speaker clarification indicates that *hou-feng* refers locally to a weak or gentle breeze, so the gloss “gentle breeze” is used here.

The contrast with *xiao* (小) is not a simple referential contrast. *Xiao* is increasingly available for neutral smallness, especially in younger peer interaction and Mandarin-aligned contexts, as in *xiao-wa-wa* (小娃娃) “small child”, *xiao-ge-zi* (小个子) “small person”, or *xiao-feng* (小风) “light wind”. *Hou*, by contrast, remains tied to local address, idioms, teasing, and expressions for which speakers do not give a straightforward Standard Mandarin replacement. The proverb *gege laoren qin hou-r* (个个老人亲猴儿) “every parent is especially fond of the youngest child” is one such idiomatic case.

The younger-speaker materials show that *xiao* is used more widely for neutral smallness, especially in peer and Mandarin-aligned contexts, as in *xiao-feng* (小风), *xiao-ge-zi* (小个子), and *xiao-wa-wa* (小娃娃). *Hou* remains active in local address terms, idioms, and expressions without direct Standard Mandarin equivalents, such as *hou-da* (猴大), *hou-r* (猴儿), *hou-hou-r* (猴猴儿), and *gege laoren qin hou-r* (个个老人亲猴儿). The shift is therefore strongest where the meaning is only neutral smallness; it is weaker where *hou* carries address, teasing, or idiomatic value.

8. The Leng (冷)/Liang (凉) Temperature Contrast

The distinction between *leng* (冷) [lən³¹²] and *liang* (凉) [liɑŋ²⁴] illustrates scope asymmetry in a tactile domain. Speakers use *leng* as the unmarked form for tactile and bodily coldness, including cold hands, cold water, and cold objects. *Liang* appears more readily in environmental descriptions, fixed expressions, careful speech, or Mandarin-aligned contexts.

In Wubu County, for example, a speaker distinguished drinking *leng-shui* (冷水) “cold water” from sitting in a *liang-sou-sou* (凉飕飕) “cool” shade. Although Standard Mandarin may allow *liang-shui* (凉水) for cool water, local speakers maintained *leng* as the default form for direct sensory contact. The analysis here concerns usage distribution, collocation, and speaker judgment; it does not make a claim about psycholinguistic processing.

Among younger speakers, *liang* has expanded in peer interaction. The contrast between *leng-shui* (冷水) and *liang-shui* (凉水) nevertheless remains semantically relevant: *liang-shui* tends to refer to boiled water that has cooled down, whereas *leng-shui* refers more broadly to unheated or cold water. Tactile contexts such as *shou-leng* (手冷) “cold hands”, *shui-leng* (水冷) “the water is cold”, and *mo-zhe-leng* (摸着冷) “cold to the touch” also continue to favor *leng*. What younger speakers are losing is not the word *leng*, but part of its environmental range.

9. The Generalized Classifier *Ge* (个)

Quantification provides another domain in which local organization remains visible. Northern Shaanxi Jin shows broad use of a generalized classifier *ge* (个) [kuəʔ³] across humans, animals, objects, body parts, and events, whereas Standard Mandarin often distributes these contexts across more specialized classifiers.

In everyday speech, speakers use *ge* broadly in expressions such as *yi ge mao* (一个猫) “one cat”, *yi ge niu* (一个牛) “one cow”, *yi ge ren* (一个人) “one person”, *yi ge che* (一个车) “one car”, and *yi ge shi* (一个事) “one matter”. This broad constructional scope does not eliminate reference distinctions; contextual information supplies the nominal interpretation. Specialized Mandarin classifiers are available to speakers and may appear in careful speech, school-based expression, or cross-dialectal communication, but they do not displace the generalized classifier in ordinary local talk in the data examined here.

The classifier data show little generational differentiation. The generalized classifier *ge* remains robust among younger speakers, with 489 elicited responses compared with 15 for specialized classifiers. Young speakers still use *ge* across animals, humans, and objects in everyday speech, including *yi ge mao* (一个猫) “one cat” and *yi ge niu* (一个牛) “one cow”. Specialized classifiers are available but appear mainly in school-based, careful, Mandarin-oriented, or cross-dialectal speech. The robustness of *ge* should not be interpreted as evidence of weak contact pressure. It sits at the intersection of a pan-Sinitic preference for generalized classifiers and high-frequency morphosyntactic routinization. In the present Northern Shaanxi Jin data, this combination makes the classifier domain less permeable to Mandarin specialized forms than lexical domains such as neutral diminutives or topic nouns; field observations from the Zhongyuan Mandarin transitional belt to the south suggest that specialized classifiers are more readily accepted there in careful and cross-dialectal speech.

10. Secondary Layered Pairs: Material and Sun Terms

The pairs *mu-shi* (木植)/*mu-tou* (木头) and *ri-tou* (日头儿)/*tai-yang* (太阳) are not included in Table 1 or Table 2, so they should not be treated as main statistical evidence. They are nevertheless useful for illustrating how shared written forms and age-sensitive distribution can mislead a purely form-based account of lexical convergence.

10.1. *Mu-Shi* (木植) and *Mu-Tou* (木头)

In Northern Shaanxi Jin, *mu-tou* (木头) does not have the same meaning as Standard Mandarin *mùtōu* “wood/timber”. The second syllable *tou* bears a full tone rather than a neutral tone, and the local form refers to a coffin. The Standard Mandarin meaning “wood, timber” is expressed by *mu-shi* (木植) in local speech. This contrast is not simply a difference between a dialect word and a standard word written differently; it is a case where the graphically familiar form *mu-tou* has undergone a different semantic specialization in the local system.

The local meaning can be explained historically. In Northern Shaanxi Jin, *mu* (木) retains an older sense connected with coffin or coffin timber, while *tou* functions as a nominal

suffix. Classical examples show that mu could be associated with coffin or coffin materials, as in expressions related to jiu-mu (就木) “to approach the coffin, to die”. By contrast, in Standard Mandarin, mùtōu developed as a disyllabic noun for “wood/timber”; the coffin-related meaning of mu is no longer the common semantic base of the word. The Northern Shaanxi Jin form mu-tou therefore preserves a different semantic layer from Standard Mandarin mùtōu.

Mu-shi (木植), by contrast, is the ordinary local form for “wood/timber”. The form mu-shi for “wood/timber” is already noted in riverine Jin materials (Xing et al., 2012); the present point is not first attestation, but the way the graphically familiar mu-tou is split into non-interchangeable functions across the local and Standard Mandarin systems. The form is historically and phonologically informative. According to Xing et al. (2012), the character 植 derives from a Middle Chinese voiced obstruent category and its fricative reflex in the riverine Jin materials follows regular historical correspondences. In localities with literary and vernacular readings, such as Qingjian, Yanchuan, and Jia County, the [ʂ]-initial form belongs to the vernacular layer, while a more Mandarin-aligned affricate reading belongs to the literary layer. This supports the article’s broader claim that inherited lexical organization must be identified by combining semantics, collocation, and phonological layer.

In use, both older and younger speakers report mu-shi for “wood/timber” and mu-tou for “coffin”. Fixed expressions further support the distinction: shanshang mu-shi chou, yaoli bu chou chuan (山上木植稠，窑里不愁椽) “when timber is dense on the mountain, one need not worry about rafters for the cave dwelling” uses mu-shi for timber, while mu-tou chuan zhao (木头穿罩) refers to the coffin and burial clothing after death. Some points, including Yulin, Jia County, and Mizhi, also show layering between mu-shi and an incoming or Mandarin-aligned mu-tou in the “wood/timber” sense. Because the distribution of this overlap is not quantified here, the pair is treated as secondary evidence rather than a core statistical case.

10.2. Ri-Tour (日头儿) and Tai-Yang (太阳)

The pair ri-tour (日头儿) and tai-yang (太阳) shows a different type of age-sensitive redistribution. Among speakers aged sixty and above, ri-tour remains common for “sun” in ordinary speech and is especially stable in proverbs, agricultural sayings, xiehouyu, and folk songs. The form is morphologically layered: ri-tou (日头) is an older compound in which tou functions as a nominal suffix, and Northern Shaanxi Jin further adds the suffix -r (儿). Historical texts from the Sui-Tang and Five Dynasties period already contain ri-tou in the meaning “sun”, as in the line wu shi an nei zuo, shi jue ri-tou tun 午时庵内坐，始觉日头暎 (“Sitting in the hermitage at noon, I first feel the sun’s warmth”) from Hanshan 寒山 (Liu & Liu, 1992). Related ri-tour forms are also reported in riverine Jin materials (Xing et al., 2012); the contribution here is the age- and genre-sensitive distribution of the form in Northern Shaanxi Jin.

The local distribution is genre-sensitive. In older speakers’ everyday speech, both ri-tour and tai-yang can be used in some referential contexts: speakers may say ri-tour luoshan le (日头儿落山了) “the sun has set”, ri-tour shanglai le (日头儿上来了) “the sun has risen”, jinr mei tai-yang (今儿没太阳) “there is no sun today”, or jinr mei ri-tour (今儿没日头儿) “there is no sun today”. In culturally marked or formulaic contexts, however, ri-tour is preferred. For example, in the Spring Festival custom of not carrying water before sunrise, the relevant expression is ri-tour mei shanglai le, buneng danshui (日头儿没上来了，不能担水) “the sun has not risen; water cannot be carried.” Proverbs and sayings also preserve ri-tour, such as yunli de ri-tour, dongli de feng, xiezi de weiba, houmuqin de xin (云里的日头，洞里的风，蝎子的尾巴，后母亲的心) “the sun behind clouds, the wind inside a cave, a scorpion’s tail, and a stepmother’s heart”, which

names four harmful or dangerous things. The xiehouyu liuyue de ri-tour—du shi le (六月的日头儿——毒实了) “the June sun—extremely harsh/poisonous” similarly uses ri-tour to express scorching heat and, figuratively, harshness.

Among younger speakers aged 18–30, tai-yang has largely replaced ri-tour in ordinary referential use. Younger speakers still recognize or reproduce ri-tour in proverbs, sayings, xiehouyu, and folk-song lines, but in ordinary conversation they tend to use tai-yang. This differs from the tian/tiānqì pattern: ri-tour shows stronger generational retreat and genre-bound retention, whereas tian remains active in fixed weather expressions among younger speakers. The pair therefore illustrates why semantic reallocation and replacement must be distinguished by domain and age group; different domains show different rates of replacement across age groups. The comparison is summarized in Table 3.

Table 3. Secondary layered pairs added from supplementary field notes.

Pair	Local/ Inherited Form	Mandarin- Aligned Form	Semantic Scope of Local Form	Constructional/ Collocational Range	Typical Local Examples	Younger-Speaker Response	Replacement Judgment
mu-shi/mu-tou	mu-shi (木植) for “wood/timber”; local mu-tou (木头) for “coffin”	Standard Mandarin mùtóu (木头) “wood/timber”	mu-shi denotes wood/timber; local mu-tou denotes coffin/coffin materials.	mu-shi occurs in timber/cave- dwelling building expressions; local mu-tou occurs in death/burial contexts.	shanshang mu-shi chou, yaoli bu chou chuan (山上木植稠, 窑里不愁椽); mu-tou chuan zhao (木头穿罩)	Both older and younger speakers report mu-shi for “wood/timber” and mu-tou for “coffin”; some points also show overlap with Mandarin-aligned mu-tou for “wood”.	Not inter- changeable in core local usage; replacing mu-shi with Mandarin mùtóu may be ambiguous or Mandarin- aligned.
ri-tour/tai-yang	ri-tour (日头儿)	tai-yang (太阳)	ri-tour denotes the sun and is strongly tied to local formulaic, customary, and oral- traditional contexts.	Common in proverbs, agricultural sayings, xiehouyu, folk songs, and older speakers’ ordinary speech.	ri-tour luoshan le (日头儿落山了); ri-tour mei shanglai le, buneng danshui (日头儿没上来了, 不能担水); liuyue de ri-tour—du shi le (六月的日头儿— 毒实了)	Young speakers mostly use tai-yang in ordinary conversation, but retain ri-tour in proverbs, songs, sayings, and fixed expressions.	Tai-yang replaces ri-tour in many young-speaker ordinary contexts; ri-tour remains genre-bound and culturally embedded.

Note. Table 3 is based on supplementary field notes rather than the 810 min corpus counts. It is included to show why secondary pairs should be analyzed through semantic scope, phonological layer, and genre or age distribution, not through graphic form alone.

11. Discussion

The cases differ in the degree of generational pressure. The generalized classifier *ge* is the most stable, since it functions as a high-frequency morphosyntactic routine in ordinary speech. The *tian* network remains stable in fixed eventive expressions such as *bian-tian* (变天) and *tian-kai* (天开), even though *tiānqì* (天气) has expanded as a topic noun. *Hou*-based forms show stronger pressure from *xiao* (小) in neutral diminutive contexts, but remain active in address terms, idioms, and teasing. *Ri-tour* (日头儿) represents a stronger retreat: it survives mainly in proverbs, songs, *xiehouyu*, and customary expressions, while *tai-yang* (太阳) has become the ordinary young-speaker form. These differences suggest that lexical reallocation depends on frequency, constructional embedding, cultural load, and the availability of a neutral Mandarin-aligned alternative.

Semantic reallocation is useful for describing the sorting of forms across use conditions. Standard Mandarin forms appear most often in reportive, institutional, public-facing, cross-dialectal, or younger peer contexts. Inherited forms retain stronger local functions where they are part of fixed expressions, tactile contrasts, idiomatic address, or productive morphosyntactic patterns. The apparent-time and secondary-pair evidence further

shows that not all inherited forms move along the same path: some remain general, some are redistributed by setting, and some retreat into fixed genres.

Table 2 shows that younger speakers do not simply reproduce the older pattern. They use *xiao*, *liang*, and *tiānqì* more freely in peer interaction, while retaining *tian*, *hou*, *leng*, and *ge* in the contexts specified above. The apparent-time pattern is uneven across lexical domains rather than uniformly conservative or uniformly replacing.

Section 10 points to the same problem from another angle. *Mu-shi/mu-tou* shows that shared characters may hide different semantic layers: local *mu-tou* “coffin” is not the same functional item as Standard Mandarin *mùtōu* “wood/timber”. *Ri-tour/tai-yang* shows that a local inherited form may retreat from ordinary young-speaker usage while remaining active in fixed genres and cultural expressions. These cases support the methodological point that dialect vocabulary should be documented through meaning, phonological layer, collocation, genre, and age distribution together.

Methodologically, these findings imply that dialect lexicography should move beyond documenting only dialect-only words. Because functional divergence can hide behind shared characters, future documentation should record phonological layer, semantic scope, constructional scope, collocational networks, and interactional frame. This approach is better suited to describing inherited lexical organization under contact than simple translation-equivalent lists. A consolidated list of the illustrative lexical sets discussed in the article is provided in Appendix A.

12. Conclusions

Northern Shaanxi Jin does not show a single outcome of contact-induced lexical change. In the domains examined here, Mandarin-aligned forms have expanded, especially in school, media, cross-dialectal communication, and younger peer interaction. At the same time, inherited forms remain active where they are embedded in local expressions, tactile distinctions, address terms, idioms, or generalized morphosyntactic patterns.

The article does not argue that all inherited vocabulary is stable. It shows that replacement pressure is filtered through semantic scope, constructional scope, network productivity, and interactional context. Semantic reallocation is the term used here for this distributional outcome: some local forms remain central, some retreat into fixed genres, and some Mandarin-aligned forms become ordinary in younger or outward-facing use.

The younger-speaker supplement supports this domain-specific account. Young speakers have incorporated more Mandarin-aligned forms in peer interaction, but they still preserve local distinctions in tactile experience, fixed weather expressions, idiomatic address, and generalized classifier use. Material and sun terms also show that graphic identity may conceal divergent local meanings or genre-bound survivals. Further work should test the same diagnostics in motion, evaluation, address, and expressive vocabulary, and should compare subareas within Northern Shaanxi Jin with denser token counts and broader demographic sampling.

Funding: This research was funded by the Key Project of the 2026 Education and Teaching Reform Program (Chinese Language and Culture Special Program) of the Teaching Steering Committee for Education Majors in Vocational Colleges, Ministry of Education of China, grant number 2026JGZD004.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data supporting the findings of this study were obtained through the author’s fieldwork and are not publicly available because the recordings and elicitation materials contain information that may compromise participant confidentiality. The data may be made available by the corresponding author upon reasonable request, subject to privacy and ethical restrictions.

Conflicts of Interest: The author declares no conflict of interest.

Appendix A. Illustrative Lexical Sets

- A. Weather family (base tian): tian (天) “weather/day”; hao-tian (好天) “good weather”; tian-bu-hao (天不好) “bad weather”; tian-dao (天道) “climate/seasonal conditions”; tian-zao (天早) “weather-caused crop loss”; tian-hong (天红) “dry air”; tian-shai (天晒) “scorching sun”; tian-wu (天捂) “muggy weather”; tian-kai (天开) “clearing up”. These are used in everyday planning and causal explanation.
- B. Diminutive/stance family base hou (猴): hou-wa-wa (猴娃娃) “small child”; hou-nü-nü (猴女女) “little girl”; hou-ge-zi (猴个子) “small-sized person”; hou-dan-dan (猴蛋蛋) “small/insignificant person or thing” (often evaluative); hou-da (猴大) “youngest uncle”; hou-r (猴儿) “youngest son”; hou-hou-r (猴猴儿) “the youngest child or animal offspring”; hou-nao (猴脑) “small head”; hou-feng (猴风) “gentle breeze”. These forms encode size as well as local address, stance, and idiomatic meanings.
- C. Secondary layered pairs: mu-shi (木植) vs. mu-tou (木头); ri-tour (日头儿) vs. tai-yang (太阳). These pairs are not used as primary statistical evidence, but they illustrate two further outcomes: graphic identity may mask semantic divergence, and inherited forms may retreat into fixed genres among younger speakers.
- D. Temperature distinction: leng (冷) “cold to the touch” vs. liang (凉) “cool, often environmental or Mandarin-aligned”. Leng-shui (冷水) remains the unmarked form for unheated or cold water; liang-shui (凉水) tends to refer to boiled water that has cooled down among younger speakers.
- E. General classifier: ge (个) functions as a generalized classifier across humans, animals, objects, body parts, and events in everyday speech; specialized classifiers are available but are less frequent and more context-bound in the data examined here.

References

- Aikhenvald, A. Y. (2007). Grammars in contact: A cross-linguistic perspective. *Journal of Linguistics*, 43, 1–66. [\[CrossRef\]](#)
- Auer, P. (2005). Europe’s sociolinguistic unity, or: A typology of european dialect/standard constellations. In N. Delbecque, J. van der Auwera, & D. Geeraerts (Eds.), *Perspectives on variation: Sociolinguistic, historical, comparative* (pp. 7–42). De Gruyter Mouton. [\[CrossRef\]](#)
- Backus, A. (2004). Convergence as a mechanism of language change. *Bilingualism: Language and Cognition*, 7, 179–181. [\[CrossRef\]](#)
- Britain, D., & Trudgill, P. (1999). Migration, new-dialect formation and sociolinguistic refunctionalisation: Reallocation as an outcome of dialect contact. *Transactions of the Philological Society*, 97, 245–256. [\[CrossRef\]](#)
- Bybee, J. (2010). *Language, usage and cognition*. Cambridge University Press.
- Liu, Y., & Liu, Z. (Eds.). (1992). *Wang Fanzhi Hanshan shi quanji* [Complete collection of Wang Fanzhi and Hanshan poems] (p. 73). Hainan Publishing House. (In Chinese)
- Matras, Y. (2009). *Language contact*. Cambridge University Press.
- Mufwene, S. S. (2001). *The ecology of language evolution*. Cambridge University Press.
- Thomason, S. G., & Kaufman, T. (1988). *Language contact, creolization, and genetic linguistics*. University of California Press.
- Xing, X. (2006). *Shaanbei jinyu yufa bijiao yanjiu* [Comparative study of northern shaanxi jin grammar]. Commercial Press. (In Chinese)
- Xing, X. (2007). Guanyu shenhua hanyu fangyan cihui yanjiu de sikao [Thoughts on deepening the study of Chinese dialect vocabulary]. *Journal of Shaanxi Normal University (Philosophy and Social Sciences Edition)*, 2, 117–122. (In Chinese)
- Xing, X., Wang, L., Zhang, W., & Li, X. (2012). *Qin jin liangsheng yanhe fangyan bijiao yanjiu* [A comparative study of riverbank dialects in shaanxi and shanxi]. The Commercial Press. (In Chinese)

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.