The Historical Development of Korean Indefinites: A Corpus-Based Study

by

Rok Sim

Bachelor of Arts
Kyung Hee University, 2015

Master of Arts
Kyung Hee University, 2017

Submitted in Partial Fulfillment of the Requirements

for the Degree of Doctor of Philosophy in

Linguistics

McCausland College of Arts and Sciences

University of South Carolina

2026

Accepted by:

Stanley Dubinsky, Major Professor

Emily Manetta, Major Professor

Anne Bezuidenhout, Committee Member

Brett Sherman, Committee Member

Jiwon Yun, Committee Member

Lara Lomicka Anderson, Interim Vice Provost and Dean of the Graduate School

Dedication

Although this dissertation bears my name, it was never the product of my efforts alone. It is the result of many encounters, many acts of guidance and encouragement, and many forms of generosity. For all of my 인연 (因緣)s, I am deeply thankful.

Acknowledgments

인연 (因緣) refers to the ties that connect people and to the causes and conditions through which events unfold. It suggests that no outcome is shaped by one person alone. In that spirit, I begin this work by acknowledging the many forms of 인연 that sustained me throughout my dissertation journey.

First and foremost, I offer my deepest gratitude to my advisor, Dr. Stanley Dubinsky. More than anyone else, he shaped my intellectual growth during this journey. He gave generously of his time, reading my writing closely and offering careful feedback and guidance. Among the many lessons I learned from him, the most enduring was the importance of expressing complex ideas with clarity, precision, and care. Without his dedication and guidance, I could not have brought this work to completion.

I am also deeply grateful to Dr. Emily Manetta and Dr. Anne Bezuidenhout, whose encouragement came at a moment when I had begun to doubt the value of my work. Their reassurance restored my confidence and helped me see the project differently. I also thank Dr. Brett Sherman, who reminded me to move thoughtfully between detail and broader perspective, helping me attend to both the finer points and the larger intellectual picture.

I am especially thankful to my outside committee member, Dr. Jiwon Yun. It was a privilege to have her as part of this dissertation. Her insights brought both structure and inspiration to this work, helping me refine its direction and understand it within a broader intellectual frame.

Beyond my committee, I have been fortunate to receive support from many faculty members in the linguistics program. I thank Dr. Elaine Chun for encouraging me to think more boldly and to imagine broader possibilities for my work. I am also grateful to Dr. Amanda Dalola and Dr. Eric Holt for their encouragement and for the opportunity to work with them on other projects, and to Dr. Nina Moreno for coordinating my Korean classes and for her steady support over the years. I am equally thankful to my foreign language lab family—Bill, Henning, and all the lab members—whose warmth and camaraderie gave me a strong sense of belonging.

I would also like to express my deep gratitude to Dr. Jong-Bok Kim, whose guidance during my bachelor' s and master' s years first made this path imaginable to me. His example as an enthusiastic and deeply committed linguist inspired me to pursue this field more seriously. I also thank Jungsoo Kim for patiently listening to my questions and encouraging me through moments of uncertainty, as well as Okgi Kim and Seulkee Park, my first *seonbae*, who introduced me to the world of linguistics and inspired me to dream of studying abroad.

My heartfelt thanks go to Jiyeon, Keunhyung, and Drew, with whom I shared many conversations about linguistic ideas as well as the struggles of graduate life. I also thank Cheng, Dawson, Danielle, Lesley, Ruthanne, Sam, Shana, Kaleigh, Sarah, and Altyn for their friendship and support. Graduate school can be deeply solitary, but they made this journey far less lonely.

Last, and most importantly, I thank my family in Korea. I also thank Hyeseon, whose care and steady presence sustained me throughout this journey, and my four-legged, purring companion, Sophy. No words can fully express what I owe them. Their love, patience, and sacrifices carried me through every stage of this journey, especially in moments of uncertainty, exhaustion, and distance from home. Whatever I have been able to accomplish rests on the foundation they gave me.

Abstract

Korean offers an unusually transparent window into how indefinite paradigms are built and reshaped over time. Two independent bases—a *wh*-indeterminate root (*nwukwu* 'who') and a dedicated indefinite root (*amo*/*amwu*)—combine with a small, reused set of particles (e.g. -(*i*)*nka*, -*to*, -(*i*)*na*) to yield a dense paradigm spanning interrogative, existential/epistemic, polarity-sensitive, and free-choice functions. The system is especially revealing because it exhibits stable asymmetries and gaps (notably **amwu-inka*) and because forms that are morphologically parallel do not freely alternate, but instead show persistent specialization.

This dissertation develops a usage-based, corpus-driven account of how this modern division of labour emerges. I treat change as movement and competition within a structured semantic space, combining Haspelmath-style semantic maps for indefinites with polarity-oriented models of diachrony. Empirically, I track *nwukwu-* and *amo-*/*amwu-* (bare and particle-marked) across three centuries of written Korean using the History corpus (1700–1890) and the Yonsei corpus (1900–1990). Tokens are manually curated, annotated for map-based function, and analyzed using normalized frequencies, (segmented) logistic regression, and entry-threshold diagnostics that identify when a function becomes a stable part of a series.

The results are organized along three dimensions. First, roots: bare *nwukwu-* expands from interrogatives into adjacent indefinite territory, while bare *amwu-* narrows along a polarity cline, retreating from earlier existential and free-choice

uses into predominantly negative-polarity distributions anchored in strong negative environments. Second, grammaticalization: wh+-(*i*)*nka* undergoes a clause-to-DP reanalysis that yields a productive epistemic indefinite series (*nwukwunka*), while the absence of **amwu-inka* follows from the timing of available source configurations and from paradigm-internal niche saturation. Third, competition: in shared particle series, *amwu-* and *nwukwu-* partition the map rather than duplicate each other, with *amwu*-forms concentrating in stronger negative and indiscriminative zones and *nwukwu*-forms occupying more neutral and extensional zones.

In sum, these findings argue that Korean indefinites are best explained at the level of the paradigm: usage-driven reweighting of functions, clause-to-DP grammaticalization, and competition jointly produce the modern system, and they do so in ways that largely respect semantic-map adjacency.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

This dissertation investigates the historical development of Korean indefinites built from two core bases, the *wh*-form *nwukwu* 'who' and the dedicated indefinite root *amwu*. In present-day Korean, these bases combine with a small set of particles (notably -(*i*)*nka*, -*to*, -(*i*)*lato*, and -(*i*)*na*) to yield a tightly structured paradigm of existential, polarity-sensitive, and free-choice expressions. Korean is especially revealing because the same small inventory of roots and particles gives rise to a dense but asymmetric system: some forms are used as interrogatives and plain existentials, others are confined to polarity-sensitive or free-choice environments, and morphologically parallel expressions do not simply alternate.

The central claim of this dissertation is that the modern division of labour in this paradigm is best explained at the level of the paradigm rather than at the level of isolated items. More specifically, I argue that three processes jointly produce the modern system: (i) long-term semantic reorganization of the two bases, (ii) clause-to-DP grammaticalization of particles, and (iii) competition within partially overlapping particle series. More broadly, the Korean case shows how indefinite paradigms can be assembled diachronically through competition over a structured semantic space: source constructions, usage skew, and paradigm-internal differentiation together determine which forms stabilize where.

This chapter lays out the empirical domain, the core asymmetries, and the analytic tools used throughout the dissertation. Section 1.1 introduces Korean *wh*-indefinites and shows how particle choice systematically shapes existential,

polarity-sensitive, and free-choice readings. Section 1.2 turns to the dedicated indefinite root *amwu* and highlights the key synchronic asymmetry that recurs throughout the dissertation, namely the absence of a plain existential **amwu-inka* and the concentration of *amwu*-forms in polarity and free-choice uses. Section 1.3 reviews a clause-to-DP grammaticalization perspective on the particles and motivates a diachronic approach to paradigm structure. Section 1.4 distills three puzzles and states working hypotheses about base reorganization, particle grammaticalization, and competition. Section 1.5 identifies the research gap and formulates the research questions guiding the corpus study. Section 1.6 introduces the theoretical tools used to model change across the paradigm. Section 1.7 previews the corpus methodology, annotation scheme, and quantitative strategy, and summarizes the main findings. Section 1.8 briefly outlines the remainder of the dissertation.

1.1 *Wh*-indefinites

In English, *wh*-words like *who*, *what*, and *where* are used almost exclusively in interrogative contexts. In Korean, by contrast, *wh*-forms such as *nwukwu* 'who' can also be used as indefinites meaning 'someone', depending on the sentence type.[1] Compare:

(1) a. Nwu-ka w-ass-ni?
who-NOM come-PST-QUE
'Who came?'
[Interrogative]

[1] Note that *nwu* is a contracted form of *nwukwu*.

b. Nwu/*amwu-ka w-ass-ta.
who/amwu-NOM come-PST-DECL
'Someone came.'

[Declarative]

In (1a), *nwukwu* behaves like English *who*; in (1b) it is interpreted as 'someone'. The form itself does not encode a fixed interrogative vs. indefinite split, which motivates the term *wh*-indeterminate (Kratzer & Shimoyama, 2002; Kuroda, 1965; Shimoyama, 2001). In this dissertation, I use the term *wh*-indefinite for *wh*-forms when they are used with indefinite readings.

A second defining property of Korean *wh*-indefinites is that particles can attach to them, systematically shaping their interpretation. The items in (2a)–(2d) illustrate four common particles—-(*i*)*nka*, *-to*, -(*i*)*lato*, and -(*i*)*na*—which yield existential, negative-polarity, and free-choice interpretations.[2,3]

(2) a. Nwukwu/*amwu-nka w-ass-ta.
who/amwu-NKA come-PST-DECL
'Someone came.'

[Existential]

b. Nwukwu/amwu-to an w-ass-ta.
who/amwu-TO not come-PST-DECL
'No one came.' (lit. 'anyone did not come')

[NPI, negation]

[2]Here *existential* refers to a plain ∃-quantified indefinite in upward-entailing contexts (Heim, 1982). A *negative polarity item* (NPI) is restricted to licensing environments such as negation and other downward-entailing and/or nonveridical contexts, and is interpreted in the scope of the relevant operator (Giannakidou, 2011; Ladusaw, 1980). A *free-choice item* (FCI) is an indefinite that yields a free-choice inference in modal, imperative, or generic environments (often glossed as free-choice 'any') (Dayal, 1998).

[3]Throughout the dissertation I cite these particles as -(*i*)*nka*, -(*i*)*na*, and -(*i*)*lato*. After a vowel-final stem the initial *i-* is typically absent (e.g. *nwukwu-nka*), while after a consonant-final stem it surfaces (e.g. *amwu-inka*). The particle *-to* does not show this alternation.

c. Nwukwu/amwu-lato w-ass-ni?
who/amwu-LATO come-PST-QUE
'Did anyone come?'

[NPI, question]

d. Nwukwu/amwu-na ha-l.swu.iss-ta.
who/amwu-NA do-can-DECL
'Anyone can do it.'

[Free choice]

In (2a), -(*i*)*nka* yields a plain existential reading. With *-to* and -(*i*)*lato* in (2b)–(2c), the *wh*-indefinite behaves as an NPI in negative and interrogative environments. With -(*i*)*na* in (2d), it yields a free-choice interpretation. This particle sensitivity is the starting point for the asymmetries discussed below: a shared set of particles combines with both bases, yet the resulting paradigms are not symmetric.

1.2 *Amwu*-indefinites

Alongside *wh*-indefinites, Korean also has a dedicated indefinite root *amwu*. Unlike *nwukwu*, *amwu* is never interrogative: it appears only in indefinite expressions. Structurally, *amwu* participates in many of the same particle series as *nwukwu*, but with an important gap in the existential domain:

(3) a. *Amwu-nka w-ass-ta.
amwu-NKA come-PST-DECL
'Someone came.'

[Ungrammatical]

b. Amwu-to an w-ass-ta.
amwu-TO not come-PST-DECL
'No one came.' (lit. 'anyone did not come')
[NPI, negation]

c. Amwu-lato w-ass-ni?
amwu-LATO come-PST-QUE
'Did anyone come?'
[NPI, question]

d. Amwu-na ha-l.swu.iss-ta.
amwu-NA do-can-DECL
'Anyone can do it.'
[Free choice]

Examples (3a)–(3d) make two points. First, *amwu* cannot combine with -(*i*)*nka* to form a plain existential, in contrast to *nwukwu* in (2a). Second, in NPI and free-choice contexts *amwu* combines with the same particles as *nwukwu* and yields closely related truth-conditional readings.

Table 1.1 Functional distribution of *nwukwu* vs. *amwu*.

Function	*nwukwu*	*amwu*
Interrogative	✓	×
Existential (-(*i*)*nka*)	✓	×
NPI (negation; -*to*)	✓	✓
NPI (question; -(*i*)*lato*)	✓	✓
Free choice (-(*i*)*na*)	✓	✓

Table 1.1 highlights the key synchronic asymmetry that recurs throughout the dissertation: *nwukwu* spans interrogative, existential, and polarity/free-choice uses, whereas *amwu* lacks interrogative and plain existential uses and is largely

confined to polarity-sensitive and free-choice territory. The next section reviews an influential way of connecting this synchronic pattern to diachrony: a clause-to-DP grammaticalization approach to the particles.

1.3 From clause to DP

A widely adopted perspective treats the modern particle-marked forms (e.g. *nwukwu-nka*, *nwukwu-na*, *nwukwu-tunci*) as synchronically DP-like expressions whose particles originated as clausal operators and were later reanalysed as DP-internal markers (J.-M. Yoon, 2005; Yun, 2011). On this view, the particle contributes the relevant nominal operator (existential, scalar, concessive, or disjunctive), but its meaning is historically rooted in a clause-level function.

More concretely, existing work argues that DP-internal particles preserve semantic "residues" of earlier clausal uses: interrogative -(*i*)*nka* is traced to clause-typing morphology and yields epistemic or existential indefinites with *wh*-roots (Kang, 2014; J.-M. Yoon, 2005); disjunctive -(*i*)*na* supports free-choice readings (Yun, 2011); scalar *-to* 'even' yields NPI-like effects; and concessive -(*i*)*lato* 'even if' yields concessive–scalar indefinites (J. Choi, 2007; J. Kim, 2020a, 2020b; C. Lee, 1996; C. Lee et al., 2000; J.-H. Lee, 2010; Lim, 2017). The important point here is not a full derivation, but the architectural consequence: if these particles enter the nominal domain by grammaticalization, then asymmetries in the modern paradigm may reflect which source configurations were historically available to which bases.

Comparable clause $\rightarrow$ DP pathways are well attested cross-linguistically. Japanese indeterminates are built from *wh*-roots plus particles reanalysed as nominal operators (Hiraiwa & Nakanishi, 2021; Kratzer & Shimoyama, 2002; Kuroda, 1965; Nakanishi, 2021; Shimoyama, 2006). In Indo-Aryan and Romance, free-choice expressions often derive from concessive or headless-relative structures (Dayal,

1998; Giannakidou & Quer, 2013). These parallels motivate treating Korean as a particularly clear case of a broader typology: a small inventory of particles is recycled from clausal operators and comes to delimit nominal indefinite meanings.

This clause-to-DP view provides a natural bridge from the descriptive facts in the preceding sections to the core puzzles of the dissertation. In particular, it reframes the existential gap in Table 1.1—the absence of **amwu-inka*—as a diachronic question about which grammaticalization paths were historically available to *amwu* and *nwukwu*.

1.4 Puzzles and working hypotheses

The asymmetries introduced above point to three tightly connected mismatches in the modern paradigm. Each mismatch raises a diachronic puzzle and motivates a working hypothesis.

- **Bases.** *nwukwu* functions as an interrogative and can also receive indefinite readings, whereas *amwu* is never interrogative and is strongly tied to polarity-sensitive and free-choice uses.

  Puzzle 1: How did these base profiles arise historically?
  H1: *amwu* narrows over time toward polarity-sensitive territory, while *nwukwu* expands from interrogative use into adjacent indefinite functions.

- **-(*i*)*nka*.** *wh*+-(*i*)*nka* forms (e.g. *nwukwu-nka*) yield existential or epistemic indefinites, but there is no productive **amwu-inka*.

  Puzzle 2: Why does a *wh*+-(*i*)*nka* series emerge, but no parallel *amwu*+-(*i*)*nka* series?
  H2: -(*i*)*nka* is recruited into the nominal domain via source configurations

available to *wh*-roots (especially clausal question structures), yielding epistemic 'someone-or-other' indefinites. *Amwu* does not participate in the relevant source configurations, and by the time *wh*+-(*i*)*nka* becomes productive, the relevant epistemic or existential niche is already stably occupied elsewhere in the paradigm.

- **Shared particles.** When *nwukwu* and *amwu* combine with the same particle (notably *-to* and -(*i*)*na*), the resulting forms do not freely alternate; instead, they show systematic specialization in strength and pragmatic profile (Giannakidou & Yoon, 2016; M. Kim & Kaufmann, 2006).

  Puzzle 3: What governs the division of labour in shared particle series?
  H3: *amwu*-based forms concentrate in stronger negative and indiscriminative zones, while *nwukwu*-based forms remain in more neutral, extensional, and epistemic zones.

These puzzles define the empirical and theoretical core of the dissertation. The next section situates them in relation to prior historical and synchronic work and formulates the research questions that guide the corpus study.

1.5 Research gap and research questions

This dissertation takes a diachronic, corpus-based approach to Korean indefinites. Although prior work identifies major ingredients of the modern system and sketches key historical facts (A.-R. Kim, 2000; J.-H. Park, 2007; C.-S. Suh, 1989; C.-M. Suh, 1987, 1989; Yun, 2022), it remains unclear how the present paradigm is assembled over time as an internally structured system. What is still missing is a quantitatively grounded, paradigm-level account of how root change, particle grammaticalization, and competition interact to produce the modern distribution.

A particularly influential strand of prior work (summarized by Yun (2022), building on earlier historical surveys) suggests a broad developmental split between the two bases. In early stages, *amwu* (older *amo*) is described as the primary resource for indefinite meanings, whereas *wh*-words are reported to be restricted to interrogative uses. Later, *wh*-based indefinites emerge and expand, yielding the modern situation in which both series can express indefinites but do not overlap uniformly. Figure 1.1 schematizes this reported divergence. The key point of the figure is not frequency but ordering: it highlights that modern overlap and competition in shared particle paradigms are historically late outcomes that require an explanatory account.

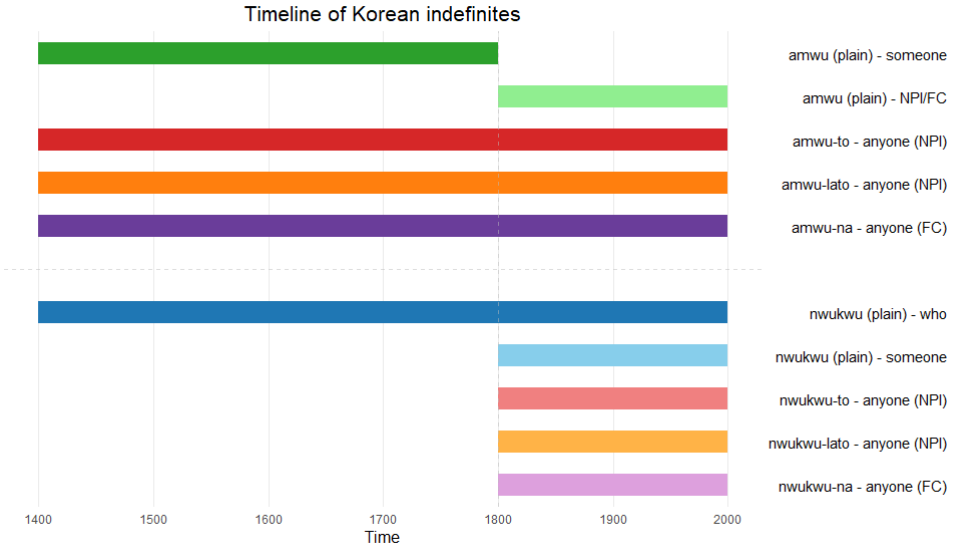


Figure 1.1 Timeline of Korean indefinites

Even with these descriptive generalizations in place, three major gaps remain. First, the literature lacks a quantitatively grounded trajectory for how *amwu* narrows and *nwukwu* expands. Second, we still need an explicit account of how the grammaticalization of -(*i*)*nka* reshapes semantic-map coverage while preserving

the categorical gap **amwu-inka*. Third, we lack a clear explanation of how competition is resolved when both bases are morphologically available in the same particle series.

These gaps motivate the following research questions:

- **RQ1.** What is the timeline and mechanism by which *amwu* shifts from a general existential ('someone') to a polarity-sensitive indefinite ('anyone'), and by which *nwukwu* expands from an interrogative ('who') to an indefinite ('someone')? When does functional overlap first arise, and what mechanisms drive these shifts?

- **RQ2.** How does the distribution of bare *amwu* evolve inside polarity-sensitive environments? In particular, is there evidence that its range narrows over time, becoming increasingly restricted to negative contexts?

- **RQ3.** When and how do *amwu* and *nwukwu* first begin to share particles such as -(*i*)*nka*, -(*i*)*na*, -*to*, and -(*i*)*lato*? How does the grammaticalization of -(*i*)*nka* contribute to the semantic-map distribution of *wh*-based indefinites, and why does this expansion still yield the gap **amwu-inka*?

- **RQ4.** In Modern Korean contexts where both bases are available with the same particle, how are the two forms distributed? Which semantic and discourse-pragmatic factors govern the choice between them, and what does this reveal about paradigm-internal competition?

Together, these questions shift the focus from isolated lexical items to the larger organization of the paradigm. The aim is not merely to ask when a particular form first appears, but to explain how an indefinite system comes to be partitioned across a structured semantic space.

1.6 Theoretical background

To model the three dimensions above—base-level change, particle grammaticalization (especially -(*i*)*nka*), and competition in shared series—I use two cross-linguistic tools.

First, Haspelmath (1997) organizes indefinite functions on a semantic map, where adjacency corresponds to natural pathways of extension and contraction; diachronic change is therefore expected to proceed largely contiguously across neighboring regions of the map (Haspelmath, 2003; Hopper & Traugott, 1993). This provides a geometric way to state and test hypotheses about where *amwu* and *nwukwu* move over time, and where a newly grammaticalized marker such as -(*i*)*nka* comes to locate *wh*-based indefinites.

Second, Jäger (2010) proposes a polarity cline along which indefinites may shift from broad existential uses toward increasingly restricted negative-polarity distributions (Hoeksema, 2010; Horn, 1989). This cline offers a direct way to track the narrowing component of the paradigm, particularly the tendency for *amwu* to become concentrated in negative and other polarity-sensitive environments.

Together, the semantic map and the polarity cline make it possible to ask not only *where* each form is licensed synchronically, but also *how* its functional coverage is reorganized over time. They also provide a shared metric for competition: when two base+particle forms coexist, specialization can be modeled as a partitioning of nearby regions on the map.

At the same time, these tools make only constrained predictions. They are well suited to predicting likely directions of change—especially local extensions and retreats across adjacent nodes—but they do not by themselves determine timing, frequency, or which competitor ultimately wins a given niche. Those outcomes depend on additional factors such as source constructions, usage skew, and paradigm-internal competition.

Korean is especially revealing because the two bases are reported to move in opposite directions: *amwu* narrows toward polarity-sensitive uses, whereas *nwukwu* expands from interrogative use into adjacent indefinite territory. This internal asymmetry makes it possible to evaluate map contiguity, polarity narrowing, and competition within a single paradigm.

1.7 Methodology and preview of findings

To address the research questions in Section 1.5, this dissertation adopts a usage-based, diachronic corpus method. The core assumption is that the organization of an indefinite paradigm is visible in speakers' and writers' choices over time: shifts in relative frequency, the emergence of new form–function pairings, and the stabilization of distributional gaps are treated as evidence for entrenchment, reanalysis, and paradigm-internal competition. Empirically, I track the two bases *nwukwu-* and *amo-/amwu-* across roughly three centuries of written Korean using two corpora: the History corpus (ca. 1700–1890) and the Yonsei corpus (1900–1990). From these corpora I extract tokens of the bare bases as well as particle-marked forms that are central to the Modern Korean paradigm, including -(*i*)*nka*, -*to*, and -(*i*)*na* (and, where relevant, related particles such as -(*i*)*lato* and -*tunci*). All automatically retrieved hits are manually screened to remove homographs and lexicalized expressions, and each token is retained with its decade metadata and sufficient context for functional interpretation.

Each token is annotated for its semantic function on an extended Haspelmath-style semantic map (Degano & Aloni, 2022). The annotation inventory distinguishes the core nodes and polarity environments needed for the analysis (e.g. SK/SU for specific indefinites, IR/CA/Q/CO for nonveridical and concessive contexts, and DN/AA/AM for increasingly strong negative licensing). A step-by-step decision procedure, including the diagnostics used to assign labels, is provided in

Appendix A. For analyses that target broad reorganizations, the fine-grained labels are additionally collapsed into five macro-environments: Q_WH (wh-questions), PPI (SK, SU), NEUTRAL (IR, Q, CA), NPI (DN, AA, AM), and FREE-CHOICE (FC, GEN, UFC, IND). This two-level coding makes it possible to compare (i) coarse shifts between interrogative, existential/neutral, negative, and free-choice regions and (ii) finer redistributions within those regions.

Change is quantified in two complementary ways. First, token counts are normalized by corpus size and tracked by decade, and I fit logistic models for contrasts that correspond to major cuts in the space (e.g. Q_WH vs. non-Q, NPI vs. non-NPI). Where the trajectories show clear regime shifts, segmented logistic models are used to estimate approximate change points. Second, to connect timing of change to semantic-map geometry, I use an entry-threshold method: for each series and each label ℓ, the *entry decade* is defined as the first decade in which ℓ reaches both (i) at least five tokens and (ii) at least 5% of that series. This provides a conservative criterion for when a function becomes a stable part of a paradigm rather than a rare outlier, and it allows a direct comparison between entry timing and adjacency on the map.

The methodology is developed in detail in Chapter 3 and applied in the three empirical chapters, each corresponding to one dimension of the dissertation. The main findings, previewed here in schematic form, are as follows.

- **Roots** (**Chapter 4**). The two bases move in opposite directions. Bare *nwukwu-* expands from predominantly interrogative use into adjacent indefinite functions, especially in specific-unknown and neutral or nonveridical environments, while bare *amo-*/*amwu-* narrows along a polarity cline, losing earlier existential and free-choice uses and becoming increasingly concentrated in negative licensing environments, particularly stronger negative contexts.

- **Grammaticalization of -(*i*)*nka*** (**Chapter 5**). *Wh*+-(*i*)*nka* develops through a clause-to-DP pathway: analytic *nwukwu-i-nka* remains primarily clausal (question-like), while a reduced form (*nwukwunka*) stabilizes as a DP-level epistemic indefinite ('someone-or-other'). The persistent absence of a productive **amwu-inka* series follows from the interaction of timing, source availability, and paradigm-internal niche occupation rather than from an arbitrary morphological gap.

- **Competition in shared particle series** (**Chapter 6**). When *amwu-* and *nwukwu-* share the same particle, they do not simply duplicate one another; instead, they partition the semantic space. In the negative *-to* series, *amwu-to* concentrates toward the strongest negative region, while *nwukwu-to* shows a broader distribution that connects negative and nearby nonveridical or concessive uses. In the *-na* series, *nwukwu-na* dominates extensional free-choice and generic uses, while *amwu-na* shows a stronger pull toward indiscriminative free choice. Across series, specialization respects semantic-map adjacency: expansions and retreats proceed largely by movement to neighboring nodes rather than by jumps across unrelated functions.

In sum, these results support the dissertation's central claim: the asymmetries of Modern Korean indefinites are best explained at the level of the paradigm. Usage-driven reweighting of functions, clause-to-DP grammaticalization of particles, and competition in overlapping series jointly determine where each form 'lives' on the semantic map and why certain gaps and specializations persist. More broadly, the Korean case argues that indefinite systems can be modeled as diachronic partitions of a structured semantic space, not merely as collections of independent form–meaning pairs.

1.8 Roadmap

This introduction has identified the empirical domain and articulated three connected problems that structure the dissertation: (i) lexical change in the bases *amwu* and *nwukwu* over time, (ii) the grammaticalization of particle morphology —especially -(*i*)*nka*—and its consequences for the semantic-map distribution of indefinites, including the persistent absence of **amwu-inka*, and (iii) paradigm-internal competition in Modern Korean when both bases are available in the same particle series.

The remainder of the dissertation develops and tests these ideas as follows. Chapter 2 (*Indefinites: Form, Function, Polarity, and Change*) builds the theoretical and typological background needed for the analysis: it introduces cross-linguistic strategies for forming indefinites, semantic-map approaches to indefinite functions, the core notions in polarity theory, and the main pathways of change that link clausal operators to DP-internal morphology. Chapter 3 then presents the corpora, annotation scheme, and statistical procedures used to track change on the semantic map. Chapter 4 investigates the diachronic trajectories of the bare bases *amwu* and *nwukwu*. Chapter 5 examines the emergence of the *wh*+-(*i*)*nka* series and explains why no productive **amwu-inka* develops. Chapter 6 turns to Modern Korean and analyzes how competition is resolved in shared particle series such as -*to* and -(*i*)*na*. Chapter 7 concludes.

I now turn to Chapter 2, which establishes the conceptual tools—form, function, polarity, and change—needed to interpret the diachronic patterns in the corpus.

Chapter 2

Indefinites: Form, Function, Polarity, and Change

This chapter assembles the conceptual and descriptive tools needed for the three core questions that the dissertation addresses in the Korean domain. Those questions all concern how indefinites are distributed across the functional space that Haspelmath (1997) models with a semantic map, how they behave in polarity-sensitive environments, and how those distributions change over time. What differs across the three empirical chapters is the point in the system at which change is observed.

The *first puzzle* (Chapter 4) concerns a *base shift*. Diachronically, bare *nwukwu-* 'who' comes to be used as an indefinite meaning roughly 'someone', while *amwu-*, which earlier descriptions treat as a 'someone'-type existential, becomes specialized for polarity-sensitive 'anyone'-type meanings. In the terms developed later in Section 2.4.1, this raises the question of how the two bases move in opposite directions along what I will call a *polarity path*: *nwukwu-* expands from a purely interrogative root into an indefinite, whereas *amwu-* narrows from a plain existential into an NPI-like item.

The *second puzzle* (Chapter 5) is what I will call the *inka asymmetry*. Synchronically, Korean has indefinites formed by attaching *-(i)nka* to *wh*-roots, such as *nwukwu-inka* 'someone (or other)'; these are commonly analyzed as existential or referentially vague indefinites. In contrast, the combination **amwu-inka* is not available in the ordinary system. This raises two related questions: what the

internal semantics of *wh-inka* forms is (compositional analyses vs. grammaticalization accounts that derive them from clausal Q-morphemes), and why this pattern cannot be replicated with *amwu-* as the base.

The *third puzzle* (Chapter 6) concerns *competition between pairs* when both bases are available with the same particle. In the modern language, *nwukwu-* and *amwu-* both combine with *-to*, yielding *nwukwu-to* and *amwu-to* in strongly negative environments, and they also combine with *-na* in free-choice environments. This invites the question of how the two series share or divide up functional space when they are morphologically parallel. The empirical focus of Chapter 6 will be on the *-to* and *-na* pairs, with *-tunci* treated more briefly and left for future work.

The present chapter establishes three strands of background that feed directly into these puzzles. Section 2.1 reviews general theories of indefinites, polarity, and semantic maps, using English as a baseline and then broadening to a typological perspective. Sections 2.2–2.4.1 review major approaches to polarity licensing and diachronic change, with special attention to feature-based paths and grammaticalization sources. Section 2.5 then introduces the Korean paradigm itself, built from *wh*-roots, *amwu-*, and a small inventory of particles, and lays out the three asymmetries that structure the dissertation. The larger goal is not to provide an encyclopedic survey of indefinites, but to build a focused toolkit for asking how Korean forms move, stabilize, and compete in a shared semantic space.

2.1 The landscape of indefinites: semantics and typology

Before turning to Korean, it is useful to recall how indefinites are modelled in formal semantics and how they are organized typologically. This section provides the general toolkit that will be applied to the Korean case in later chapters.

Section 2.1.1 briefly reviews three standard semantic perspectives on indefinites —existential quantification, choice functions, and dynamic discourse reference—in

order to fix terminology and to distinguish plain existential uses from the enriched readings that will matter for Korean. Section 2.1.2 then turns to the main morphological strategies for building indefinites and to semantic-map approaches. The typological payoff of that section is central for the rest of the dissertation: it places plain existentials, polarity-sensitive items, free-choice items, and ignorance indefinites in a single structured space, and it makes it possible to ask how Korean series move within that space over time.

2.1.1 Indefinites as existential quantifiers and beyond

In what follows, I review the standard semantic treatments of indefinites—as existential quantifiers, choice functions, and dynamic discourse referents—in order to fix terminology and to distinguish plain existential uses from the enriched readings that will be central to the Korean data.

English indefinites provide a particularly clear case study because their morphological alternations reveal otherwise hidden aspects of licensing. The distinction between *some-* and *any-* forms has long been taken as evidence for polarity sensitivity (Giannakidou, 1998, 2011; Israel, 1996, 2004; Kadmon & Landman, 1993; Krifka, 1995; Ladusaw, 1979, 1980; Zwarts, 1995). Before turning to polarity, however, it is useful to establish the existential core of indefinites and the interpretive flexibility they exhibit (Montague, 1973).

In their simplest use, indefinites contribute existential force, as in (1).

(1) Someone left early.

Following Heim (1982) and Kamp (1981), such a sentence is true iff there exists at least one person who left early. This corresponds to the existential quantifier $\exists x$ and captures the intuition that indefinites assert the existence of at least one entity with the relevant property.

Very quickly, however, indefinites move beyond this simple quantificational role. A key observation in the modern semantics literature, emphasized by Fodor and Sag (1982), is that indefinites can be interpreted either specifically or non-specifically (see also Enç 1991; Farkas 2002). Consider (2).

(2) A student in my class won a prize.

This sentence has two readings. On a *specific* reading, the speaker has a particular student in mind, perhaps Alice, and uses the indefinite to refer to her. On a *non-specific* reading, the speaker simply reports that some student or other won a prize, without knowing or committing to who that student is. The contrast shows that indefinites are not reducible to bare existential quantifiers; their interpretation interacts with discourse context and scope (Heim, 1982; Kamp, 1981; Karttunen, 1976).

To capture this flexibility, Reinhart (1997) and Winter (1997) develop *choice-function* analyses. On this view, an indefinite denotes a function that selects an element from a set. This makes it possible for indefinites to take wide scope without syntactic movement (see also Kratzer 1998; Chierchia 2001; Kratzer and Shimoyama 2002). Consider (3).

(3) Everyone wants to marry a rich celebrity.

In (3), *a rich celebrity* is ambiguous. On one reading, there is a single particular celebrity that everyone wants to marry. On another, each person may have a different celebrity in mind. Under a choice-function analysis, the wide-scope reading is available because the indefinite can select a particular individual across all cases without moving syntactically.

Dynamic semantics offers a complementary perspective. Kamp (1981) and Heim (1982) treat indefinites not as quantifiers but as instructions to introduce new discourse referents.

(4) If a farmer owns a donkey, he beats it.

Here, the pronouns *he* and *it* clearly depend on the indefinites in the antecedent, even though no explicit quantifier binds them. On a dynamic view, the indefinites introduce referents into the discourse model that remain accessible in the consequent (Groenendijk & Stokhof, 1991; Kadmon, 1987).

Both approaches—choice functions and dynamic semantics—support the same core insight: indefinites may look simple, but their interpretive possibilities go beyond plain existential force. They participate in scope ambiguities, they can be specific or non-specific, and they interact with discourse in ways that ordinary quantifiers do not. This foundational complexity sets the stage for the more articulated distinctions that will matter for Korean: plain existentials, polarity-sensitive uses, free-choice readings, and ignorance indefinites.[1]

2.1.2 Morphological strategies and semantic maps

I now turn to the morphosyntactic realization of indefinites and to semantic-map approaches. The goal is to link concrete morphological series (*some-*, *any-*, *no-*, *nwukwu-*, *amwu-*, etc.) to their functional coverage and to provide a typological benchmark against which the Korean system can be measured.

[1]In this chapter I use *specific* and *non-specific* in the sense standard in formal-semantic work on indefinites (Enç, 1991; Fodor & Sag, 1982). An indefinite is *specific* when it is used to pick out a particular individual, so that subsequent reference (e.g. with a demonstrative like 'that person') is licensed, and *non-specific* when it does not introduce a single discourse referent (for example, when it ranges over possibilities or remains purely existential). This usage does not coincide exactly with the labels on Haspelmath (1997)'s semantic map, where 'specific' is split into *specific known* and *specific unknown* based on the speaker's epistemic access to the individual. More generally, the term *specificity* is well known to be used in multiple, partially independent ways in the literature (Partee, 1987). In the present study, I keep the formal-semantic contrast between specific and non-specific for the diagnostics in this chapter, while retaining Haspelmath's original function labels (including *specific unknown*) when I explicitly discuss the semantic map in later chapters, in order to facilitate comparison with the typological literature.

Although English provides a convenient entry point, it is only one instance of a broader typological landscape. Cross-linguistically, languages construct indefinites using a limited set of morphological strategies, each aligned with recurring functional behaviours. Haspelmath (1997, 2005) and Dryer and Haspelmath (2013) identify three primary strategies, with mixed systems also common (see also Haspelmath 2013; Matthewson 1999; von Heusinger 2011):

(5) a. ***Wh*-based systems.** A single interrogative root combines with different particles or clitics to yield existential, negative, or free-choice meanings. Japanese is a textbook case: *dare* 'who' forms *dare-ka* ('someone'), *dare-mo* (with negation, 'no one'), and *dare-demo* ('anyone' with a free-choice reading) (Hiraiwa & Nakanishi, 2021; Kratzer & Shimoyama, 2002; Kuroda, 1965; Nakanishi, 2021; Shimoyama, 2006; Uegaki, 2018). Korean parallels this with *nwukwu* 'who', yielding *nwukwu-nka*, *nwukwu-to*, *nwukwu-lato*, and *nwukwu-na*, alongside the dedicated root *amwu-* specialized for polarity-sensitive environments (J. Choi, 2007; Giannakidou & Yoon, 2016; Shimoyama, 2001; Yun, 2013).

b. **Generic noun compounds.** Many European languages form indefinites through compounding. English *somebody*, *anyone*, *somewhere*, French *quelqu'un*, and German *jemand*, *niemand*, among others, illustrate this path. These forms often become morphologically opaque but preserve robust functional alignments (Dryer & Haspelmath, 2013; Haspelmath, 1997, 2013).

c. **Dedicated roots.** Modern Greek has *kapjos* ('someone'), *kanenas* (NPI), and *opjosdhipote* (free choice) (Giannakidou, 1998, 2001, 2011). Dedicated indefinite roots are also attested in languages such as St' át' imcets (Matthewson, 1999) and in various Romance and Germanic varieties (von Heusinger, 2011).

Mixed systems are frequent. Korean employs both *wh*-based forms and *amwu-*, while Greek mixes dedicated forms with particles and compounding (J. Choi, 2007; Giannakidou, 1998, 2011; Giannakidou & Yoon, 2016; Yun, 2013). Across types, the same functional contrasts recur: plain existential indefinites, polarity-sensitive items, free-choice items, and ignorance or epistemic indefinites (Chierchia, 2013; Haspelmath, 1997; Menéndez-Benito, 2010). This typological stability suggests that the space of indefinite meanings is tightly constrained and that morphological strategies track stable functional clusters rather than arbitrary meanings.

Haspelmath (1997) models these clusters by arranging indefinite functions on a *semantic map*. Nodes correspond to uses such as interrogative, plain existential, specific unknown, negative, conditional, and free choice, and attested forms typically occupy *contiguous* regions of this space. The map was originally developed as a synchronic tool to describe how indefinite series are distributed across functions in different languages; later in the chapter I return to its diachronic implications.

A particularly influential instantiation of this approach is Haspelmath's semantic map for indefinite pronouns, based on a sample of 40 languages (Haspelmath, 1997). The map arranges nine functions of indefinites so that each is adjacent to semantically related ones. In the English system Haspelmath distinguishes nine core functions: specific known, specific unknown, irrealis non-specific, question, conditional, indirect negation, direct negation, comparative, and free choice. These labels will be useful shorthand in what follows.

All of these functions are instantiated by English indefinites. The examples in (6) illustrate them (Haspelmath, 1997, p. 64):

(6) a. Someone called yesterday⋯Guess who? (specific known)

b. Somebody called⋯I don' t know who it was. (specific unknown)

c. Buy me some newspapers. (irrealis non-specific)

d. Did you see anything? (question)

e. I don' t think anybody has seen it. (indirect negation)

f. I didn' t see anything. (direct negation)

g. If you hear anything, let me know. (conditional)

h. The boy runs faster than anyone in his class. (comparative)

i. Anyone can solve the problem. (free choice)

In (6a–c), *some-* indefinites realise three distinct functions: a speaker-known specific referent (a), a specific but speaker-unknown referent (b), and an irrealis non-specific referent in a directive (c). In (6d–i), *any-* indefinites realise the remaining functions: question, indirect and direct negation, conditional, comparative, and free choice. Crucially, not all functions are available to both series. As (7) shows, specific known and specific unknown uses are restricted to *some-*, while free choice is restricted to *any-*; questions and conditionals allow both.

(7) a. {Someone/*anyone} called yesterday⋯Guess who?

b. {Somebody/*anyone} called⋯I don' t know who it was.

c. *Someone/anyone can solve the problem.

d. Did you see {something/anything}?

e. If you hear {something/anything}, let me know.

Haspelmath (1997) proposes that these nine functions are not an arbitrary list but form a structured space that indefinite series can move around in. A simplified version of his map for indefinites is given in Figure 2.1.

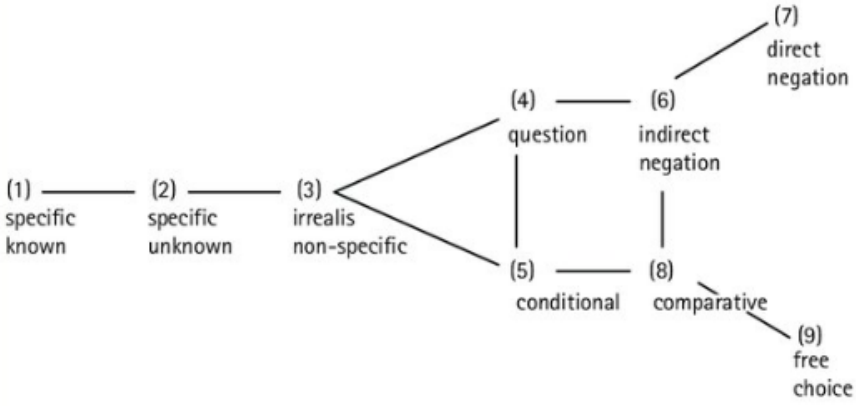


Figure 2.1 A semantic map for indefinite pronouns (Haspelmath, 1997, p. 64).

The core principle is one of *functional contiguity*. For a given morphological series S (for example, the English *some-*, *any-*, or *no-* series), let $F(S)$ be the set of functions on the map that S can realise. Contiguity requires that $F(S)$ form a connected region of the map: if S is used for two functions that are separated by an intermediate node, then S must also be available for that intermediate function. Concretely, if the *some-* series is attested at both the specific-known and irrealis non-specific nodes, it must also be possible at the specific-unknown node, which

lies between them. Likewise, if the *any-* series is used in questions and free-choice contexts, the intervening nodes—conditionals, indirect negation, direct negation, and comparatives—are predicted to be possible *any-* environments as well. This is exactly what we observe in English: the *any-* series forms a continuous chain from questions and conditionals through indirect and direct negation and comparatives to free choice.

This contiguity principle will matter later when I ask not only *where* Korean forms occur, but *how* they get there. The map is well suited to predicting local pathways of extension and retreat, but it does not by itself predict exact timing, token frequency, or which of two competing series will win a given niche. Those issues depend on additional factors such as source constructions, distributional skew, and paradigm-internal competition.

From Haspelmath to extended maps. Subsequent work has shown that Haspelmath's nine-way inventory is often too coarse to capture the full functional range of *any*-like items. Aguilar-Guevara et al. (2010) and, building on them, Degano and Aloni (2022) adopt an *extended* semantic map that keeps Haspelmath's core nodes (SK, SU, IR, Q, CA, CO, DN, FC) but adds further distinctions on both the free-choice and negation side. Their Table 3 explicitly relabels Haspelmath's functions as (a–h) and introduces additional functions (i–n), with two further labels (o–p) added for Italian-specific uses of *qualsiasi*.

The extended map I adopt is schematized in Figure 2.2, which refines the 'free choice' and 'indirect negation' sectors of Haspelmath's original map:

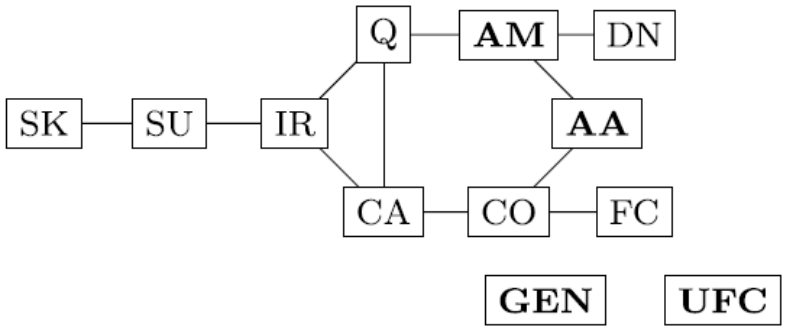


Figure 2.2 Extended semantic map for indefinites (after Aguilar-Guevara et al., 2010; Degano & Aloni, 2022).

On this extended map, the Haspelmath-type functions SK, SU, IR, Q, CA, CO, DN, and FC are supplemented by the following additional nodes (labels as in Table 3 of Degano and Aloni (2022, p. 455)): AM (anti-morphic), AA (anti-additive), UFC (universal free choice), GEN (generic), *no matter*, and IND (indiscriminacy).[2]

In Haspelmath's map, all 'any'-like uses under possibility or in generic statements are lumped under a single FC node. The extended map decomposes this domain into more fine-grained functions that are adjacent to FC but not identical to it:

- [FC] is the original Haspelmath 'free choice' node, illustrated by sentences like *You may kiss anybody*. Here the indefinite contributes a permission-like inference: the speaker asserts existence (*you may kiss someone*) and, crucially, that all contextually relevant options are permitted.

- [UFC] (universal free choice) corresponds to cases where the same form has a genuinely universal reading, typically under a necessity modal or in maximising environments, as in *John read any book on lions he could find*. In such

[2]Degano & Aloni (2022) furthermore introduce two additional labels, US ('universal strengthening') and *expression*, to annotate marginal or idiomatic uses of Italian *qualsiasi*, which lie at the periphery of the map and will not play a central role here.

uses, the indefinite ranges over all relevant books, rather than just signalling that 'any option would do'.

- [GEN] captures generic, kind-level generalizations such as *Any dog has four legs*. These sentences behave more like universal generics than like plain modal FC statements: they quantify over instantiations of a kind across time, not over the modal alternatives in a single situation.

- [IND] (indiscriminacy) is an 'off-map' function in the sense of Degano and Aloni (2022): it uses the same morphology as FC but expresses a distinct, typically existential reading with an 'average/just any' or pejorative flavour, as in *I don' t want an ordinary / just any book*. IND is usually existential and comments on the *quality* of the chosen individual rather than on the set of permitted options.

By differentiating AM and AA from DN, the extended map predicts that a morphological series might be licensed in DN and AM but not AA, or in AA but not AM, while still respecting contiguity: the negative sector itself is internally articulated. This finer structure turns out to matter for Italian *qualsiasi*, which shows a cluster of UFC and FC uses but only marginal realization in DN, AM, and AA environments, and it is precisely the kind of subtlety we will want to track for Korean indefinites as well.

In addition to the nodes just discussed, Aguilar-Guevara et al. (2010) introduce a *no matter* function to capture concessive 'no-matter-what' readings (equivalent to English *whatever*-type constructions) that do not behave like ordinary indefinites but are clearly related to the same morphological material. Degano and Aloni (2022) show that such *no matter* uses play a crucial diachronic role in the emergence of Italian *qualsiasi* as a free-choice indefinite, even though synchronically they sit

slightly off the core map. I do not treat *no matter*, US, or *expression* as core indefinite functions for Korean, but they are useful diagnostic labels for certain marginal uses.

Implications for Korean. Importantly, Haspelmath's map is not tailored to English. Haspelmath (1997, p. 75) explicitly shows how different morphological series in other languages carve up the same functional space. For Korean, he notes that:

(8) a. *-na* and *-tunci* occupy the free-choice sector of the map;

b. *-to* appears in the direct- and indirect-negation and comparative zones;

c. -(*i*)*nka* is associated with specific-known and specific-unknown uses, irrealis non-specifics, questions, and conditionals.

Already in Haspelmath's original typology, then, Korean *-na*/*-tunci*, *-to*, and -(*i*)*nka* are seen as systematically dividing up the same semantic space that English *some-*/*any-*/*no-* cover. The present dissertation takes this as a starting point but follows Degano and Aloni (2022) in adopting the extended map in Figure 2.2. This allows us to distinguish, for example, between plain FC and truly universal (UFC/GEN) uses of Korean free-choice markers, and between their behaviour in DN, AM, and AA environments, rather than lumping all of these under a single FC or 'indirect negation' label. In addition, I refine the Korean picture by adding the choice of base (*nwukwu-* vs. *amwu-*) and by examining in detail how these series are distributed across FC, UFC, GEN, IND, and the various negative subtypes in contemporary and historical Korean.

In sum, Haspelmath's semantic map provides a synchronic typological baseline: it tells us which functions tend to cluster and how different morphological series divide the space (Haspelmath, 1997). The extended map of Aguilar-Guevara et al. (2010) and Degano and Aloni (2022) further articulates the free-choice and

negation sectors into FC/UFC/GEN/IND and DN/AM/AA, respectively, giving us a more fine-grained toolkit for mapping English and Korean indefinites into the same structured domain. In Section 2.4, I return to this extended map from a diachronic perspective and ask how the Korean series move within this enriched space over time.

2.2 Polarity phenomena and licensing

The second component of the general background concerns polarity phenomena: negative polarity items (NPIs), positive polarity items (PPIs), and free-choice items (FCIs). Roughly, NPIs are expressions that prefer negative or otherwise 'weak' environments, PPIs avoid such environments, and FCIs tend to appear in modal and generic contexts with a quasi-universal flavour (Chierchia, 2013; Dayal, 1998, 2004; Giannakidou, 1998, 2011; Ladusaw, 1979, 1980; Menéndez-Benito, 2010). These categories are central to all three puzzles in the Korean system: the reanalysis of *amwu-* as an NPI-type item (Puzzle 1), the status of *wh-inka* as neither a classic NPI nor a canonical free-choice item (Puzzle 2), and the competition between *nwukwu-* and *amwu*-based series that share the same particle (Puzzle 3).

Section 2.2.1 uses the English *some/any* alternation as a baseline. It introduces the dimensions along which indefinites vary—polarity sensitivity, free-choice behaviour, and specificity—and shows how these dimensions come apart even in a single language. Section 2.3.2 then distils the main syntactic and semantic or pragmatic ideas about polarity licensing: downward entailment, focus-based alternatives, domain widening and scalar strengthening, and the broader notion of non-veridicality. These notions will be used repeatedly in later chapters, for instance when characterising what it means for *amwu-* to 'gain' polarity features or for *-to* to pattern as a strong NPI, in contrast to the weaker *-lato* and the free-choice particles *-na*/*-tunci*.

2.2.1 NPIs, PPIs, FCIs: English *some*/*any* as baseline

I begin with polarity sensitivity in English, using the *some*/*any* alternation as a model. NPIs are expressions that appear only in a restricted class of environments, most typically under negation or in other non-assertive contexts, while PPIs show the opposite tendency and avoid negation (Giannakidou, 1998, 2011; Ladusaw, 1979, 1980). FCIs, finally, tend to appear in modal and generic environments with a quasi-universal interpretation (Chierchia, 2013; Dayal, 1998, 2004; Giannakidou, 2001; Menéndez-Benito, 2010).

English provides a particularly clear illustration, because the *some-* series and the *any-* series distribute differently:

(9) a. I didn't see any mistakes.

b. Did you see anything?

c. You may pick any card.

In ordinary affirmative contexts, forms with *some-* are natural (*I saw someone at the party*), while *any-* is infelicitous (**I saw anyone at the party*). Under negation, however, *any-* is required, as in (9a). Similarly, in polar questions like (9b) and in permission contexts like (9c), *any-* appears instead of *some-*.

Standardly, *any-* in (9a) and (9b) is analyzed as an NPI: it is licensed in negative and interrogative environments but blocked in simple affirmative assertions (Giannakidou, 1998, 2011; Ladusaw, 1979, 1980). In (9c), by contrast, *any* displays a free-choice reading: *any card* is no longer a simple existential but approximates a universal, 'whichever card' choice (Chierchia, 2013; Dayal, 1998, 2004; Giannakidou, 2001; Menéndez-Benito, 2010). English thus shows in one morphological family the three categories that will be central in the Korean case: NPIs, PPIs (such as *some* in certain positions), and FCIs.

The English pattern also illustrates that polarity sensitivity, free choice, and specificity are related but distinct. Specific indefinites such as *someone I know* can appear in positive contexts without being polarity sensitive, and free-choice readings of *any* are tied to modals rather than to plain negation. In the remainder of this section I abstract away from these language-specific details and focus on general licensing conditions that have been proposed for NPIs and FCIs.

2.3 Polarity licensing

The preceding sections treated polarity sensitivity mainly from the perspective of morphological contrasts and semantic maps. I now turn to the more traditional question of *licensing conditions*: under what structural and interpretive conditions are polarity-sensitive indefinites acceptable? I begin with syntactic accounts and then move to semantic and pragmatic approaches. Throughout, the English *any*-series serves as a reference point for later comparison with Korean.

2.3.1 Syntactic licensing and scope

One important dimension of polarity behaviour is distributional: certain indefinites can appear only in a subset of syntactic environments, even when other, morphologically similar indefinites are less restricted. From a syntactic perspective, early generative work treated NPIs as elements that require a local structural relation with a licensor, typically clause-mate negation (Klima, 1964).

Later feature-based implementations often model negative concord and *n*-words in terms of syntactic dependency with a negative head, sometimes accompanied by overt movement in negative-concord systems (Giannakidou, 2000; Progovac, 1993).

In English, however, the relevant dependency is not always reflected in surface order, especially for subjects. Consider:

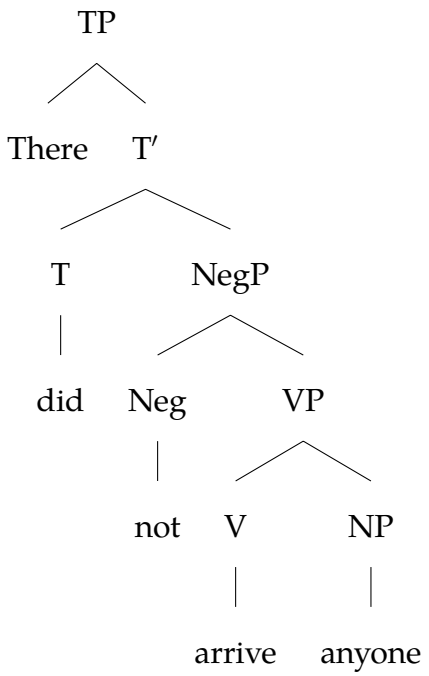
(10) *Anyone didn't leave.

The sentence in (10) is ungrammatical because the NPI *anyone,* in [Spec, TP], structurally precedes sentential negation *not*. This violates the requirement that the NPI be c-commanded by the licensor. To capture such contrasts, many theories appeal to *scope-based licensing*: NPIs must be interpreted in the scope of their licensor at Logical Form (LF), even if surface order suggests otherwise (Giannakidou, 1998; Ladusaw, 1979, 1980; Linebarger, 1980, 1987).

(11) There didn't arrive anyone.

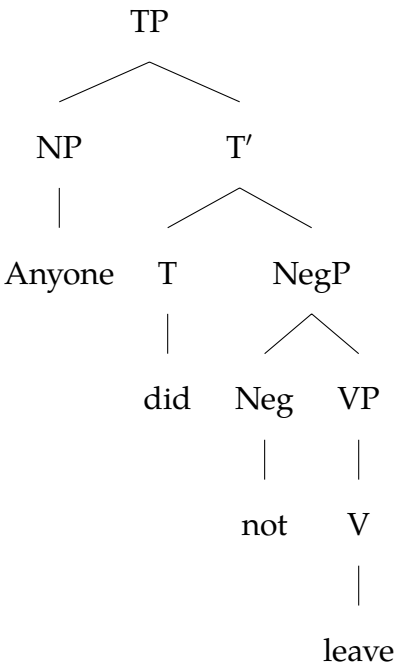
In (11), the NPI *anyone* remains within the VP and under the scope of negation, thus satisfying licensing requirements. These configurations are typically represented in terms of c-command and scope relations at LF. For instance, a simplified LF structure for (11) is given in (12):

(12) LF structure for sentence (11):



In contrast, in the ungrammatical (10), the NPI is merged too high in the structure and is not in the scope of negation:

(13) Structure for sentence (10):



In both cases, licensing hinges on syntactic factors such as c-command, scope, and structural locality. Syntactic theories of NPI licensing therefore interface with semantic theories: the same configurations that allow the relevant scope relations must also correspond to semantic environments that support NPIs (Giannakidou, 1998; Ladusaw, 1980).

These principles extend beyond simple clauses to embedded contexts such as questions and conditionals. In the conditional *If anyone arrives, call me*, the protasis introduces a non-assertive environment that licenses the NPI *anyone* even in the absence of overt negation (Giannakidou, 1998; Karttunen, 1977). In the next subsection, I build on these structural foundations by turning to semantic and pragmatic accounts of polarity licensing.

2.3.2 Semantic and pragmatic licensing

If syntactic licensing provides the necessary structural conditions, semantics and pragmatics supply the interpretive principles that determine when those conditions yield acceptability. Three broad semantic strands are standardly distinguished—monotonicity-based licensing, alternative or scalar licensing, and nonveridicality-based licensing—with free-choice phenomena forming a partly overlapping line (Chierchia, 2013; Dayal, 2004; Giannakidou, 1998; Kadmon & Landman, 1993; Ladusaw, 1980; Lahiri, 1998). This subsection briefly revisits these strands in a more licensing-oriented perspective, complementing the earlier discussion in Section 2.1.1.

Downward entailment

Ladusaw (1980) proposes that NPIs are licensed in *downward-entailing* (DE) environments, where inferences from supersets to subsets are valid (see also Zwarts 1995). Negation is the classic case: from (14) one can infer that John did not see a dog, did not see a black dog, and so on.

(14) John didn't see anything.

Let $\varphi[\alpha]$ be a sentence with predicate α. A context is DE in the position of α iff whenever $\alpha \subseteq \beta$, we have $\varphi[\beta] \Rightarrow \varphi[\alpha]$. Under this view, NPIs are licensed under negation and in other DE environments (Ladusaw, 1980; Zwarts, 1995). However, questions and many modal contexts license NPIs without being DE in any straightforward sense, motivating generalizations beyond monotonicity (Giannakidou, 1998).

Focus and alternatives

Lahiri (1998) develops a focus-based account that unifies NPIs, minimizers, and *even*. On his view, many NPIs are built from a minimal element (such as *one*) combined with a focus particle *even*. The resulting expression carries a scalar presupposition over a set of alternatives and is licensed only where this presupposition matches independent expectations about likelihood.

In Hindi, the NPI *ek bhii* 'even one' consists of *ek* 'one' plus *bhii* 'even'. In a simple positive sentence, focus on *one* yields alternatives like {one came, two came, three came, ⋯}; *even* presupposes that 'one came' is the least likely among them. This clashes with the monotonicity of cardinals, since 'one came' is entailed by all stronger alternatives, and the result is infelicitous.

Under negation, by contrast, *ek bhii* is well formed. If *even* scopes above negation at LF, the relevant alternatives are {one did not come, two did not come, three did not come, ⋯}, and now it is natural to treat 'one did not come' as the least likely. Negating this most expected positive outcome yields a strengthened negative ('not even one' ≈ 'no one').

Lahiri's focus-and-alternatives architecture makes explicit how scalar presuppositions, minimal elements, and negation interact to derive NPI behaviour. In later chapters I will adopt this framework as a template for the compositional analysis of Korean polarity items, in particular the *amwu-+-to* series.

Domain widening and scalar strengthening

Kadmon and Landman (1993) argue that English *any* combines *domain widening* with *scalar strengthening*. First, it widens the domain of quantification to include even marginal or unexpected individuals; second, it asserts that the predicate holds across this widened domain.

(15) a. I didn't see any mistakes.

b. *I saw any mistakes.

In (15a), this yields a strengthened denial: the speaker did not see even the smallest or least expected mistakes. In (15b), domain widening in a simple positive statement is pragmatically odd; nothing in the context calls for such a strong claim. This perspective connects NPIs to scalar items such as *even* and minimizers (*at all*, *a shred*), which likewise operate on scales of likelihood or informativeness (Chierchia, 2013; Israel, 1996; Kadmon & Landman, 1993; Lahiri, 1998).

Free choice and alternatives

Free-choice uses of *any* form a closely related but partly distinct pattern. In permission contexts such as (16), *any* receives a near-universal interpretation: the addressee may choose freely among the relevant alternatives.

(16) You may take any seat.

Dayal (1998, 2004) analyse free-choice items as inherently modal and universally quantified under modals. Chierchia (2013), by contrast, derives the universal flavour via alternatives and exhaustification; related alternative-based accounts are developed by Menéndez-Benito (2010) and Giannakidou (2001). Both approaches predict that free-choice items cluster in modal and generic contexts and resist plain episodic assertions.

Nonveridicality

Finally, Giannakidou (1998, 2011) propose *nonveridicality* as a unifying condition on polarity licensing. An operator F is *veridical* if $F(p)$ entails p; *antiveridical* if $F(p)$ entails $\neg p$; and *nonveridical* if it entails neither p nor $\neg p$. Negation is antiveridical;

questions, conditionals, and many modals are nonveridical, since they do not commit the speaker to the truth of their embedded proposition.

On this view, NPIs are licensed in the scope of nonveridical operators, while the strongest NPIs may be restricted to antiveridical ones. Nonveridicality thus covers classic negation cases and also the broader range of environments where NPIs and free-choice items are found, such as questions, conditionals, and possibility modals (Giannakidou, 1998, 2011).

To summarize, syntactic accounts of polarity licensing emphasize c-command, scope, and locality between NPIs and their licensors, while semantic and pragmatic approaches refine this picture by identifying interpretive conditions: downward-entailing environments, focus-based scalar structures, domain widening plus strengthening, free-choice alternatives, and nonveridicality. The discussion so far has been essentially synchronic. I now turn to how such patterns arise and change over time.

2.4 Diachrony: feature change and grammaticalization

The final strand of the general background concerns diachrony. I start from Jäger's (2010) feature-based model of polarity change and then embed it in a broader view of the diachrony of indefinites based on semantic maps.

Section 2.4.1 presents Jäger's (2010) feature-based approach. In this model, polarity-sensitive items are classified by features such as [affective] and [negative], and diachronic change proceeds by adding or dropping such features in response to skewed input distributions. I adopt this feature vocabulary and its associated *polarity cline* as the polarity-based dimension of the analysis, and reuse it in Chapter 4 to describe the shifts in *nwukwu-* and *amwu-*.

Section 2.4.2 then broadens the view to grammaticalization sources. Typological work has repeatedly noted that interrogative markers, disjunctions, minimisers, and focus particles are recycled as indefinite and polarity markers in many languages (Haspelmath, 1997, 2003). This observation is particularly relevant for Korean: *-nka* and *-na* are historically clausal markers that develop into DP-internal indefinite markers, and *-to* has clear affinities with additive and even-like particles. These pathways will be central to the analysis of *wh-inka* in Chapter 5 and to the comparison of *nwukwu-* and *amwu*-based series in Chapter 6.

To situate these developments in the larger typology of indefinites, I assume a semantic-map approach in the sense of Haspelmath (1997, 2003), enriched with the functional labels and off-map categories proposed in the extended-map work of Aguilar-Guevara et al. (2010) and Degano and Aloni (2022). Throughout this thesis I use their compact label set (e.g. SU, IR, Q, CA, AM, AA, DN) for cross-linguistic comparability.

In sum, Haspelmath's semantic map and Jäger's (2010) polarity cline define a two-dimensional diachronic space in which the Korean series will be located. In later chapters I argue that the historical trajectories of *nwukwu-* and *amwu-* can be seen as different ways of moving along this cline while staying within contiguous regions of the semantic map.

2.4.1 Feature-based polarity paths

To connect the various strands, I adopt Jäger's (2010) account of how polarity behaviour of indefinites can change over time. Her model links two ingredients: (i) lexical underspecification for polarity features and (ii) skewed distributions in the input that lead learners to resolve that underspecification. The result is a usage-based view of polarity: categories such as 'plain indefinite', 'NPI', and 'negative indefinite' are outcomes of diachronic reanalysis driven by frequency patterns.

Jäger formalizes this intuition using two features, [AFFECTIVE] and [NEGATIVE]. Very roughly, [AFFECTIVE] corresponds to items that prefer 'weak' environments such as questions and conditionals, while [NEGATIVE] corresponds to items that prefer explicitly negative environments. An item that is initially underspecified for both features may appear broadly; if its distribution becomes skewed, learners can resolve the underspecification by assigning [AFFECTIVE][3] or [NEGATIVE], turning it into an NPI or a negative indefinite (Hoeksema, 1998, 2010; Jäger, 2010).

A polarity path in Jäger's system can be summarized as in (17).

(17) plain existential (PPI) → non-negative NPI contexts (preferring questions/conditionals) → negative NPI contexts → inherently negative item (n-word, nothing).

Once we overlay this feature system onto the semantic map of indefinites, the polarity cline can be reinterpreted as a particular route through the SU–IR–CA–Q–AA–AM–DN region. Plain existentials that are still PPIs or only weakly polarity-sensitive are at home in SU-type uses (specific unknown) and more generally in SU/IR contexts. The emergence of [AFFECTIVE] corresponds to expansion from SU into IR and CA/Q on the map—that is, preferential occurrence in irrealis, conditional, and interrogative environments. The subsequent strengthening of [NEGATIVE] pushes items further along the map into AA and AM (anti-additive and antimorphic contexts), and finally into DN-type environments (direct negation). In other words, Jäger's sequence 'plain existential → non-negative NPI → negative NPI → n-word' can be seen as a diachronic path that runs continuously across the SU–IR–CA–Q–AA–AM–DN subspace of the indefinite map, without skipping intermediate nodes.

[3]The term 'affective' was coined by Klima (1964) to designate the class of grammatical environments— including negation, questions, and conditionals—that license indefinite pronouns like *any* and other negative polarity items.

Crucially, Jäger's case studies mostly track *single* lexical items as they move along this cline. The Korean system is richer: it contains two indefinite bases, *nwukwu-* and *amwu-*, plus multiple particle series that partly overlap in morphological shape and semantic function. Changes in one item's distribution therefore have consequences for its competitors. In later chapters I use Jäger's feature system together with Degano & Aloni's map labels to describe how *amwu-* narrows towards a strong NPI (moving rightwards into AA/AM/DN) while *nwukwu-* expands beyond interrogative uses (spreading from Q into neighbouring SU/IR/CA cells). Shifts in their relative frequencies across these mapped environments are interpreted as evidence for competition along the same polarity cline.

This diachronic perspective also fits well with Haspelmath's map-based continuity hypothesis: expansions and specializations proceed through *neighbouring* functions on the map rather than by skipping intermediate ones. In the present dissertation, the evidence bears most directly on a usage-based, learner-facing account of how such local movements are realized: forms become reanalyzed as the normal choice for parts of the map because they are repeatedly encountered there. This does not rule out a deeper representational explanation in which adjacent nodes also share underlying featural structure; it only means that the corpus evidence developed here speaks most directly to the distributional mechanism by which adjacency is diachronically implemented.

2.4.2 Grammaticalization: source construction and reanalysis

Interrogative-based indefinites arise along a small set of recurrent grammaticalization paths (Haspelmath, 1997, §6.2). The *dunno*-type is one such path, alongside the *want/pleases* type and the *it may be* type. All three start from biclausal constructions in which a *wh*-clause is embedded under a higher predicate, and all three end up as compact markers attached to interrogative stems. What distinguishes

the *dunno*-type is (i) its epistemic source ('I don't know' or 'who knows?'), and (ii) its strong tendency to specialise for the *specific-unknown* function, in contrast to the free-choice meanings typical of the *want* and *be* types (Haspelmath, 1997, §6.2).

The canonical source construction for *dunno*-indefinites is an indirect *wh*-question embedded under a verb of ignorance, most often 'know' in the scope of sentential negation:

(18) I don't know who/what ⋯

This yields structures of the type in (19):

(19) I don't know [*wh*-X was / did Y].

In discourse, the embedded clause is highly predictable and therefore prone to sluicing-like ellipsis in Ross's sense (Ross, 1969). Haspelmath (1997, pp. 131–133) sketches the following stages; I adopt his terminology but recast them here for later use:

- **Stage 1: full indirect question**
  I don't know what it was.

- **Stage 2: sluicing**
  (*She told him something,*) *I don't know what* (*it was*).

  At this stage, *I don't know* is still a full matrix clause but begins to behave prosodically and informationally like a parenthetical: it carries low stress, can be inserted clause-medially, and contributes primarily the speaker's stance rather than new propositional content.

- **Stage 3: syntactic amalgam**
  She told him I don't know what.

Here the embedded *wh*-phrase simultaneously satisfies the selectional needs of the verb in the main clause (*told*) and of the verb in the parenthetical (*know*). From a structural point of view, this is the 'amalgam' configuration discussed by Lakoff (1974) and Espinal (1991): two clause skeletons sharing a single *wh*-constituent. Such biclausal hybrids are unstable and provide an ideal context for reanalysis.

- **Stage 4: reanalysis as an indefinite marker**
  The sequence 'I-don't-know + *wh*' is reinterpreted as a single NP-internal complex, modifying the interrogative pronoun rather than introducing its own proposition. Negation and the verb 'know' lose their compositional meaning and are recast as a prefixal indefiniteness marker, yielding forms like Old English *náthwā* 'someone', Old High German *neizwer*, Old Church Slavonic *nekůto*, or French *je-ne-sais-quel* 'some kind of N'. Phonological reduction (cliticization, erosion of pronominal and verbal morphology) further strengthens the new morphological bond between the epistemic predicate and the *wh*-root (Haspelmath, 1997, pp. 131–133).

A parallel, more expressive subpath starts from rhetorical questions of the form 'who knows *wh*-' and from invocations such as 'God knows *wh*-' or 'the devil knows *wh*-'. These behave semantically like 'I don't know *wh*-' and can likewise be reduced to frozen indefiniteness markers in some languages (e.g. Lithuanian *kaz-kas* 'someone' < 'who knows who') (Haspelmath, 1997). In most languages, however, such 'God-knows' expressions remain marginal, expressive indefinites and never fully grammaticalise into neutral, all-purpose forms (Haspelmath, 1997).

Outside Europe, Indonesian *entah* 'I don't know / maybe' (in *entah di-mana* 'somewhere', *entah bagaimana* 'somehow') exemplifies a slightly different implementation of the same pathway: here the epistemic element is already a particle

rather than a full verb meaning 'know', but its function as a routinized 'ignorance' marker is analogous to the *dunno*-sources just discussed (Haspelmath, 1997).

Because the source construction literally asserts the speaker's ignorance about a *particular* individual or situation, *dunno*-indefinites naturally specialise for the *specific-unknown* reading on the semantic map (Haspelmath's SU node) (Haspelmath, 1997). Synchronically, they pattern with contexts where:

- there is a concrete, contextually fixed referent, and
- the speaker does not (or presents themself as not) know which one it is.

This explains why *dunno*-based forms are fine in roughly 'I heard something, I don't know what'-type contexts, but infelicitous in canonical non-specific or free-choice environments, where there is no fixed individual that could in principle be known. Haspelmath (1997, pp. 132–133) illustrates this with contrasts like:

(20) a. Susanne is thinking about something. I don't know what. (specific)

b. She wants to marry an Ainu speaker. I don't know whom. (specific)

* Take some apple. I don't know which one. (irrealis, non-specific)

* Did you see anybody? I don't know whom. (question, non-specific)

* You can take any apple. I don't know which one. (free choice)

In the non-specific cases it would be pragmatically infelicitous for the speaker to assert ignorance about the referent, because if the referent is non-specific, nobody could possibly know it in the first place. When we look at the way *dunno*-indefinites are actually used—Old English *náthwā*, Lithuanian *kaz-kas*, French *je-ne-sais-quel*—we indeed find them in specific-unknown contexts: 'some man or other crept inside', 'some kind of vague desire', 'someone came', etc. (Haspelmath, 1997).

Different languages instantiate different points along this path:

- **Weak grammaticalization.** The epistemic verb is still segmentable and optionally present; the construction retains an expressive flavour. Modern French *je ne sais quel N* is transparent and stylistically marked.
- **Strong grammaticalization.** The epistemic element has become a bound morpheme or opaque particle that no longer functions as a verb 'know'. Old Norse *nekkver*, Old Church Slavonic *nekůto*, Bulgarian dialectal *namkoj*, and Lithuanian *kaz-kas* illustrate this end-point: the former matrix clause has been phonologically eroded and now serves purely as an indefinite marker, often across a full series (person, thing, place, etc.).

Even at this stage, traces of the original 'ignorance' meaning often persist: *dunno*-indefinites frequently retain a slightly vague 'some X or other' flavour and are dispreferred where the speaker wants to present the referent as fully identifiable or discourse-prominent.

Comparison with other interrogative-based paths. Seen against other interrogative-based grammaticalization types, the *dunno* path is structurally similar but semantically distinct (Haspelmath, 1997, §6.2).

- '**want/pleases**' **type:** (Latin *quilibet*, Spanish *cualquier*, Russian *kto-ugodno*, Albanian *kushdo*, etc.)

  Source construction: non-specific free relatives like 'take what you want / what pleases you'. Here the embedded clause denotes a *choice set* under a volitional predicate. The construction paraphrases a free-choice statement ('you may take any X'), and accordingly the resulting markers specialise for

free-choice and related non-specific functions. Morphologically, since the volitional predicate sits inside the subordinate clause, it typically surfaces as a suffix or postposed element on the *wh*-word (Haspelmath, 1997).

- '**it may be**' **type:** (Russian *kto-nibud*' < 'whoever it may be', Bulgarian *kojto i da e*, French *qui que ce soit*, etc.)

  Source construction: parametric concessive conditionals ('whoever it may be'), structurally close to unconditional 'no matter *wh*-' clauses. These introduce a set of mutually exclusive alternatives and state that the main proposition holds for each alternative; this universal-over-alternatives structure is a well-known cradle for free-choice meanings (Haspelmath, 1997). Recent diachronic work on Italian *qualsiasi* argues precisely for this kind of path: from unconditional 'whatever X is' to a fully grammaticalized free-choice determiner (Degano & Aloni, 2022).[4]

- '**dunno**' **type:** In contrast, the *dunno* source construction does not quantify over all values of the *wh*-variable; it asserts ignorance about the actual value. Consequently, it does not naturally yield free-choice or generic functions. Instead, it carves out a corner of the indefinite space that the other paths do not directly target: specific-unknown referents whose identity is either genuinely unknown or pragmatically backgrounded (Haspelmath, 1997).

The contrast is summarized in Table 2.1.

This typological perspective helps explain why *dunno*-indefinites are relatively infrequent cross-linguistically, often geographically clustered (well attested in Europe, sparse elsewhere), and semantically 'narrow' compared to the large, multifunctional *any*-like systems that emerge from the other two sources (Haspelmath, 1997). At the same time, in languages where they do grammaticalise strongly,

[4]See Degano and Aloni (2022) for detailed discussion of Italian *qual si sia* and *qualsiasi*.

Table 2.1 Interrogative-based grammaticalization paths for indefinites.

Type	Source meaning	Primary diachronic target
dunno	'I don't know which X'	specific-unknown (SU)
want/pleases	'(You) may take what you want/like'	free choice / irrealis
it may be	'whoever/whatever it may be'	free choice / generic (FC/UFC/GEN)

dunno-forms fill an otherwise under-served semantic niche and can become the default way of signalling that the speaker has a particular but non-identifiable referent in mind.

From a semantic-map perspective (Haspelmath, 1997, 2003), these grammaticalization pathways are not random. Interrogatives first expand into epistemically weak or irrealis uses ('someone or other, I don't know who'), then into polarity-sensitive uses under negation and in if-clauses and modals ('anyone at all'), and finally, in some languages, into fully lexicalized free-choice or negative-indefinite items. Free-choice forms like Spanish *cualquiera* and Polish *ktokolwiek* occupy the 'free choice' and 'irrealis' zones; NPI-like forms like Greek *kanenas* inhabit the 'polarity-sensitive' zone; and simple wh+Q forms such as Japanese *dare-ka* correspond to 'plain existential' or 'specific unknown' regions. The crucial point is that interrogative pronouns, disjunction markers, and focus or additive particles provide the raw material for these shifts. Once they are recruited into the nominal domain and reanalyzed as DP-internal markers, they yield stable paradigms of indefinites whose distributions can be captured using the same semantic and polarity notions introduced earlier in this chapter (Haspelmath, 1997, 2003).

In sum, this typological perspective gives us a concrete set of expectations for Korean. Because -(*i*)*nka* is a clause-typing Q-element, the Korean wh+-(*i*)*nka* series is a natural candidate for an ignorance-based, 'dunno-like' path. If Korean follows

this path, then wh+-(*i*)*nka* should primarily serve specific-unknown, ignorance-marking uses ('someone-or-other') and should not automatically develop the broad free-choice and generic range that is more typical of want/pleases and it-may-be type sources.[5]

2.5 The Korean system and three core asymmetries

The previous sections have introduced the basic tools for thinking about indefinites: their existential core, how they differ in specificity and identifiability, how polarity-sensitive items are licensed in particular environments, and how clause-level markers can grammaticalize into DP-level particles. In this section, I instantiate these tools in Korean and lay out the empirical landscape for the rest of the dissertation. I focus on three core asymmetries that recur throughout the system. Together, these asymmetries locate the Korean paradigm in the two-dimensional space defined by Haspelmath's semantic map of indefinite functions (Haspelmath, 1997, 2003) and Jäger's polarity cline (Jäger, 2010).

2.5.1 Inventory of bases and particles

Korean indefinites are built from two main kinds of bases (Chang, 1996; Sohn, 2001). One is a set of *wh*-based roots such as *nwukwu* 'who', *mwues* 'what', and *e-ti* 'where'. These are used as interrogative pronouns in questions and, in many contexts, also as indefinites (E.-H. Park, 2009; Yun, 2013, 2022). The other is a dedicated indefinite root *amwu-*, roughly 'any (person)' or 'any X', which does not function as a question word in the modern language and appears only in indefinite expressions (An, 2007; J. Choi, 2007; C. Lee, 1996).

[5]Chapter 5 evaluates this expectation directly, using diachronic corpus evidence and semantic-map labels to track where wh+-(*i*)*nka* forms end up in the Korean indefinite system.

These bases combine with a small set of DP-level particles, including *-(i)nka*, *-to*, *-na*, *-tunci*, and *-lato*, to form morphologically compositional indefinite expressions (J. Choi, 2007; Kang, 2014; Lim, 2017; Yun, 2022). Table 2.2 summarises the core series that are productive in contemporary Korean (cf. Yun, 2022).[6]

Table 2.2 Core Korean indefinite series by base and particle.

Base(s)	**Particle**	**Typical gloss**	**Typical environment**
wh-roots	*-(i)nka*	'someone / something or other'	simple declaratives; vague or unknown referent
wh, *amwu-*	*-to*	'no one / nothing'	negation-sensitive ('not anyone', 'not anything')
wh, *amwu-*	*-na*	'anyone / anything'	permission, modals, conditionals (free choice)
wh, *amwu-*	*-tunci*	'whoever / whatever'	concessive 'no matter who/what' uses
wh, *amwu-*	*-lato*	'any X at all'	questions, conditionals, imperatives

Two observations about this paradigm are worth highlighting from the outset. First, the nature of the base matters. In modern Korean, *nwukwu* can be both an interrogative 'who?' in questions and an indefinite 'someone' in declaratives (E.-H. Park, 2009; Yun, 2022), whereas *amwu-* cannot be used as a question word and does not by itself mean 'who?'; it only appears as an indefinite component (An, 2007; J. Choi, 2007; C. Lee, 1996). Historical evidence, discussed in Chapter 4, indicates that Middle Korean *amo* (the ancestor of *amwu-*) did function more broadly as a 'someone'-type expression, but this use has been lost (C.-M. Suh, 1987; J.-M. Yoon, 2005).

Second, many particles form parallel series with both wh-roots and *amwu-*, as in the case of *-to*, *-na*, *-tunci*, and *-lato* (Yun, 2022). This morphological symmetry will

[6]This table abstracts away from number, case, and honorific markers, which do not affect the basic generalizations here.

matter for the competition patterns in the third asymmetry below. In contrast, only wh-bases appear with -(*i*)*nka*; there is no productive form **amwu-inka* alongside *nwukwu-inka* and *mwues-inka* (Y.-J. Choi, 2011; Kang, 2014; Yun, 2022).

What follows highlights three recurring contrasts in the Korean indefinite paradigm. Each contrast compares forms that are close in shape but uneven in use: one member is broadly available and natural, while the other is restricted or unavailable. I illustrate each asymmetry with a small set of representative examples in this section; the later chapters provide fuller documentation and analysis.

Asymmetry 1 compares the two bases in their bare form. Bare *nwukwu* is attested both as a question word and as 'someone' in simple declaratives, whereas bare *amwu* is not used this way and typically requires an additional particle (J. Choi, 2007; C. Lee, 1996; Yun, 2022). The evidence comes from contemporary usage and acceptability contrasts, supplemented in Chapter 4 by historical attestations of Middle Korean *amo* (C.-M. Suh, 1987; J.-M. Yoon, 2005).

Asymmetry 2 concerns the -(*i*)*nka* series. Wh-roots form productive indefinites like *nwukwu-inka*, but there is no parallel **amwu-inka* (Kang, 2014; Yun, 2022). Here I combine modern examples with the historical record for clausal *-nka*, developed in Chapter 5 (Y.-J. Choi, 2011).

Asymmetry 3 turns to pairs where both bases combine with the same particle. I focus on *-to* under negation (*amwu-to* vs. *nwukwu-to*) and *-na* in permissive or modal environments (*amwu-na* vs. *nwukwu-na*) (J. Choi, 2007; Yun, 2022). The later chapters evaluate these contrasts with corpus distributions across environments, alongside well-known minimal contexts reported in the literature (An, 2007; Giannakidou & Yoon, 2016; M. Kim & Kaufmann, 2006).

2.6 Asymmetry 1: Base morphemes *amwu* and *nwukwu*

Asymmetry 1 compares the two bases *amwu-* and *nwukwu-*. The contrast is visible already in their bare forms, and it becomes sharper once the bases combine with particles and occur in environments like negation and intensional predicates. Section 2.6.1 establishes the basic synchronic split with minimal pairs: bare *nwukwu* appears both in questions and in simple declaratives with an indefinite interpretation, whereas bare *amwu* does not show the same distribution and is typically realized in particle-marked expressions (J. Choi, 2007; C. Lee, 1996). Section 2.6.2 summarises the scope- and negation-related diagnostics in S. Yoon (2008). Section 2.6.3 briefly relates these contrasts to the specific/non-specific ambiguity of English *a/an* discussed by Fodor and Sag (1982). Section 2.6.4 adds distributional facts about bare *amwu* in negative constructions from E.-H. Park (2009). Section 2.6.5 turns to earlier Korean and summarises the diachronic background as reported in descriptive and historical work on Korean interrogatives, indefinites, and the pronominal system (A.-R. Kim, 2000; J.-H. Park, 2007; C.-S. Suh, 1989; C.-M. Suh, 1987, 1989; J.-M. Yoon, 2005; Yun, 2022). Finally, Section 2.6.6 states the diachronic problem that motivates the corpus study in Chapter 4.

2.6.1 Bare *amwu* and *nwukwu* in contemporary Korean

In contemporary Korean, *nwukwu* and *amwu-* differ already at the bare level (J. Choi, 2007; C. Lee, 1996). The base *nwukwu* is primarily an interrogative pronoun 'who?', but it also has a well-attested indefinite use in declaratives ('someone') (E.-H. Park, 2009; S. Yoon, 2008). As (21) and (22) show, bare *nwukwu* can occur both in questions and in simple declaratives.

More concretely, *nwukwu* is first and foremost an interrogative pronoun 'who?', but it also has well-attested indefinite uses ('someone, anyone') in the right environments. The minimal contrast between its interrogative and indefinite uses is illustrated in (21) and (22).

(21) Nwukwu-ka w-ass-ni?
who-NOM come-PST-Q
'Who came?' ⇒ bare *nwukwu* as interrogative

(22) Na-nun nwukwu-lul manna-ss-ta
I-TOP who/someone-ACC meet-PST-DECL
'I met someone.' ⇒ bare *nwukwu* as indefinite

In (21), *nwukwu* behaves as an ordinary question word. In (22), with accusative case, it is naturally understood as 'someone', though in the right context it can also be read as an echo-question ('I met who?').

By contrast, *amwu* is not used as a question word at all. It is a dedicated indefinite base glossed 'any (person), any X'. In present-day standard Korean, it cannot freely appear with case markers on its own. The attempted bare uses in (23) and (24) are ungrammatical in the intended 'anyone/someone' sense.

(23) *Amwu-ka w-ass-ni?
amwu-NOM come-PST-Q
intended: 'Did anyone come?' ⇒ ungrammatical bare *amwu* as subject

(24) Na-nun *amwu-lul manna-ss-ta.
I-TOP any-ACC meet-PST-DECL
intended: 'I met some/any person.' ⇒ ungrammatical bare *amwu* as object

In practice, *amwu* surfaces almost only when followed by a particle such as *-to*, *-na*, or *-lato* (*amwu-to*, *amwu-na*, *amwu-lato*, ⋯), and in a few highly constrained bare expressions. Those expressions always have a non-specific flavour: they are about 'any person' rather than 'some particular person', and, as we will see next, they gravitate towards negative or otherwise non-assertive contexts.

Even at this bare level, the split is clear. The form *nwukwu* is a flexible base that can be used both to ask 'who?' and to say 'someone'. The form *amwu*, by contrast, is a specialized indefinite base: it cannot ask 'who?', and it cannot by itself mean 'someone'; it lives in a narrower morphological and semantic niche.

2.6.2 Yoon (2008): scope and polarity of *amwu-* vs. *nwukwu-*

Yoon (2008) sharpens this basic contrast by looking at two issues: which readings the two bases support in intensional and quantificational contexts (scope), and how they behave in negative environments (polarity).

Her first key observation is that, in intensional contexts like 'want', *amwu-* and *nwukwu-* systematically differ. Consider (25) and (26).

(25) Con-un hankwuk-yeca amwu-wa kyelhonha-ko
John-top Korean-woman *amwu*-with marry-comp
sipehan-ta.
want-decl
'John wants to marry any Korean woman.' ⇒ *amwu*-series under *sipeha-* 'want'

[Yoon 2008]

(26) Con-un hankwuk-yeca nwukwu-wa kyelhonha-ko
John-TOP Korean-woman *nwukwu*-with marry-COMP
sipehan-ta.
want-DECL
'John wants to marry a Korean woman.' ⇒ *nwukwu*-series under *sipeha-* 'want'

[Yoon 2008]

In (25), the natural interpretation is that John is not picky: he just wants to marry a Korean woman, and it does not matter which one. In (26), the natural interpretation is that there is a particular Korean woman he wants to marry. Yoon analyses this in terms of scope: *amwu*-based indefinites are forced to stay inside the scope of *want* (non-specific), while *nwukwu*-based indefinites can take wide scope over *want* (specific).

Yoon then shows that the same pattern appears in sentences like 'Every woman talked to a child in fifth grade'. With a *nwukwu*-series object, only a strong, specific reading ('there is some particular fifth-grader that every woman talked to') is available; with an *amwu*-series object, only a weak, non-specific reading ('for each woman, there is some child or other') is available. In effect, Korean splits the work that English does with a single determiner *a/an*: English *a child* can mean either 'some child or other' or 'a particular child', but in Korean the 'some child or other' reading is confined to *amwu*-indefinites, and the 'a particular child' reading is confined to *nwukwu*-indefinites.

The same division of labour appears in polarity behaviour. Under clausal negation, both series can be used, as in (27) and (29), but outside negation their distributions diverge.

(27) Mina-nun amwu-to manna-ci anh-ass-ta

Mina-TOP *amwu*-TO meet-CONN neg-PST-DECL

'Mina didn' t meet anybody.' ⇒ *amwu-to* under sentential negation

[Yoon 2008]

(28) Mina-nun amwu-to manna-ss-ta

Mina-TOP *amwu*-TO meet-PST-DECL

intended: 'Mina met just anybody.' ⇒ ungrammatical *amwu-to* in simple affirmative

[Yoon 2008]

(29) Mina-nun nwukwu-to manna-ci anh-ass-ta

Mina-TOP *nwukwu*-TO meet-CONN neg-PST-DECL

'Mina didn' t meet anybody.' ⇒ *nwukwu-to* under sentential negation

[Yoon 2008]

(30) Mina-nun nwukwu-to manna-ss-ta

Mina-TOP *nwukwu*-TO meet-PST-DECL

intended: 'Mina met just anybody.' ⇒ ungrammatical *nwukwu-to* in simple affirmative

[Yoon 2008]

Both forms are fine under negation but excluded in simple positive sentences. What distinguishes them is what happens in non-negative, nonveridical environments (certain questions, conditionals, modals). Yoon shows that *amwu-to* is essentially restricted to strictly negative contexts and always conveys a very strong 'not even anyone' meaning, while *nwukwu-to* can appear in a somewhat wider range of sensitive environments and still allows readings where *-to* keeps its ordinary 'also/

too' meaning. Taken together, Yoon's scope and polarity data suggest that *amwu-* is inherently non-specific and tightly tied to negative environments, while *nwukwu-* can introduce specific referents and is less tightly bound to negation.

2.6.3 Fodor & Sag (1982) and the English determiner *a/an*

Yoon explicitly relates her Korean contrast to Fodor and Sag's (1982) analysis of the English indefinite determiner *a/an*. Their key point is that a single form *a/an* can support both specific and non-specific readings. The difference comes from how the determiner interacts with other operators, not from there being two different lexical items.

A familiar example is (31).

(31) Every woman talked to a child in fifth grade.

This sentence can mean that for each woman, there is some child or other she talked to (non-specific: $\forall > \exists$), or that there is a particular child such that every woman talked to that one child (specific: $\exists > \forall$). Fodor and Sag argue that we do not need two determiners a_1 and a_2. Instead, a single *a/an* is flexible enough to take wide or narrow scope; the specific vs. non-specific difference is a matter of structure, not of lexical duplication.

In Korean, Yoon's proposal is that this flexibility is lexically split. The work that English does with a single determiner *a/an* is divided between two bases. The *amwu*-series always behaves like the 'some child or other' reading (narrow scope, non-specific), and the *nwukwu*-series always behaves like the 'a particular child' reading (wide scope, specific). That is, English allows *a child* to cover both readings, but Korean distributes them: *amwu* for the weak, non-specific reading and *nwukwu* for the strong, specific reading.

2.6.4 Park (2009): bare *amwu* in negative contexts

Park (2009) looks more closely at where bare *amwu* appears in modern Korean. Her conclusion is simple: all of her clear bare-*amwu* examples occur in negative environments, and they always involve the idea of 'not just anyone' or 'no X at all'.

One context involves lexical negation, as in (32).

(32) Amwu-sayngkak-i eps-ta.
amwu-thought-NOM not.exist-DECL
'I don't have any idea.' ⇒ bare *amwu* under lexical negation
[Park 2009]

Here *amwu-sayngkak* literally means something like 'any-thought'. Combined with *eps-ta* 'there is not', the sentence means 'there is not any thought', i.e. 'I don't have any idea'. A corresponding positive sentence with *amwu-sayngkak-i iss-ta* is not natural in the intended 'I have some kind of idea' reading.

Another context involves the negative copula *ani-*. Park discusses expressions of the form *amwu-ka ani-n X*. A typical example, paraphrased in (33), is used in Christian discourse.

(33) Amwu-ka ani-n hananim-i inchenha-n salam
amwu-NOM be.NEG-ADN God-NOM acknowledge-ADN person
'a person who is not just anyone but is recognized by God' ⇒
bare *amwu* with copular negation
[Park 2009]

In (33) the string *amwu-ka ani-n* introduces the idea 'not just anyone', and the rest of the phrase specifies who this special person is ('a person recognized by God'). Bare *amwu* here provides the indiscriminate baseline 'any person', which is then explicitly denied by *ani-*.

Park's examples all have this shape. Bare *amwu* appears either under a negative verb like *eps-ta* 'there is not' or with the negative copula *ani-*. There are no straightforward positive, neutral sentences with bare *amwu*. This matches Yoon's picture: in the modern language, *amwu-* is not only non-specific but also clearly polarity-sensitive. It shows up precisely where the grammar wants an 'anyone' base in the scope of negation or in a 'not just anyone' contrast.

In sum, the synchronic data from Yoon (2008) and Park (2009) point in the same direction. *Nwukwu-* behaves like a flexible wh/indefinite base that can introduce specific or non-specific individuals and is not intrinsically tied to negative contexts. *Amwu-* behaves like a dedicated, non-specific base whose natural habitat is negative or other nonveridical environments, and whose bare uses are restricted to 'no X at all' or 'not just anyone' readings.

2.6.5 Earlier Korean: *amo* and *nwukwu*

So far, the discussion has been synchronic: we have looked only at modern Korean. A range of descriptive and diachronic work discusses how interrogatives and indefinite forms are realized in earlier Korean and how the pronominal system is organized across periods (A.-R. Kim, 2000; J.-H. Park, 2007; C.-S. Suh, 1989; C.-M. Suh, 1987, 1989; J.-M. Yoon, 2005; Yun, 2022). Against that background, the historical examples below illustrate the core point needed for Asymmetry 1: earlier *amo* (the predecessor of *amwu-*) is attested with ordinary case marking in environments where present-day Korean would normally use *nwukwu* for an indefinite 'someone'.

In the historical materials considered here, *amo* is attested with ordinary case marking in a range of declarative environments, as in (34)–(36). Middle Korean orthography is simplified here, with focus on *amo* and case.

(34) Icyen syeysyang-ey-nun nam-uy mal-i
formerly world-LOC-TOP other-GEN word-NOM
na-lul cimok-ha-mye amo-uy
I-ACC point.out-do-CONN someone-GEN
cason-i-yo amo-uy cichin-i-la
descendant-COP-CONN amo-GEN kin-COP-DECL
syuta kwensyol tali-ko cyuk-ul ...
gossip rumor carry-CONN die-ADN...

'Formerly, in this world, other people' s words singled me out, saying (I was) someone' s descendant and someone' s relative, dragging along gossip and rumours, to the point of dying.' ⇒ *amo-uy* with genitive (historical)

[<toklipsinmwun, 1898>]

(35) Amo-ka pyenghwan-i phi-kyengha-may
amo-NOM illness-NOM not.light-CONC

'someone's illness is not light ⋯' ⇒ Middle Korean *amo* with nominative

(36) Amo-lul uylyon-chi mal-ko
amo-ACC discuss-NEG say-AND

'⋯do not discuss some person ⋯' ⇒ Middle Korean *amo* with accusative

In (34)–(36), *amo* refers to people in context and combines freely with case markers. These are positions where present-day Korean would normally use *nwukwu-* as an indefinite ('someone'), rather than *amwu-*. On the descriptive level needed here, the important point is that earlier *amo* patterns like a general-purpose indefinite

that can appear bare with case, rather than like the restricted particle-dependent *amwu-* of the modern language.

The same diachronic literature also reports that *nwukwu* in earlier stages is primarily interrogative 'who?', and that its indefinite uses are not the default profile associated with the form in present-day Korean (J.-H. Park, 2007; C.-S. Suh, 1989; C.-M. Suh, 1987, 1989; J.-M. Yoon, 2005). Only over time does the *nwukwu*-series expand into the 'someone/anyone' space that earlier *amo* could cover. The historical background, then, is consistent with a reorganization of labor between the two bases: older *amo* is broadly available as 'someone', while *nwukwu* is primarily 'who?'; in the modern language, *amwu-* is restricted and *nwukwu* spans both 'who?' and 'someone'.

2.6.6 Interim summary and a diachronic puzzle

The synchronic work by Yoon (2008) and Park (2009) gives a fairly unified picture of present-day Korean. *Nwukwu-* is a flexible base: it can be used as a question word ('who?'), it can be used as an indefinite ('someone'), it can take wide or narrow scope, and it is not confined to negative sentences. *Amwu-*, by contrast, is much more specialized. It cannot appear as a bare 'who?'; it cannot appear as a simple bare 'someone'; and wherever it does occur, it favours non-specific 'anyone'-type readings and sits very comfortably inside negative or otherwise 'sensitive' environments.

Yoon's examples show that this difference is not just a matter of where the forms can appear, but also of what readings they support. Only *nwukwu*-series indefinites give rise to strong, specific readings in contexts like 'want' and 'every'; *amwu*-series indefinites are confined to weaker, non-specific readings. Park's examples then add that bare *amwu* itself is only attested in clearly negative contexts, either under a negative verb like *eps-ta* 'there is not', or inside a 'not just anyone' construction like

amwu-ka ani-n ⋯. Together, Yoon and Park point to the same conclusion: in modern Korean, *amwu-* is not only non-specific but also strongly polarity-sensitive.

The diachronic literature and the historical examples in Section 2.6.5 suggest that the distribution of the predecessor *amo* was broader than the distribution of present-day *amwu-*, including bare uses with case marking in environments that today would normally favor *nwukwu* for 'someone' (A.-R. Kim, 2000; J.-H. Park, 2007; C.-S. Suh, 1989; C.-M. Suh, 1987, 1989; J.-M. Yoon, 2005; Yun, 2022). At the same time, earlier descriptions treat *nwukwu* primarily as an interrogative 'who?', with its indefinite uses becoming robust only later.

Putting these together, the direction of change can be stated as follows. Earlier Korean shows a broader use of *amo* as 'someone', alongside *nwukwu* as 'who?'. Modern Korean shows a narrowed *amwu-* that is largely particle-supported and polarity-leaning, alongside a broadened *nwukwu* that now serves as both 'who?' and 'someone'.

This gives us a concrete diachronic problem.

- **Puzzle 1** (**Base asymmetry and change**). How did Korean move from a system in which *amo* functioned as a general 'someone' indefinite and *nwukwu* as 'who?', to the modern system in which *amwu* is a non-specific, polarity-bound 'any-root' and *nwukwu* is the main 'who/someone-root'?

2.6.7 Haspelmath's continuity and Jäger's polarity path

To sharpen this puzzle, it is helpful to place the Korean facts in two broader frameworks: Haspelmath's (1997) continuity hypothesis for indefinites and Jäger's (2010) polarity cline, both introduced earlier in this chapter.

Haspelmath proposes that the different uses of indefinite forms across languages—interrogative, 'plain' existential, specific-unknown, modal or conditional, free-choice, under negation, and so on—can be arranged on a single semantic map.

A key idea is continuity: each indefinite form tends to cover one connected region of this map, and historical change tends to move in small steps along adjacent functions. For instance, a form might start as a plain ‘someone’ indefinite, then extend to uses in questions and conditionals, then extend further to ‘anyone under negation’, rather than jumping directly from ‘someone’ to an n-word.

Viewed this way, the diachronic picture summarized above suggests that earlier *amo* occupied a central ‘someone’ region on Haspelmath’s map, whereas modern *amwu-* clusters near the edge where negative and free-choice uses are concentrated. Over the same period, *nwukwu-* expands outward from the interrogative node into the neighboring ‘someone’ area (Haspelmath, 1997, 2003; Jäger, 2010).

Jäger (2010) zooms in on the polarity part of this map and asks how items become more and more tied to contexts like questions, conditionals, and negation. She describes a cline of four stages, repeated in (37).

(37) PPI $\rightarrow$ non-negative NPI $\rightarrow$ negative NPI $\rightarrow$ NCI.

A PPI (positive polarity item) is an ordinary indefinite that is happy in simple positive sentences and does not show any clear preference for questions, conditionals, or negation. A non-negative NPI is an item that starts to prefer ‘weak’ environments such as questions, if-clauses, comparatives, and certain modals, but is still not tied to negation. A negative NPI is an item that now prefers genuinely negative environments: it sounds most natural with clausal negation or strongly negative predicates, though it may still occasionally appear in questions or conditionals. An NCI (negative concord item, often called an n-word) is an item that has effectively become inherently negative: it can mean ‘no one, nothing’ on its own, and when it co-occurs with a clausal negator the sentence still expresses only a single negation.

If we place Korean *amo*/*amwu* on this cline, a plausible story emerges. Earlier *amo* patterns like a broadly available 'someone' form, while modern *amwu-* is heavily concentrated in negative and other polarity-sensitive environments. Yoon's and Park's synchronic diagnostics show that *amwu-* is most at home in negative contexts and in closely related 'sensitive' environments, while *nwukwu-* remains comparatively flexible (E.-H. Park, 2009; S. Yoon, 2008). This distribution is compatible with a rightward movement on Jäger's cline for *amwu-*, from a broad existential profile toward the negative end.

On this view, the Korean development for *amwu-* can be sketched as in (38).

(38) *amo* (PPI) → *amwu* in question/conditional/permission contexts → *amwu* in strongly negative contexts (and partly NCI-like).

Stage I is earlier *amo* as a general 'someone' indefinite with no strong polarity bias. Stage II would be a period in which *amo*/*amwu* starts to appear more often in questions and if-clauses (and possibly other non-negative 'affective' environments), moving into the non-negative NPI region on Jäger's cline and into the neighboring polarity or conditional area on Haspelmath's map.

At present, Stage II remains a hypothesis that needs corpus confirmation. The synchronic descriptions (Yoon 2008; Park 2009) and the diachronic or descriptive literature on earlier Korean establish the endpoints, while Jäger's cline tells us what an intermediate Stage II should look like in terms of questions, conditionals, and negation (Jäger, 2010; A.-R. Kim, 2000; J.-H. Park, 2007; C.-S. Suh, 1989; C.-M. Suh, 1987, 1989; J.-M. Yoon, 2005). Haspelmath's continuity hypothesis and Jäger's polarity path together provide a way of phrasing the core diachronic question for *amwu-*: did Korean move from a broad 'someone' indefinite towards an 'anyone under negation' item by local shifts in questions, conditionals, and negatives, or does the Korean pattern require additional intermediate steps?

Chapter 4 will address this question empirically, by placing attested tokens of *amo*/*amwu* and *nwukwu* on Haspelmath's map and classifying their polarity environments along Jäger's cline.

2.7 Asymmetry 2: Korean -(*i*)*nka* indefinites and two puzzles

Asymmetry 2 concerns a systematic contrast in the particle domain. Korean -(*i*)*nka* combines productively with *wh*-roots to form epistemic indefinites such as *nwukwu-inka* 'someone or other', but there is essentially no corresponding series **amwu-inka*. This section provides the descriptive and analytical background on -(*i*)*nka* that will be used in the later chapter devoted to Asymmetry 2.

The discussion brings together three strands of evidence. Section 2.7.1 introduces the basic distributional profile of *wh-inka* forms in contemporary Korean and highlights their ordinary DP behaviour. Section 2.7.2 summarises the characteristic ignorance or non-identification effects associated with *wh-inka*, which motivate the description of these items as anti-specific. Section 2.7.3 reviews existing semantic work that treats -(*i*)*nka* as an epistemic indefinite marker rather than as a polarity item (Kang, 2014).

Sections 2.7.4–2.7.5 then review proposals that relate nominal *wh-inka* to clausal uses of *-nka*. A recurrent idea in the literature is that wh+Q strings can arise from clausal configurations, often discussed under the Unified Clausal Approach, and that ellipsis and reanalysis in such environments can yield DP-like expressions (Jang, 1999; A.-R. Kim, 2000; C.-M. Suh, 1987; J.-M. Yoon, 2005). More recent work discusses a grammaticalization trajectory specifically for referentially vague *wh-inka* (Kang, 2014, 2021).

On this basis, Section 2.7.6 formulates two puzzles. Puzzle 2.1 asks how well a clause-to-DP trajectory for *wh-inka* is supported by diachronic evidence and how

it fits with continuity expectations on the semantic map (Haspelmath, 1997). Puzzle 2.2 asks why the -(*i*)*nka* series is confined to *wh*-bases and does not extend to *amwu-*, even though other particles combine freely with both base types.

2.7.1 *wh-inka* in contemporary Korean

The most discussed member of the *wh-inka* series is *nwukwu-inka*, built from the *wh*-word *nwukwu* 'who' and the particle -(*i*)*nka*. Its basic use is illustrated in (39).

(39) Nwukwu-inka-ka w-ass-ta.
who-INKA-NOM come-PST-DECL
'Someone or other came.' ⇒ epistemically vague indefinite

The sentence in (39) asserts that there exists some person who came, and it additionally signals that the referent is not identified in the current discourse. In ordinary use, the speaker presents the individual as not being specified, and typically not being promoted as a stable discourse topic, yielding the familiar 'someone-or-other' effect (Kang, 2014).

Distributionally, *nwukwu-inka* behaves like an ordinary noun phrase. It can appear as subject, object, or in oblique positions, and it can occur in both main and embedded clauses. Crucially, it is acceptable in simple affirmative sentences and does not require negation or any special licensing environment. It is therefore not an NPI in the usual sense.

A second basic property is that *nwukwu-inka* is not interrogative. Bare *nwukwu* can be used as a question word, as in (40), but once -(*i*)*nka* attaches, the *wh*-form is interpreted only as an indefinite.

(40) Nwukwu-ka w-ass-ni?
who-NOM come-PST-Q
'Who came?' ⇒ bare *nwukwu* as interrogative

In sum, *nwukwu-inka* is a *wh*-based indefinite that is morphologically transparent, distributionally unrestricted by polarity, and interpretively associated with epistemic vagueness about the referent.

2.7.2 Ignorance and anti-specificity

Beyond its basic existential reading, *nwukwu-inka* systematically conveys ignorance or non-identifiability on the part of the speaker. A speaker who chooses *nwukwu-inka* signals that they do not, or will not, identify the referent (Kang, 2014). This is clear in dialogues such as (41)–(42), where a *wh-inka* statement is followed by a request for identification.

(41) Nwukwu-inka cenhwa-ha-ess-ta.
who-INKA phone-do-PST-DECL
'Someone or other called.' ⇒ *wh-inka* statement

(42) Nwukwu-i-ess-ni?
who-COP-PST-Q
'Who was it?' ⇒ follow-up question (infelicitous)

The follow-up question in (42) is typically judged infelicitous as an immediate response to (41). The first utterance already presents the caller as not being identified; asking 'who exactly?' clashes with that stance. Similar clashes arise when *nwukwu-inka* is followed by naming, pointing, or detailed description, as in the English continuations in (43) and (44), modelled on *some X or other*.

(43) I want to meet some professor or other. # His name is John.

(44) I want to meet some professor or other. # It' s that guy over there.

Under intensional predicates such as 'want', *nwukwu-inka* behaves in line with this anti-specific character. Compared with bare indefinites that permit both specific and non-specific readings, *nwukwu-inka* strongly favours the non-specific reading and resists interpretations in which a particular, discourse-salient individual is introduced as the main referent (Kang, 2014).

2.7.3 Existing analyses: epistemic indefinite, not an NPI

Existing analyses of *wh-inka* converge on the idea that -(*i*)*nka* is an epistemic indefinite marker. The *wh*-root contributes a set of candidates, and -(*i*)*nka* yields an existential claim while adding a speaker-oriented condition of non-identification (ignorance or non-disclosure) (Kang, 2014).

Two points are important for the remainder of this chapter. First, -(*i*)*nka* is not a polarity item: it is compatible with simple affirmatives and does not require negation or any special licensor. Second, *wh-inka* items are not used as free-choice items or as NPIs of the *amwu-* type; instead, they cluster in the 'specific unknown / someone-or-other' region of the indefinite space (Haspelmath, 1997).

2.7.4 From interrogative clause to existential *nwukwu-nka*

A long-standing idea in the Korean literature is that non-interrogative wh+Q strings such as *nwukwu-nka* can be related to clausal wh+*-nka* configurations. In approaches often grouped under the Unified Clausal Approach, Q-morphemes are treated as clausal elements (e.g. complementisers), and apparent nominal wh+QPs are analyzed as arising from reduced clause structures and ellipsis in discourse (Jang, 1999; A.-R. Kim, 2000; C.-M. Suh, 1987). Diachronic discussion of changes in wh-constructions provides additional motivation for connecting earlier clausal uses of *-nka* to later non-interrogative distributions (J.-M. Yoon, 2005).

One recurrent clausal analysis derives existential *nwukwu-nka* from a structure that includes an implicit head noun such as *salam* ‘person’ and a covert ignorance predicate (‘not know’), yielding an interpretation close to ‘a person whose identity is not known’. A simplified surface example is given in (45), and a schematic representation of the older clausal source is given in Figure 2.3.

(45) Nwukwu-nka w-ass-ta
who-Q come-PST-DECL
‘Somebody came.’ ⇒ ≈ ‘A person whose identity is not known came.’
[clausal-source approach]

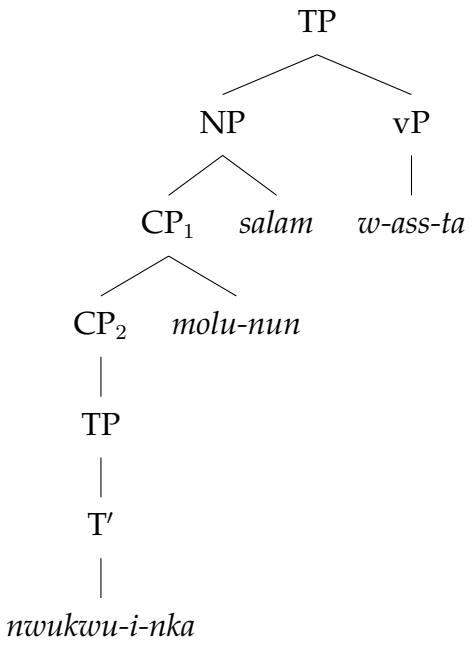


Figure 2.3 Stage 1: relative-clause style source for *nwukwu-nka* ‘someone’.

A second recurrent idea derives existential *nwukwu-nka* from concessive constructions built on wh-copular questions (roughly, ‘though I don’t know who he is ⋯’), as in (46). In such environments, ellipsis of the higher clausal material can

leave the wh+-*nka* string in a position where it can be reanalyzed as a nominal expression (Kang, 2021; J.-M. Yoon, 2005).

(46) e e_i nwukwu-i-nka mol-ato
pro pro_i who-COP-Q not.know-CONC
e_i w-ass-ta
pro_i come-PST-DECL
'Though I don't know who he is, he came.' ⇒ concessive clausal source

A schematic structure for (46) is given in Figure 2.4.

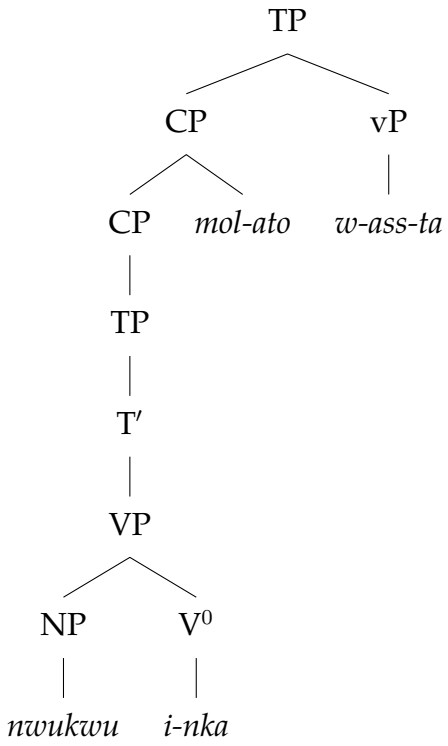


Figure 2.4 Stage 2: concessive CP source for *nwukwu-nka*.

A similar perspective is applied to PP-internal wh+Q strings such as *nwukwu-eykey-nka* in (47), where the wh+Q sequence is embedded in a concessive clause and may surface as a residual string after ellipsis.

(47) ku-nun chayk-ul e_i nwukwu-eykey-i-nka
he-TOP book-ACC pro who-DAT-COP-Q
mol-ato e_i cwu-ess-ta
not.know-CONC pro_i give-PST-DECL
'He gave a book (to someone), though he doesn't know who it is.' ⇒ concessive clausal source

2.7.5 Clause–to–DP reanalysis: Q as delimiter

A natural next step in the clausal-source perspective is reanalysis: wh+Q strings that repeatedly surface in nominal positions can come to be treated as DP-internal expressions in the synchronic grammar. For expository convenience, I represent this DP-internal function of *-nka* using a *Del* head ('delimiter') that attaches to a nominal projection. This captures the surface DP behaviour noted in Section 2.7.1 while maintaining a link to clausal *-nka* sources discussed in Section 2.7.4 (Kang, 2021).

On this view, the existential in (45) can be represented as a DP whose nominal core is *nwukwu* and whose edge hosts a DP-internal *-nka*, as in Figure 2.5. The PP case is represented as a PP containing a nominal complement with *-nka*, as in Figure 2.6.

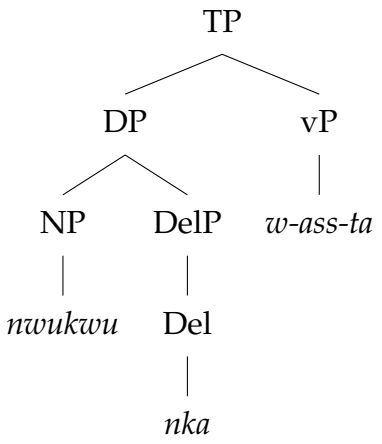


Figure 2.5 Stage 3: DP with delimiter *-nka*.

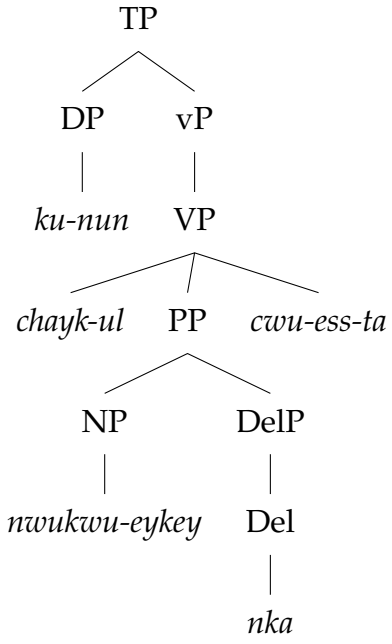


Figure 2.6 Stage 3: PP with delimiter *-nka*.

For present purposes, the crucial point is that the proposed source configurations are wh-based: they are built on wh-copular questions and their embedded or concessive variants. This matters because the dedicated indefinite base *amwu-* does not appear in those wh-question templates.

2.7.6 Interim summary and two puzzles

Up to this point, three descriptive generalizations are in place.

First, in present-day Korean, *wh-inka* forms such as *nwukwu-inka* behave like ordinary noun phrases. They occur in argument positions, in main and embedded clauses, and they are acceptable in simple affirmative sentences. Their characteristic effect is non-identification: the speaker presents the referent as not being specified in the discourse (Kang, 2014).

Second, existing analyses treat -(*i*)*nka* as an epistemic indefinite marker, not as a polarity item. The wh-part contributes alternatives (candidate individuals), and -(*i*)*nka* yields an existential claim while adding an ignorance or non-disclosure component (Kang, 2014).

Third, several lines of work relate nominal *wh-inka* to clausal uses of *-nka*. Earlier discussions of changes in wh-constructions provide a diachronic backdrop (J.-M. Yoon, 2005), and recent work makes the grammaticalization question explicit for referentially vague *wh-inka* (Kang, 2021). A recurring structural feature of these proposals is that the source configurations are wh-based.

These points lead directly to two puzzles.

2.7.7 Puzzle 2.1: a clausal path and continuity

Puzzle 2.1 concerns the diachronic path to *wh-inka*. If nominal *wh-inka* is historically related to clausal *-nka*, then the historical record should show bridging contexts where wh+*-nka* occurs in embedded or concessive environments that plausibly support reanalysis as a DP-like expression (Kang, 2021; J.-M. Yoon, 2005).

This expectation can be stated in semantic-map terms using Haspelmath (1997). If change proceeds stepwise through adjacent functions, then a wh-based series should support a path of the following sort:

(48) interrogative ('who?') → embedded or concessive ('though I don't know who ⋯') → specific unknown ('someone or other').

Puzzle 2.1 is therefore an empirical question about the distribution of wh+-*nka* across historical environments: do attested uses support a clear sequence from clausal questions, to embedded or concessive bridging contexts, to DP uses? Later chapters return to this question with diachronic evidence designed to distinguish clausal uses, bridging contexts, and DP uses.

2.7.8 Puzzle 2.2: the missing *amwu-inka*

Puzzle 2.2 concerns the present paradigm. In modern Korean, -(*i*)*nka* combines productively with wh-roots (e.g. *nwukwu-inka*, *mwues-inka*), but there is no productive **amwu-inka* series. This is striking because other particles combine freely with both base types: we find *nwukwu-na* / *amwu-na* and *nwukwu-tunci* / *amwu-tunci*, but not *amwu-inka*.

One likely ingredient is structural. The clausal sources that plausibly feed -(*i*)*nka* are wh-based (wh-copular questions and their embedded or concessive variants). Since *amwu-* is not a wh-question word, it does not occur in these templates, and it therefore lacks an obvious clausal configuration that could be reanalyzed as a DP-level *amwu-inka*.

A second ingredient is paradigmatic, but the issue is not overlap by itself. Transitional overlap is common in the history of indefinite systems. The more precise question is whether a new *amwu-inka* series would have both (i) a plausible source configuration and (ii) some stable functional contribution distinct enough to survive within a paradigm already supplied by *wh-inka* on one side and the *amwu-* polarity and free-choice series on the other. In this sense, the problem is not that overlap would be impossible, but that it is not obvious what durable niche a new *amwu-inka* series would fill.

Puzzle 2.2 is to make these intuitions precise. The later chapter on Asymmetry 2 addresses this by combining the synchronic distribution of *wh-inka* with diachronic evidence about the environments in which *-nka* occurs and the contexts that plausibly support reanalysis (Y.-J. Choi, 2011; Kang, 2014, 2021).

Chapter 5 explores these questions in detail.

2.8 Asymmetry 3: Synchronic variation and competition

Asymmetries 1 and 2 focused on contrasts where either the base varies with a fixed particle (*amwu-* vs. *nwukwu-*) or the particle varies with a fixed *wh*-base (*-(i)nka* vs. other Q-morphemes). Asymmetry 3 turns to the configuration in which both bases are available with the *same* particle in the modern language. These shared-particle environments are where *amwu-* and *nwukwu-* compete most directly, and where differences in licensing and interpretive flavour are most transparent.

This section is a targeted synchronic literature review. Its goal is to establish a baseline description of the two best-studied shared-particle pairs that the diachronic analysis in Chapter 6 will later build on. I concentrate on (i) the *-to* pair, which is discussed in the Korean literature as a strong polarity domain (An, 2007; J. Choi, 2007; Chung, 2012; C. Lee, 1996), and (ii) the *-na* pair, which is discussed as a free-choice domain with systematic subtype differences in domain restriction and indiscriminacy (J. Choi, 2007; M. Kim & Kaufmann, 2006; M. Kim, 2015).

2.8.1 The *-to* series: *amwu-to* and *nwukwu-to*

Asymmetry 3 refines the base contrast by examining what happens when *amwu-* and *nwukwu-* combine with the same particle *-to*. On the surface, *amwu-to* and *nwukwu-to* appear to be close competitors: under sentential negation, both are natural and yield a 'no one' interpretation (J. Choi, 2007; C. Lee, 1996), as in (49)–(50).

(49) Mina-nun amwu-to manna-ci
Mina-TOP any-TO meet-CONN
anh-ass-ta
neg-PST-DECL
'Mina didn't meet anybody (at all).' ⇒ *amwu-to* under sentential negation

(50) Mina-nun nwukwu-to manna-ci
Mina-TOP who-TO meet-CONN
anh-ass-ta
neg-PST-DECL
'Mina didn't meet anybody.' ⇒ *nwukwu-to* under sentential negation

In simple negative sentences, both forms are acceptable and clearly tied to clause-final negation *anh-* 'not'. For this reason, earlier discussion often treated them as members of the same class of strong NPIs. Once we move beyond simple negation, however, the two series diverge in systematic ways that are repeatedly noted in the Korean literature (An, 2007; J. Choi, 2007; Chung, 2012).

Strength and flexibility

For *amwu-to*, three generalizations are widely reported. First, it is excluded from plain affirmative clauses:

(51) *Amwu-to w-ass-ta
amwu-TO come-PST-DECL
intended: 'Someone came.' ⇒ ungrammatical *amwu-to* in simple affirmative

Second, it is severely degraded in weak, non-negative environments such as questions and conditionals (J. Choi, 2007; C. Lee, 1996):

(52) *Amwu-to w-ass-ni?

amwu-TO come-PST-Q

intended: 'Did anyone come?' ⇒ ungrammatical *amwu-to* in questions

(53) *Amwu-to o-myen ⋯

amwu-TO come-COND ⋯

intended: 'If anyone comes, ⋯' ⇒ ungrammatical *amwu-to* in conditionals

Third, *amwu-to* is strongly concentrated under clausemate negation or under clearly negative predicates and is commonly described as yielding a strengthened inference close to 'not even anyone' (An, 2007).

For *nwukwu-to*, the distribution is broader. It is fully grammatical under clausal negation, and many speakers judge it more acceptable than *amwu-to* in certain weak environments (J. Choi, 2007). The contrast in (54) is representative:

(54) ?Nwukwu-to w-ass-ni?

who-TO come-PST-Q

'Did anyone (at all) come?' ⇒ *nwukwu-to* in questions

Speakers also allow *nwukwu-to* in some conditionals and modal contexts, and it can retain an additive interpretation of *-to* ('someone too / as well') alongside scalar uses in the right discourse conditions. The emerging synchronic generalization is that *amwu-to* is closer to a negative-only item, while *nwukwu-to* remains polarity-sensitive but less restricted (J. Choi, 2007).

Polarity status: NPI vs. NCI-like behaviour

Recent work sharpens the contrast by applying NPI vs. negative-concord diagnostics directly to Korean *-to* items (Chung, 2012; J.-B. Kim, 2024). On these diagnostics (e.g. fragment answers, interaction with degree modifiers, availability in non-negative questions or conditionals, and locality), *amwu-N-to* shows more negative-concord-like behaviour than *nwukwu-to* (J.-B. Kim, 2024). The *nwukwu-to* series still shows strong pressure toward clausemate negation, but it does not align with the NCI-like pattern as uniformly.

Scalar account of *amwu-to*

A prominent Korean analysis derives the exceptional strength of *amwu-to* from a scalar configuration in which *amwu-* contributes a minimal quantity and *-to* introduces a scalar inference that is satisfied most naturally under negation (An, 2007). On this view, affirmative clauses produce a clash, while negated clauses yield a stable 'not even one' inference, which explains the strong concentration of *amwu-to* in direct-negation environments.

Interim summary: competition in the *-to* series

The *-to* series therefore shows a division of labour rather than free alternation. Both *amwu-to* and *nwukwu-to* are negative-related, but they specialize differently. The *amwu-to* series is concentrated in direct negation and tends toward a strengthened 'not even anyone' interpretation (An, 2007; J.-B. Kim, 2024). The *nwukwu-to* series overlaps with direct negation but is more permissive at the edges of weaker environments and can preserve non-scalar readings of *-to* more readily (J. Choi, 2007).

2.8.2 The *-na* series: *nwukwu-na* and *amwu-na*

The second half of Asymmetry 3 concerns competition in the free-choice domain. Both *nwukwu-na* and *amwu-na* are standardly analyzed as free-choice items, but the Korean literature argues that they realise distinct subtypes of free choice. In particular, the two series differ in how the relevant domain is understood and in whether an indiscriminative inference is favoured (J. Choi, 2007; M. Kim & Kaufmann, 2006; M. Kim, 2015).

Basic contrasts

A standard starting point is that both forms are natural in modal, conditional, and generic environments and support an 'anyone / no matter who' reading. A simple modal example is given in (55)–(56).

(55) I kongyen-un nwukwu-na po-l.swu.iss-ta.
this performance-TOP who-NA see-can-DECL
'Anyone can see this performance.' ⇒ *nwukwu-na* in a modal context

(56) I kongyen-un amwu-na po-l.swu.iss-ta.
this performance-TOP any-NA see-can-DECL
'Just anyone can see this performance.' ⇒ *amwu-na* in a modal context

The contrast lies in the preferred flavour. *nwukwu-na* is commonly understood relative to a salient, contextually restricted set, while *amwu-na* more readily supports an indiscriminative 'just anyone at all' interpretation (M. Kim & Kaufmann, 2006; M. Kim, 2015).

Distributional tendencies

A recurring observation is that *amwu-na* is most natural in intensional environments (modals, conditionals, directive frames) and is degraded in simple episodic assertions, especially in subject position. The contrast is illustrated in (57), and it is consistent with analyses that connect Korean *-na* to choice-related and evaluative uses (J. Choi, 2007; M. Kim, 2015).

(57) Nwukwu-na Seoul-tay-ey iphakha-yss-ta
who-NA Seoul-university-at enter.SCH-PST-DECL
*Amwu-na Seoul-tay-ey iphakha-yss-ta
intended: amwu-NA Seoul-university-at enter.SCH-PST-DECL
'Anyone got into SNU.' / intended: 'Just anyone got into SNU.'
⇒ *nwukwu-na* vs. *amwu-na* as subjects

Indiscriminacy vs. domain restriction

The best-known articulation of the difference comes from M. Kim and Kaufmann (2006), who argue that *amwu-na* more readily supports an indiscriminative inference (including below-norm candidates), whereas *nwukwu-na* favours quantification over a salient, norm-based domain. This aligns with later discussion of the semantic–pragmatic connections of *-na* in Korean (M. Kim, 2015).

Interim summary: competition in the *-na* series

The synchronic division of labour can be stated as follows. *nwukwu-na* behaves as a free-choice item that is comparatively compatible with domain restriction and with a wider range of clausal environments, including episodic assertions. *amwu-na* behaves as a free-choice item that more strongly prefers intensional environments and more readily supports an indiscriminative 'just anyone' interpretation (J. Choi, 2007; M. Kim & Kaufmann, 2006; M. Kim, 2015).

In terms of the semantic-map and polarity-cline perspective adopted in this dissertation, the two *-na* series therefore occupy the same broad free-choice region but differ in where they sit within it: *nwukwu-na* is closer to a domain-restricted free-choice use, while *amwu-na* is closer to an indiscriminative, often evaluative free-choice use. Chapter 6 returns to these synchronic facts and asks how this division of labour emerges diachronically.

2.9 Chapter summary

This chapter has developed the theoretical and descriptive background for the present study. I first surveyed how indefinites are modelled in formal semantics and how they are organized typologically. Indefinites were shown to involve more than plain existential force: they may support specific vs. non-specific readings, scope variability, and discourse-linking. On the typological side, Haspelmath (1997)'s semantic map—together with later extensions designed for finer-grained empirical work (Aguilar-Guevara et al., 2010; Degano & Aloni, 2022)—provided a structured space of indefinite functions (e.g. specific, irrealis, interrogative, negative, free choice, indiscriminative) in which morphological series tend to occupy contiguous regions.

The second strand of the chapter addressed polarity phenomena and diachrony. I reviewed syntactic and semantic approaches to NPIs, PPIs, and FCIs, and I introduced Jäger (2010)'s feature-based model of polarity change, in which items may shift from polarity-neutral existential uses toward increasingly restricted polarity-sensitive distributions and, in some cases, toward negative-indefinite behaviour. Combined with the semantic-map perspective and with standard grammaticalization pathways from interrogative and concessive sources, this yields a two-dimensional framework for reasoning about how indefinites change over time.

The third strand applied this toolkit to Korean and extracted three core asymmetries that structure the paradigm. Asymmetry 1 contrasted the bases *nwukwu-* and *amo-*/*amwu-*, highlighting a diachronic split in which a historically broader 'someone' root (*amo*) narrows into a polarity-leaning 'any' base (*amwu-*), while *nwukwu-* extends beyond interrogative use into the central existential domain. Asymmetry 2 focused on the *wh-inka* series and the absence of **amwu-inka*, linking this gap to a wh-based clause→DP grammaticalization pathway. Asymmetry 3 examined competition in shared particle series (*-to*, *-na*) and showed that *amwu-* and *nwukwu-* do not simply alternate: they divide the negative and free-choice sectors of the extended functional space.

The larger theoretical point is that semantic maps and polarity clines are not merely descriptive backdrops. They provide a constrained way of stating which changes and specializations are plausible, even though they do not by themselves determine timing or outcomes. Chapter 3 turns to the empirical core of the dissertation. Building on the asymmetries and the puzzles developed here, it formulates explicit hypotheses about how Korean indefinites move in semantic and polarity space, and it introduces the corpus data, annotation scheme, and statistical models used to test those hypotheses.

Chapter 3

Present Study

This chapter turns the theoretical claims of the dissertation into a concrete empirical research design. The central goal is to show how its three working hypotheses can be tested with diachronic corpus data, semantic-map labels, and transparent quantitative summaries of change.

The dissertation' s core claim is that the modern Korean indefinite system is best understood as a paradigm whose members compete for regions of a structured semantic space (Haspelmath, 1997, 2003; Kroch, 1989; Yang, 2002). On this view, change is not simply a matter of individual forms acquiring new meanings in isolation. Rather, *amwu-* and *nwukwu-*, in combination with particles such as *-to*, *-na*, and -(*i*)*nka*, form a system whose members shift, specialize, and compete across neighboring sectors of a semantic map. The map is therefore used here as a *constrained hypothesis space*: it predicts likely local pathways of extension and retreat, but it does not by itself determine timing, frequency, or which competitor ultimately wins a given niche.

Within this framework, I adopt an underspecification view of polarity change, in the spirit of Jäger (2010) for NPIs, and extend it to the broader domain of indefinites. Polarity values are not taken to be fixed in advance; instead, they can be strengthened or weakened over time, and skewed frequency in particular environments is assumed to be a crucial driver of such shifts (Bybee, 2010; Labov, 1994; Yang, 2002). Using corpus data and a usage-based perspective, I pursue this claim along three empirical dimensions:

- the semantic change of *nwukwu-*, as it extends from 'who?' into 'someone/anyone' and comes to overlap with *amwu-* on the semantic map;
- the grammaticalization of *-(i)nka* from a clausal Q-particle to a DP-internal epistemic marker, with *wh-inka* forms but no **amwu-inka*;
- paradigm-internal competition in shared particle series such as *amwu-to*/*nwukwu-to* and *amwu-na*/*nwukwu-na*.

The introduction (Section 1.4) states the three working hypotheses (H1–H3) that follow from the asymmetries in the modern paradigm. The purpose of the present chapter is methodological: it specifies how those hypotheses are translated into observable predictions and how they are evaluated with annotated corpus data.

The chapter proceeds as follows. Section 3.1 turns H1–H3 into explicit, measurable predictions and links each prediction to the relevant dataset and analysis. Section 3.2 then describes the corpora, retrieval procedure, screening, and semantic-map labelling scheme. Section 3.3 outlines the descriptive statistical tools used to quantify trajectories, change points, and competition across series (Baayen, 2008; Bates et al., 2015; Muggeo, 2003, 2008).

3.1 Hypotheses and operational predictions

This section does not restate the motivations for H1–H3, which are developed in Chapter 1. Its purpose is to translate each hypothesis into observable predictions in the annotated corpus and to specify where those predictions are evaluated in the dissertation. The predictions are stated using the semantic-map labels and, where useful, the broader macro-environment groupings introduced in Sections 3.2 and 3.3.

H1 (**base trajectories**). If *amwu* narrows toward polarity-sensitive territory while *nwukwu* expands beyond interrogative use, then the diachronic distributions of the two roots should diverge systematically.

For *amwu*, the relative share of negative-polarity functions (DN/AA/AM) is expected to rise over time, with a corresponding reduction of non-negative existential and neutral uses. In particular, bare *amwu* should become increasingly concentrated in the negative sector of the map rather than remaining broadly distributed over positive and neutral nodes.

For *nwukwu*, the proportion of *wh*-question uses (Q_WH) is expected to decline as non-question uses increase, especially in adjacent existential and neutral regions of the map. The relevant increase is not expected to be random: new uses should cluster first in neighboring functions such as SU and IR, not jump directly into strongly negative territory. These predictions are tested in Chapter 4.

H2 (**-(*i*)*nka* path and the missing series**). If *wh*+-(*i*)*nka* develops via a clause-to-DP pathway and stabilizes as an epistemic or existential marker, then the corpus should show a staged development.

Earlier *wh*+-(*i*)*nka* tokens are expected to be predominantly clausal or question-like. Later tokens are expected to include DP uses, especially in the specific-unknown and nearby irrealis region of the map (e.g. SK/SU/IR/CA), but the series should not develop a robust NPI or free-choice profile. In addition, the corpus should show no stable, productive **amwu-inka* series beyond accidental or marginal attestations, if any. These predictions are evaluated in Chapter 5.

H3 (**competition in shared series**). If competition partitions the map when both bases occur with the same particle, then the *-to* and -(*i*)*na* series should exhibit consistent specialization rather than free alternation (Kroch, 1989; Yang, 2002).

For *-to*, *amwu-to* is expected to cluster more tightly in the strongest negative functions (DN/AA/AM) than *nwukwu-to*, which should show relatively more weight in nearby non-core or weaker regions, especially at the edge between the NPI sector and neighboring comparative or free-choice territory.

For *-(i)na*, *amwu-na* is expected to concentrate in indiscriminative free-choice uses (especially FC/IND), while *nwukwu-na* is expected to concentrate in more extensional free-choice and generic uses (FC/GEN/UFC). These predictions are tested in Chapter 6.

3.2 Data, labelling, and methods overview

To test these hypotheses, the study combines a diachronic corpus analysis with a semantic-map based labelling scheme and transparent descriptive modelling (Baayen, 2008; Bates et al., 2015; Haspelmath, 1997, 2003). Using the Korean Historical Corpus (History) and the 20th Century Korean Corpus (Yonsei)[1], I extract tokens of the relevant series: bare *nwukwu-* and *amo-/amwu-*, *wh-inka*, *amwu-to/nwukwu-to*, *amwu-na/nwukwu-na*, and related forms.

Each retained token is then assigned a single functional label from the set SK, SU, IR, Q, CA, CO, DN, AA, AM, FC, GEN, UFC, IND introduced in Section 2.1.2. The labelling follows a decision-tree procedure that encodes tests for specificity, force, clause type, and type of negative or modal environment (Figure A.1); the full procedure, with diagnostics and example sentences, is given in Appendix A. The result is a diachronic dataset in which each token is characterized by its root (*nwukwu-* vs. *amo-/amwu-*), its morphological series (bare, *wh-inka*, *-to*, *-na*, etc.), and its functional label on the semantic map (Haspelmath, 1997, 2003).

Subsequent chapters then track how the distribution of these functions changes over time and across series, and how those changes bear on Asymmetries 1–3 and

[1] Available at https://ilis.yonsei.ac.kr/corpus/.

Puzzles 1–3. For some analyses, the fine-grained labels are further collapsed into coarser macro-environments (Q_WH, PPI, NEUTRAL, NPI, FREE-CHOICE), but the core empirical work is always grounded in the full SK/SU/IR/Q/CA/CO/DN/AA/AM/FC/GEN/UFC/IND inventory.

The rest of this section describes the corpora and the workflow in more detail. Section 3.3 then summarizes the statistical tools used to quantify change and competition.

3.2.1 CORPUS DESCRIPTION AND DATA COLLECTION

The empirical base consists of two complementary corpora of historical and modern Korean, covering roughly three centuries (1700–1990): the MOTU History corpus (HISTORY, 1700–1890) and the Yonsei corpus (YONSEI, 1900–1990). Together, these corpora provide the historical depth needed to study long-term change while also offering enough 20th-century material to examine competition among modern series. At the same time, they require caution with respect to temporal imbalance and genre comparability (Biber, 2012; Biber & Conrad, 2019). The data represent written usage only; spoken Korean, informal conversation, and prosodic distinctions are largely outside the scope of the present corpus design.

Table 3.1 summarizes corpus size by decade. As in most diachronic corpora, the distribution is temporally imbalanced: in HISTORY, several early decades are sparsely represented (for instance 1700, 1800, 1820), with substantial data density only in the latter half of the 19th century. In contrast, YONSEI exhibits rapid growth throughout the 20th century; the 1990s alone contain over 77 million tokens, accounting for more than half of the modern dataset.

All frequency measures reported in later chapters are normalized per million words (pmw) to mitigate the effect of differing corpus sizes across decades. In decades with extremely small denominators (e.g. 1710, 1810), normalized values

Table 3.1 Corpus size by decade. Note the different orders of magnitude between the pre-modern and modern periods.

MOTU History (1700–1890)		**Yonsei (1900–1990)**	
Decade	**Tokens**	**Decade**	**Tokens**
1700	0	1900	1,630,851
1710	19,997	1910	377,959
1720	44,362	1920	4,612,910
1730	28,546	1930	6,385,476
1740	62,619	1940	1,993,145
1750	56,153	1950	2,936,043
1760	138,785	1960	8,938,621
1770	67,440	1970	11,549,300
1780	24,679	1980	20,614,607
1790	94,758	1990	77,014,276
1800	0		
1810	389		
1820	0		
1830	1,659		
1840	15,703		
1850	90,664		
1860	72,574		
1870	170,652		
1880	223,026		
1890	928,047		

may be driven by very few texts and are interpreted with caution; where necessary, quantitative patterns are checked against qualitative inspection of the underlying examples.

Data collection proceeds in four stages: retrieval, manual screening, annotation, and aggregation.

Retrieval. For History, the absence of morphological tagging and the presence of orthographic variation make lemma-based queries infeasible. Python scripts with regular expressions are used to match major spelling variants of the two bases:

(1) a. 아무 (amwu ‘any, anyone’) and 아모 (amo ‘someone, anyone’) for the *amwu*-type;

b. 누구 (nwukwu ‘who’) and 뉘 (nwi ‘who’) for the *nwukwu*-type.

For each hit, the scripts record document ID, publication year, and genre, so that tokens can later be aggregated by decade and, where relevant, compared qualitatively across genres. For Yonsei, the ILIS web interface and morphological annotation allow lemma-based queries for 누구, 아무, and 아모, retrieving both bare forms and base+case combinations; metadata are exported alongside the concordance lines.

Manual screening. All automatic hits then undergo manual screening to exclude tokens that do not function as indefinites. Three main types of tokens are removed: (i) lexicalized expressions where the indefinite component is semantically bleached (e.g. 아무려나 ‘anyhow’, 아무려기로 ‘after all’, 아모록 ‘commentary’), (ii) derivational homographs such as 뉘웃치다 ‘to sneer’, built on 뉘웃 rather than interrogative 뉘 ‘who’, and (iii) unrelated homographs such as 누에 ‘silkworm’, where 누 does not function as an interrogative or indefinite base. The screening is deliberately conservative: borderline cases are retained only when the indefinite function of the token is clear from context.

Annotation. Each retained token is assigned a single best functional label using the decision-tree procedure in Appendix A. The aim is not to encode every nuance of interpretation, but to place each token in a reproducible region of the semantic map.

Aggregation. After annotation, tokens are aggregated by decade and by series. These aggregated counts form the input to the decade-wise plots, logistic models, segmented regressions, and entry-threshold diagnostics used in the later chapters.

After retrieval and screening, the final dataset for the two bases is as follows:

(2) a. ***nwukwu*-type** (누구, 뉘): 1,210 tokens in History and 133,639 tokens in Yonsei.

b. ***amwu*-type** (아무, 아모): 874 tokens in History and 45,297 tokens in Yonsei.

These counts include bare bases and base+particle combinations (with case and topic markers) and form the empirical backbone for the functional labelling and subsequent analyses in Chapters 4–6.

3.3 Statistical methods

To quantify change and competition, I use simple but informative statistical tools. The models are deliberately descriptive: they summarize broad trends in how often different functions are realized and how those functions drift across the semantic map over time (Bybee, 2010; Haspelmath, 1997; Labov, 1994). The aim is not to fit the most elaborate possible model, but to describe change in a way that stays closely tied to the semantic-map framework and to the token-level annotation procedure.

In all analyses, the basic data points are the labelled tokens described in Section 3.2.1: occurrences of *nwukwu-* and *amo-/amwu-* (and their particle series), each tagged with a semantic-map label (SK, SU, IR, Q, CA, CO, DN, AA, AM, FC, GEN, UFC, IND) and, for some analyses, with a coarser macro-environment label (Q_WH, PPI, Neutral, NPI, Free-choice).[2]

[2]The macro-environment labels are introduced for expository purposes and to track coarse-grained functional shifts over time. Q_WH covers interrogative *wh*-uses; PPI covers positive specific or existential uses (roughly SK/SU); Neutral covers non-negative, non-committal uses such

3.3.1 Data structure and temporal aggregation

The basic unit of analysis is the annotated token. Each token carries at least:

- the root (*nwukwu-* vs. *amo-/amwu-*),
- the year and decade of the text (e.g. 1852, decade 1850),
- a fine-grained functional label (SK, SU, IR, Q, CA, CO, DN, AA, AM, FC, GEN, UFC, IND),
- and, where relevant, a macro-environment label and NPI subtype.

For frequency and proportion plots, tokens are aggregated by decade (10-year bins). Decade-specific corpus sizes (total word tokens) are taken from Table 3.1 and used both for per-million normalization and to flag decades with very little text, where normalized values must be interpreted cautiously. In regression models, decade is treated as a continuous time variable (e.g. 1710, 1720, $\cdots$), yielding a smooth temporal trend over the three centuries considered.

3.3.2 Binary contrasts: logistic regression

Several questions in Chapters 4 and 5 reduce to binary contrasts. A central example is the split between interrogative vs. non-interrogative uses of *nwukwu-*, coded as Q_WH vs. NON_Q. The logistic model for this contrast is given in (3).

(3)
$$\log \frac{\Pr(\text{Q\_WH} \mid t)}{\Pr(\text{NON\_Q} \mid t)} = \beta_0 + \beta_1 \cdot t.$$

Equivalently, one can parameterise the response as the probability of a token being NON_Q, since the two probabilities sum to one. In practice, I fit the model

as irrealis and polar questions (IR/Q/CA); NPI covers negative-polarity environments (DN/AA/AM); and Free-choice covers free-choice and generic uses (FC/GEN/UFC/IND). For details, see Appendix A.

as a binomial GLM in R (Baayen, 2008). For decade-aggregated counts, the implementation is:

(4)

```
glm(cbind(n_wh, n_nonq) ~ decade, family = binomial)
```

The sign and magnitude of β_1 summarize whether the proportion of non-question uses is increasing or decreasing over time. For *nwukwu-*, a negative slope for Q_WH (or a positive slope for NON_Q) captures the shift from an almost purely interrogative root toward a mixed interrogative–indefinite pattern of use, directly bearing on Goal 1 and H1.

3.3.3 Change-points: segmented logistic regression

A simple logistic model assumes a single linear trend across the entire period. To capture possible accelerations or slowdowns in change, I also fit segmented logistic regressions ('broken-stick' models) to decade-aggregated proportions. The segmented specification is shown in (5).

(5)

$$\log \frac{\Pr(\text{Q\_WH} \mid t)}{\Pr(\text{NON\_Q} \mid t)} = \begin{cases} \beta_0 + \beta_1 t & \text{if } t \leq \tau, \\ \beta_0 + \beta_1 \tau + \beta_2 (t - \tau) & \text{if } t > \tau. \end{cases}$$

The breakpoint τ and the slopes before and after τ are estimated using the `segmented` package in R (Muggeo, 2003, 2008). This yields a concrete estimate of when the balance between interrogative and non-interrogative uses changes most rapidly and provides a more informative summary than a single linear trend.

3.3.4 Macro-environments and map-based drift

To relate polarity and free-choice behaviour to the semantic map, I also track the distribution of five macro-environments (Q_WH, PPI, Neutral, NPI, Free-choice)

over time (Haspelmath, 1997). For each root and each decade, I compute the proportion of tokens in each macro-environment and fit logistic models of the form in (6).

(6)
$$\log \frac{\Pr(\text{macro\_env} \mid t)}{\Pr(\text{rest} \mid t)} = \gamma_0 + \gamma_1 \cdot t.$$

Here, `macro_env` stands in turn for PPI, NEUTRAL, NPI, and FREE-CHOICE (Q_WH being treated separately as in Section 3.3.2). The slope γ_1 indicates whether a macro-environment becomes more or less favoured over time.

3.3.5 ENTRY THRESHOLDS FOR NEW FUNCTIONS

Haspelmath-style continuity predictions are qualitative: functions are expected to spread to adjacent regions on the map rather than jumping across distant functions (Haspelmath, 1997). In a token-based diachronic corpus, it is useful to make this more explicit by tracking when a given function becomes well established for a root or series. I do this by computing, for each fine-grained label ℓ (e.g. SU, IR, CO, FC, DN) and each decade t, its proportion within the root's non-question uses, as in (7).

(7)
$$\text{prop}_\ell(t) = \frac{n_\ell(t)}{\sum_{\ell'} n_{\ell'}(t)}.$$

Here, $n_\ell(t)$ denotes the number of NON_Q tokens with label ℓ in decade t. I define the *entry decade* of label ℓ as the first decade in which two conditions are simultaneously met:

- $\text{prop}_\ell(t) \geq 0.05$ (at least 5% of the root's NON_Q tokens in that decade),
- $n_\ell(t) \geq 5$ (at least five tokens).

These thresholds ensure that a function is counted as 'introduced' only once it is both proportionally non-negligible and supported by more than a handful of tokens. This prevents one-off attestations from forcing an artificially early 'first appearance' date, following standard diachronic corpus practice of separating established patterns from sporadic tokens (Hilpert, 2013).

3.3.6 Map-based diagnostics for particle competition

For the shared particle series examined in Chapter 6, I add a simple map-based diagnostic that treats one node as the centre of a region and asks how quickly nearby nodes are recruited. For the NPI particle *-to*, I use the NPI cluster (AA/AM/DN) as centre, assign each label a graph distance from this core (`dist_from_npi`), and relate entry decade to `dist_from_npi`. For the free-choice particle *-na*, I do the same with the FC node as centre (`dist_from_fc`). The general form is given in (8):

(8)

$$\text{first\_decade}_{\ell} \sim \text{distance}_{\ell}.$$

In (8), `distance` is either `dist_from_npi` (for *-to*) or `dist_from_fc` (for *-na*). The aim is not to predict entry decades precisely, but to summarize whether new functions tend to emerge first near the core region of a particle or whether they appear in a more scattered fashion.

3.3.7 Summary

The quantitative side of the dissertation relies on a small set of recurring tools, each aimed at a different aspect of change. Logistic and segmented logistic regression provide compact summaries of binary contrasts and their change-points over time. Macro-environment modelling connects those trends to broad regions of the semantic map. Entry thresholds distinguish established functions from sporadic

attestations. Finally, map-based distance diagnostics provide a compact descriptive measure of how particle series recruit nearby vs. distant semantic-map nodes over time.

All of these methods are used alongside descriptive tables, plots, and example-based discussion. The aim is to describe diachronic trends in a way that stays closely tied to the semantic-map framework in Chapter 2 and the labelling procedure in Appendix A, while keeping the statistical machinery as transparent as possible.

Having defined the labelling scheme and the statistical diagnostics, I now turn to the empirical trajectory of the Korean system. Chapter 4 applies the same decade aggregation, logistic and segmented models, and map-based summaries to the bare roots *nwukwu-* and *amwu-*, establishing their baseline diachronic drift on the semantic map. That root-level baseline then serves as the reference point for Chapter 5 (the development of -(*i*)*nka*) and Chapter 6 (competition in shared particle series).

Chapter 4

Diachronic development of bare *nwukwu* and *amwu*

This chapter develops a usage-based, semantic-map approach to the diachrony of Korean indefinites. It focuses on two bare roots, *nwukwu* and *amwu*, which are central building blocks of the modern system but originate in different lexical classes and follow opposite paths of change. *Nwukwu* begins as the interrogative 'who?', while *amwu* begins as a non-*wh* indefinite meaning roughly 'someone/anyone'. Using the corpus, labelling, and statistical tools introduced in Chapter 3, I track their distributions over three centuries and locate their functions on Haspelmath's (1997) semantic map and Jäger's (2010) polarity cline. The central claim is that polarity behaviour at the root level can be derived from observed usage patterns rather than being treated as a fixed lexical primitive from the outset.

The empirical picture is asymmetric but orderly. Bare *nwukwu* is already very frequent in the earliest texts and remains so throughout the period. What changes is not whether it occurs but where on the map it occurs. In the 18th century, almost all *nwukwu* tokens are interrogative; by the 20th century, around 60% are still interrogatives, but roughly 40% are indefinites, dominated by PPI and NEUTRAL contexts, with stable FREE-CHOICE uses and only a small NPI minority (in the sense of §A.12). *Nwukwu* thus expands from a purely interrogative root into a mixed interrogative–indefinite item, but it never becomes polarity-driven: its non-question uses cluster around existential and free-choice nodes rather than in the negative corner.

Bare *amwu*, by contrast, starts out as a general indefinite and narrows into a strong negative polarity item. In the Early period, about three quarters of bare *amwu* tokens are PPI and roughly 10% are FREE-CHOICE, with only about 6% in NPI contexts. In the Mid period, NPI uses rise to around 29% and PPI drops to about 49%. In the Late period, NPI contexts account for more than 90% of tokens, with PPI and FREE-CHOICE reduced to small residues. Within the NPI domain, *amwu* is overwhelmingly licensed in AA-type environments, and within AA, lexical negation—especially *-eps-* 'not exist, be without'—is dominant, with short and long clausal negation playing only a secondary role. Synchronically, the bare root is therefore strongly polarity-sensitive and distributionally anchored in overtly negative configurations.

In sum, these developments yield a complementary partition of the indefinite space. On Haspelmath's map, *nwukwu* ends up covering the interrogative node plus adjacent existential and free-choice zones, while *amwu* ends up concentrated in the strong NPI zone, mainly in AA environments with lexical negation. On Jäger's polarity cline, this corresponds to opposite movements: *amwu* runs down the cline toward the NPI end, while *nwukwu* moves away from the pure interrogative end toward general indefinites without becoming an NPI itself. The chapter thus offers a concrete, corpus-based implementation of map- and cline-based approaches to polarity: polarity-dependence and interrogative–indefinite ambiguity emerge as root-level specializations carved out by distribution, not as arbitrary lexical stipulations.

More broadly, this chapter also clarifies what a semantic-map analysis can and cannot predict. The map predicts *local directions of change*: it tells us which extensions and retreats are plausible because they proceed through adjacent functions. It does not, by itself, predict exact timing, exact token frequencies, or which competitor will ultimately stabilize in a given niche. Those outcomes depend on additional

factors such as source constructions, licensing patterns, and usage skew. The Korean root system is particularly useful because it shows both sides of this division at once: strong conformity to local pathways on the map, but non-trivial asymmetry in where the two roots actually end up.

These root-level specializations provide the foundation for the particle-bearing paradigms examined in Chapter 5. Once bare *nwukwu* and bare *amwu* have diverged in this way, the series wh+-(*i*)*nka*, wh+-*na*, and *amwu*-based NPIs and FCIs can inherit, reinforce, or reshape their division of labour. Whether the same semantic-map and polarity pressures continue to govern the grammaticalization of full paradigms is the question taken up in the next chapter.

4.1 Hypotheses

To sharpen the descriptive contrast outlined previously in this chapter, the analysis evaluates three hypotheses about the diachronic behaviour of bare *nwukwu* and bare *amwu*, grounded in Haspelmath's (1997) semantic map and Jäger's (2010) polarity cline.

Following Haspelmath, I assume a *contiguity* or *continuity* constraint: diachronic changes in the functions of indefinites proceed only along adjacent nodes on the semantic map, and a root cannot 'jump over' an intermediate function. Following Jäger, I further assume that strengthening of polarity sensitivity corresponds to a diachronic drift into increasingly negative environments. These assumptions make directional rather than deterministic predictions: they tell us which pathways are plausible, but not exactly when change will occur or how completely a series will fill a given region.

(**H-ROOT i**) ***nwukwu* expands from interrogative to indefinite territory.**

Bare *nwukwu* starts as a canonical interrogative ('who?') and gradually develops non-interrogative readings, first in specific and neutral contexts and later

in generic or concessive ones. Its interrogative use remains core, but a stable minority of tokens are predicted to behave as indefinites in PPI, NEUTRAL, and FREE-CHOICE environments (as defined in §A.12). On Haspelmath's map, this corresponds to an expansion from the interrogative node into adjacent existential and, later, free-choice regions, without entering the NPI node. *Nwukwu* is not expected to develop strong polarity sensitivity.

(**H-ROOT ii**) ***amwu* narrows into a strong NPI.**

Bare *amwu* begins as a general-purpose indefinite with existential and free-choice readings in contexts such as questions, conditionals, and irrealis clauses. Over time, its distribution is predicted to become increasingly restricted to negative-polarity environments, yielding its modern status as a strong NPI. On Haspelmath's map, this corresponds to a contraction from the existential/free-choice region into the NPI node; on Jäger's polarity cline, it is a shift from positive or neutral toward maximally negative licensors.

(**H-ROOT iii**) **The two roots become functionally complementary.**

As these changes proceed, bare *nwukwu* and bare *amwu* are expected to evolve into functionally complementary items. *Nwukwu* should extend from the interrogative node into modal and generic indefinite contexts, while *amwu* should specialise in polarity-sensitive environments, especially those involving lexical negation. The resulting distributions should respect contiguity on the semantic map and avoid large- scale overlap at the level of dominant functions, even if small transitional or residual overlaps remain at the edges of the two systems.

The following sections evaluate these hypotheses using corpus-based frequency trends, per-million normalization with respect to decade and period corpus sizes,

functional annotation of licensing contexts (as introduced in §A.12), and descriptive statistical modelling of temporal change (see §3.3.3). I begin in §4.2 with bare *nwukwu* (H-ROOT i), then turn to bare *amwu* in §4.3 (H-ROOT ii), and finally compare the two roots in §4.4 (H-ROOT iii).

4.2 Bare *nwukwu* ('who') as a bare root

This section tests H-ROOT i by tracing how bare *nwukwu* develops from an almost purely interrogative item into a root that also has indefinite uses. I first look at how often *nwukwu* occurs over time, and then turn to the contexts in which it appears, using the categories in §A.12. The modelling in §4.2.3 builds on these descriptive facts.

4.2.1 Raw frequencies over time

Figure 4.1 summarizes the frequency of bare *nwukwu* from the 1700s to the 1990s. The earliest clear attestations appear in the 1710s and 1720s. In these decades the corpus is very small and contains many *wh*-questions, so normalized values are very high: in the 1710s there are 90 tokens in 0.02 million words (4501 pmw), and in the 1720s 104 tokens in 0.04 million words (2344 pmw). These figures show that *nwukwu* already functions as the main human *wh*-pronoun, but the pmw values are inflated by corpus size and text type.

If we look beyond these early decades, a more stable picture emerges. In the later 18th century, bare *nwukwu* continues to appear regularly: in the 1760s there are 100 tokens in 0.14 million words (721 pmw), in the 1770s 60 tokens in 0.07 million words (890 pmw), and in the 1790s 47 tokens in 0.09 million words (496 pmw). Very small corpora in some decades can again produce high pmw values even when raw counts are low (for example, the 1830s: 2 tokens in very little text, reported as 1206 pmw; the 1840s: 22 tokens in 0.02 million words, 1401 pmw), and these should

therefore be treated with caution rather than as evidence of sudden increases in use.

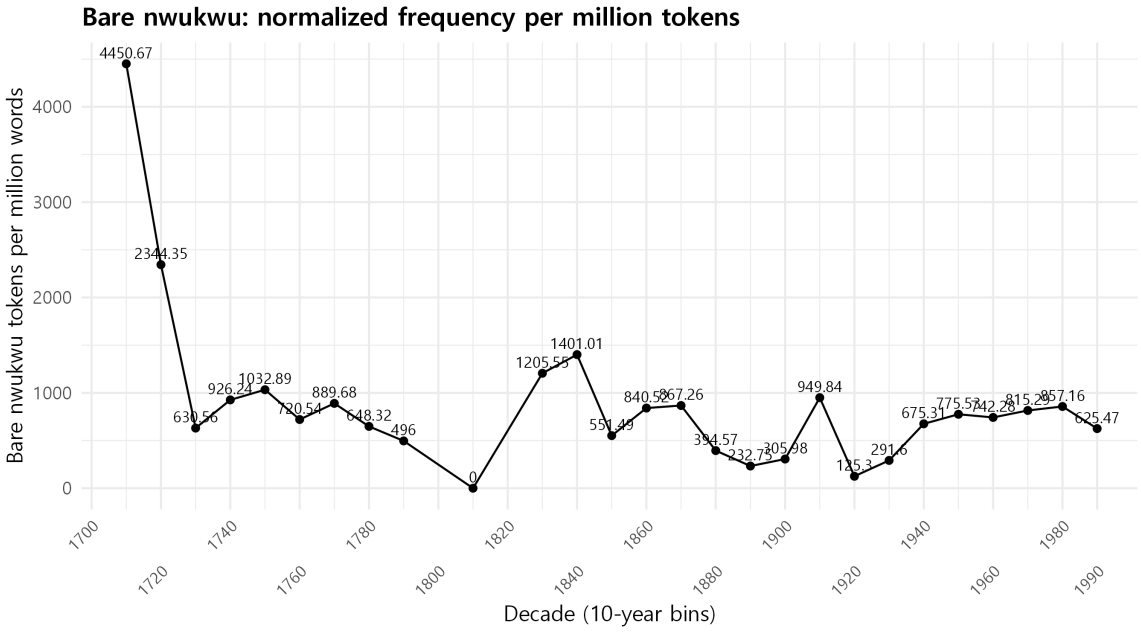


Figure 4.1 Normalized frequency (per million words) of bare *nwukwu* by decade.

From the mid-19th century onward the corpora become larger and more balanced, and both raw counts and normalized frequencies show that *nwukwu* is a stable, high-frequency item. In the 1870s there are 148 tokens in 0.17 million words (867 pmw); by the 1890s, with a much larger corpus (0.93 million words), there are still 216 tokens (233 pmw). In the 20th century, raw counts increase sharply as the corpus expands: 499 tokens in 1.63 million words in the 1900s (306 pmw), 359 tokens in 0.38 million words in the 1910s (950 pmw), 1862 tokens in 6.39 million words in the 1930s (292 pmw), and 2277 tokens in 2.94 million words in the 1950s (776 pmw). The later decades show the same pattern at a larger scale: 6635 tokens in 8.94 million words in the 1960s (742 pmw), 9416 tokens in 11.55 million words in the 1970s (815 pmw), 17,670 tokens in 20.61 million words in the 1980s (857 pmw), and 48,174 tokens in 77.01 million words in the 1990s (626 pmw).

Once early, very small decades are set aside, *nwukwu* typically appears between about 400 and 900 pmw, and there is no sign that it is disappearing. The question for the rest of this section is therefore not whether *nwukwu* survives, but how this steady frequency is divided between interrogative and indefinite uses and which environments these uses favour.

4.2.2 Licensing environments by period

In its standard interrogative use (Q_WH), *nwukwu* appears in matrix questions asking for the identity of a person. A 1704 historical text, for example, has a host addressing an unexpected guest as in (1):

(1) I sonnim-un nwukwu-si-wan-de ettehkey
this guest-TOP who.HON-COP-INT how
tul-e-wa kyey-sin-go?
enter-come be-HON-Q
'Who is this guest, and how did you come in here?' ⇒ speaker asks for the identity of the guest who has entered
[Q_WH (matrix)]

Here *nwukwu-si-* clearly behaves as a *wh*-pronoun: the clause asks for an identity and is not understood as saying simply that someone came.

Non-interrogative PPI uses appear when *nwukwu* is used to refer to a specific but unnamed individual in a positive, non-negated clause. In a 1721 historical text, for instance, a narrator speaks about certain wealthy households, as in (2):

(2) Pwukwi-han nwukwu-uy cip-tul-un
wealthy who-GEN house-PL-TOP
paksan hwullo-lul pel-e noh-ko
mountain.pass road-ACC throw-CONN put-CONN
'At certain wealthy households, the mountain-pass road lies abandoned to either side.' ⇒ speaker refers to some (unspecified) wealthy households
[PPI]

Here *nwukwu-uy cip-tul* ('somebody's houses') is understood as 'certain wealthy households', not as a question. This is the kind of use counted as PPI in Table 4.1.

NEUTRAL contexts include futures, conditionals, and other non-negative statements that talk about possibilities rather than settled facts. In a mid-19th-century text, *nwukwu* appears in the antecedent of a wish-like conditional, as in (3):

(3) Taman nwukwu-ka twitcen kwanum-kak-ul
only who-NOM rear Gwaneum-pavilion-ACC
cis-e-cwu-nun caycwu-ka iss-umyen
build-give-REL talent-NOM exist-COND
'If only someone had the skill to build the rear Gwaneum pavilion.' ⇒ conditional clause with a non-specific subject
[NEUTRAL (conditional)]

This clause does not ask 'who?'; it quantifies over possible builders. If there is some person with the relevant skill, the wish can be satisfied. Such cases are annotated as NEUTRAL.

Negative-polarity (NPI) uses of bare *nwukwu* include rhetorical questions under lexical negation and cases with clausal negation. An 18th-century text states that nobody is without faults, as in (4):

(4) Salam-i nwukwu hemwul-i
person-NOM who fault-NOM
eps-ul-i-o-manun
not.exist-FUT-Q-but
'What person has no faults?' ⇒ rhetorical question, understood as 'no one is without faults'
[NPI (LEX)]

Here *nwukwu* combines with *eps-* 'not exist' and yields a strong negative meaning: for any person, it is not true that they are without faults. Later DN examples show similar NPI readings, often with minimisers such as *hana* 'one'. For instance, in (5)–(7) from modern texts, nobody cuts a bird's claws, nobody has a clear life goal, and nobody steps forward to help:

(5) Nwuka kkak-a cwu-ci-to
who.NOM cut-CONN give-CONN-ADD
anh-ko ku sonthop-un
not-AND that fingernail-TOP
cal-alttaylo cal-ass-ta.
as_it_pleases grow-PST-DECL
'No one cuts them, so the nails just grow as they please.' ⇒ there is no person who cuts them; the nails grow unchecked
[NPI (DN)]

(6) Kutul-un nwuku hana
they-TOP who one
insayng-ey hwaksilhan mokpyo-lul
life-LOC definite goal-ACC
kac-ko iss-ci anh-ass-ta.
have-CONN exist-COMP not-PST-DECL

'None of them had any clear goal in life.' ⇒ for no person x is it true that x had a definite goal

[NPI (DN, *nwuku hana*)]

(7) Nwuku han salam
who one person
na-se-ci anh-ass-supnita.
come.out-COMP not-PST-DECL.POL

'Not a single person stepped forward.' ⇒ no person x is such that x stepped forward

[NPI (DN, *nwuku han salam*)]

In these DN cases, *nwukwu* appears under clausal negation, often with *hana* 'one'; the reading is emphatic 'no one / not a single person'.

FREE-CHOICE contexts for *nwukwu* in the corpus include concessive, generic, and indiscriminative uses. In practice, these readings are usually supported by extra morphology or particles—most often plural or distributive markers such as *-tul* or expressions like *mulonhako* 'regardless of'—rather than by the bare root alone. The examples in (8)–(9) illustrate this part of the space.

(8) Cehi hakyo-eyse nwukwu-hanthey-ken phyenci-lul
our school-at who-to-ever letter-ACC
sse-se pwuci-la-ko hay-ess-ta.
write-CONN send-IMP-COMP say-PST-DECL
'At our school, they told us to write a letter to anyone and send it.' ⇒ the addressee can be any person whatsoever
[FREE-CHOICE]

(9) Ilnyem palsim ha-key tul-myen
single.thought resolve do-INF enter-if
nwukwu nwukwu mulonhako ⋯
who who regardless.of ⋯
'If one truly sets their mind, then whoever it is, without exception, ⋯' ⇒ indiscriminative 'whoever' reading
[FREE-CHOICE (IND)]

In many FREE-CHOICE cases, especially IND-like ones, the reading depends on such additional material rather than on the bare root alone. In the data examined here, bare *nwukwu* without such support is much less often interpreted as fully free-choice or generic. These combinations therefore warrant more detailed study in future work.

The examples in this subsection show that when *nwukwu* is not interrogative, it occurs in all of the main types of contexts used in this chapter: specific and non-specific existentials (PPI, NEUTRAL), negative-polarity environments (NPI), and several kinds of free-choice and generic uses (FREE-CHOICE). Table 4.1 now gives an overview of how these non-question uses are distributed across three broad periods (Early 1700–1849, Mid 1850–1899, Late 1900–1999).

Table 4.1 Distribution of PPI, NEUTRAL, NPI, and FREE-CHOICE contexts for non-Q_WH uses of bare *nwukwu* by period.

Period	PPI	NEUTRAL	NPI	FREE-CHOICE
Early (1700–1849)	0.647	0.147	0.176	0.029
Mid (1850–1899)	0.596	0.213	0.021	0.170
Late (1900–1999)	0.396	0.264	0.121	0.219

In the **Early period** (1700–1849), non-question uses of *nwukwu* are still rare: only about 6% of all tokens are NON_Q (§4.2.3). Within this small set, PPI contexts make up about 65%, NEUTRAL and NPI each contribute around 15–18%, and FREE-CHOICE uses are around 3%. Even at this stage, then, *nwukwu* can be used as 'someone' in positive and non-negative, non-factual contexts alongside its main interrogative use.

In the **Mid period** (1850–1899), non-question uses become slightly more common, and their internal make-up shifts away from NPI. PPI remains the largest category (about 60%), NEUTRAL contexts rise to about 21%, FREE-CHOICE uses rise to 17%, and NPI contexts fall to only about 2% of NON_Q tokens. The new indefinite uses of *nwukwu* in this period are therefore mainly positive and non-negative: it is used as 'someone' in simple statements and in futures, questions, and conditionals, and as 'whoever/anyone' in free-choice constructions, while NPI-like uses under negation remain rare.

In the **Late period** (1900–1999), the non-question uses of *nwukwu* match the mixed interrogative–indefinite pattern seen in Modern Korean. Among NON_Q tokens, PPI contexts still have the largest share (about 40%), followed by NEUTRAL (26%) and FREE-CHOICE (22%), with NPI contexts at about 12%. At the same time, Q_WH uses remain the majority of all *nwukwu* tokens in each period (over 90% in Early and Mid, around 60% in Late; see §4.2.3). *Nwukwu* therefore never becomes mainly a polarity item. It keeps its interrogative core, while a smaller but stable

set of indefinite uses spreads across positive, neutral, and free-choice contexts, and only a minor part of its non-question uses involves NPI environments.

The qualitative and quantitative patterns here suggest that the main change in *nwukwu* concerns function rather than survival. Over time, the root keeps its high frequency but gains a substantial set of indefinite uses around its interrogative core. In the next subsection, I use formal modelling to describe this shift more precisely and to measure the balance between interrogative and non-interrogative uses.

4.2.3 Change-point analysis and modelling results

The descriptive patterns above suggest that bare *nwukwu* began as an almost purely interrogative item and later gained a steady set of indefinite uses, while keeping its interrogative use as the main one. In this subsection I make this picture more precise in two steps. First, I model the rise of non-interrogative uses over time. Second, I look at how those non-interrogative uses are divided among PPI, NEUTRAL, NPI, and FREE-CHOICE contexts.

Interrogative vs. non-interrogative uses of *nwukwu*

To measure the change from interrogative to mixed uses, I use a logistic regression on decade-aggregated data, classifying each token as either interrogative (Q_WH = 1) or non-interrogative (Q_WH = 0), with decade as the predictor (see §3.3.2). Table 4.2 gives the proportions of Q_WH vs. NON_Q in the three macro-periods (Early, Mid, Late), and Figure 4.2 plots the decade-wise trajectory together with the fitted curve.

Table 4.2 Q_WH vs. NON_Q uses of bare *nwukwu* by period.

Period	Type	Count	Proportion
Early (1700–1849)	Q_WH	538	0.941
	NON_Q	34	0.059
Mid (1850–1899)	Q_WH	504	0.915
	NON_Q	47	0.085
Late (1900–1999)	Q_WH	51148	0.606
	NON_Q	33212	0.394

The fitted model shows a clear increase over time in the chance that a *nwukwu* token is non-interrogative. In the Early period, only about 6% of tokens are NON_Q, which matches the idea that *nwukwu* is almost exclusively a question word. In the Mid period, the NON_Q proportion rises to about 8.5%, and in the Late period it reaches about 39.4% (Table 4.2). By the end of the period, then, *nwukwu* still appears more often in questions, but a sizable share of tokens are now indefinites.

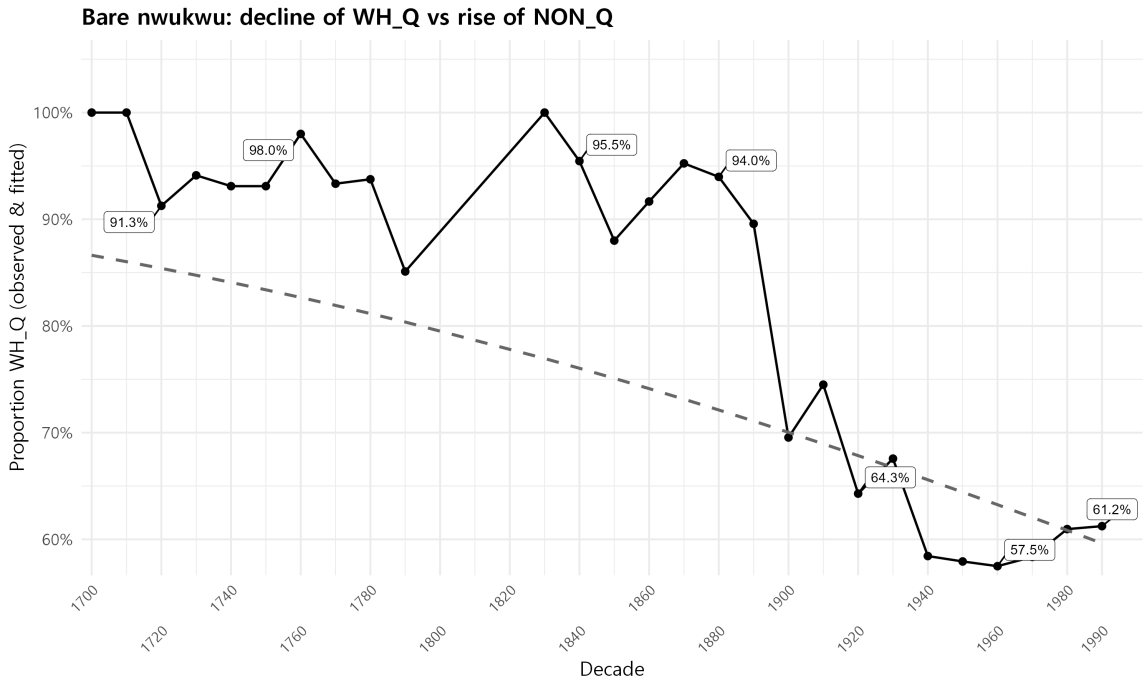


Figure 4.2 Observed and fitted proportions of Q_WH uses for bare *nwukwu* by decade (logistic regression).

To see when this shift happens, I also fit a segmented logistic regression to the same decade-wise proportions (§3.3.3). The segmented model locates a change-point in the 19th century (the estimate falls in the mid-1800s). Before this point, the proportion of Q_WH uses drops relatively quickly from decade to decade, which means that NON_Q uses are increasing at a noticeable rate. After the change-point, the slope is much flatter: by the early 20th century, the system looks fairly stable, with a minority of *nwukwu* tokens used as indefinites and a majority used as interrogatives. Figure 4.3 shows the segmented fit and the estimated breakpoint.

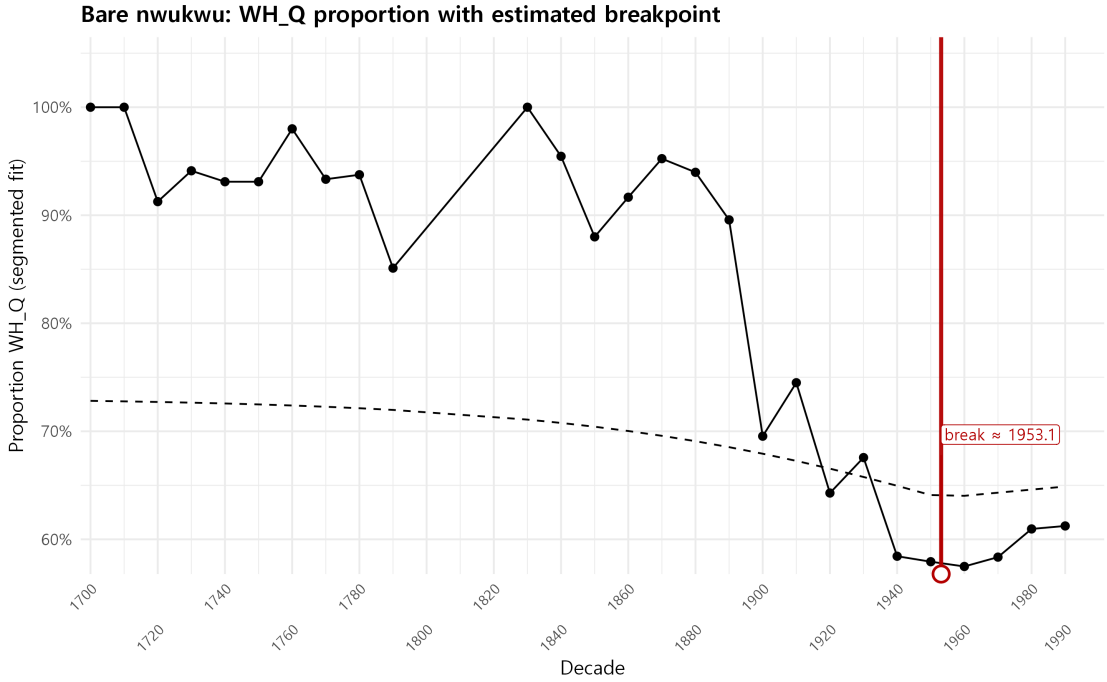


Figure 4.3 Segmented logistic regression for Q_WH vs. NON_Q uses of bare *nwukwu* by decade, with estimated breakpoint (vertical line).

For ease of comparison, Figure 4.4 plots the proportions of Q_WH vs. NON_Q by decade in a stacked format. The figure makes it visually clear that NON_Q uses slowly increase across the period, but never come to outweigh Q_WH uses.

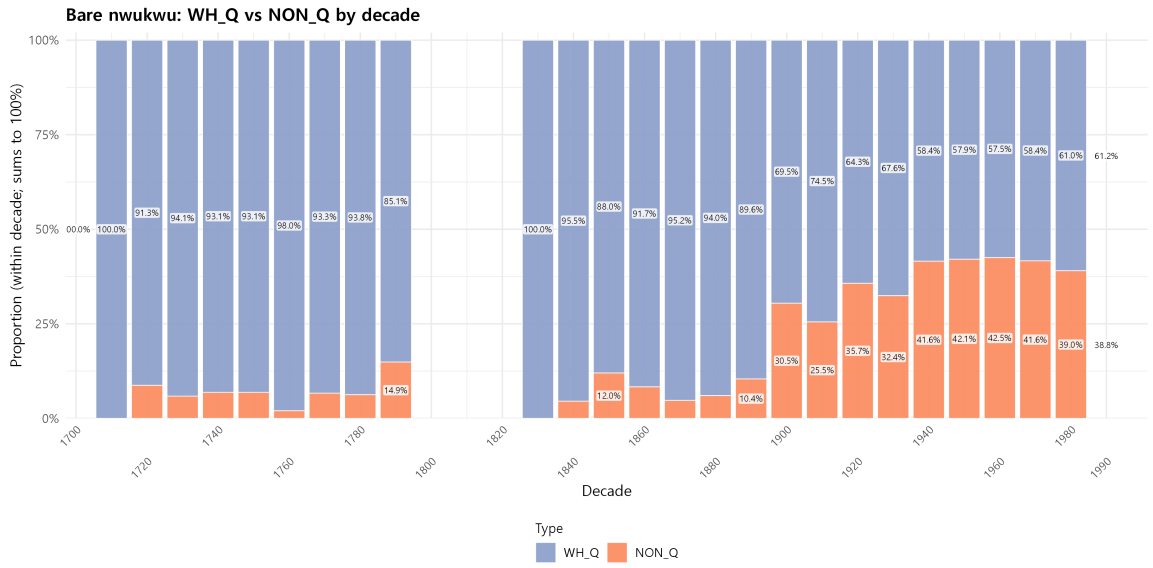


Figure 4.4 Stacked proportions of Q_WH vs. NON_Q uses of bare *nwukwu* by decade.

These models, together with the descriptive counts, show that *nwukwu* moves from being almost only a question word to a mixed use in which interrogative tokens still dominate but indefinite tokens form a clear minority. There is no sign that it moves on to a stage where negative contexts would become its main home.

Internal structure of the non-interrogative uses of *nwukwu*

The licensing analysis in §4.2.2 showed that, when *nwukwu* is used non-interrogatively, it occurs in PPI, NEUTRAL, FREE-CHOICE, and a smaller number of NPI contexts (Table 4.1). Here I use two diagnostics based on the semantic map to describe this pattern more closely: an ordered 'map state' variable and the entry-time analysis introduced in §3.3.5.

First, I group the macro-contexts into an ordered map state that reflects distance from the interrogative node: state 0 for Q_WH, state 1 for PPI and NEUTRAL (existential and near-existential uses), state 2 for FREE-CHOICE, and state 3 for NPI (§3.3.6). A cumulative logit model with decade as predictor, fitted to decade-wise

counts of these states, yields a positive slope: as time goes on, tokens are more likely to appear in the higher states (1–3) relative to state 0. Put more simply, the earliest decades consist almost entirely of state 0 (questions), and later decades contain a mix of state 0 and 1, plus a smaller band of state 2 and a thin band of state 3. The model thus captures the gradual move from purely interrogative uses toward a mix where existential and free-choice uses play an important supporting role, and NPI uses remain on the edge.

Second, I use the fine-grained functional labels (SK, SU, IR, etc.) and the graph-based semantic map introduced in §3.3.6 to ask in what order new functions become firmly established. For each label ℓ and decade t, I compute its proportion among *nwukwu* NON_Q tokens, $\text{prop}_\ell(t)$, and define the entry decade of ℓ as the first decade in which $\text{prop}_\ell(t) \geq 0.05$ and $n_\ell(t) \geq 5$ (§3.3.5). For *nwukwu*, only two labels meet this criterion: SU and IR. SU, which occupies the specific-unknown existential node next to Q_WH, reaches the threshold already in the 1720s. IR, which is one step further away in the NEUTRAL zone, does not reach the same level until the 1930s. No more distant functions (for example, free-choice or strong negative nodes) ever meet the '5% and 5 tokens' threshold for *nwukwu*.

These two diagnostics point in the same direction. The ordered map-state model shows that non-interrogative uses of *nwukwu* build up mainly in the existential and free-choice areas, and the entry-time analysis shows that SU-type existential uses appear earlier and more firmly than IR-type NEUTRAL uses, with no function in the strongly negative corner of the map becoming central for this root. Put in map terms, the change in *nwukwu* moves step by step from the interrogative node into neighbouring existential territory and only later into some NEUTRAL and FREE-CHOICE uses, without any direct move into the NPI node.

4.2.4 Conclusion for bare *nwukwu*

The evidence for *nwukwu* provides a clear answer to the central question of this part of the chapter. Bare *nwukwu* begins as overwhelmingly interrogative and only later develops a stable, minority profile as an indefinite. In the Early period (1700–1849), roughly 94% of tokens occur in *wh*-question environments (Q_WH), with only about 6% in non-question contexts (NON_Q) (Table 4.2). The small NON_Q set at this stage is concentrated in contexts that remain close to questions, such as embedded-question-like environments and ignorance-reporting frames, which is consistent with the view that *nwukwu* is historically a 'who?' item rather than an inherent indefinite.

Across the 19th and 20th centuries, *nwukwu* acquires more non-interrogative uses and then settles into a mixed distribution. The modelling in §4.2.3 shows that the proportion of NON_Q tokens rises from about 6% in the Early period to about 39% in the Late period (Table 4.2), while Q_WH remains the majority throughout. The segmented regression further indicates a 19th-century change-point, with relatively rapid growth in NON_Q uses before that point and slower change thereafter. The resulting trajectory is therefore not a shift into polarity-sensitive territory, but an expansion from a purely interrogative zone to a mixed interrogative–indefinite zone.

Within the NON_Q tokens, negative environments never become dominant. Period-level distributions (Table 4.1) show that most non-question uses fall in PPI and NEUTRAL contexts, with FREE-CHOICE contributing a smaller but stable share and NPI remaining the smallest category across periods. The map-state model in §4.2.3 captures this as a gradual redistribution of tokens from the question state into neighbouring existential and free-choice states. The entry-time results are consistent with this picture in showing that SU-type existential uses

become robustly attested earlier than IR-type NEUTRAL uses, and no strongly negative function ever emerges as central.

Overall, *nwukwu* expands from the interrogative node to a wider but still connected region of the semantic space. It remains the primary question word 'who?', while developing an increasingly substantial set of indefinite uses, especially in positive and neutral environments and, to a lesser extent, in free-choice and negative contexts. The next section turns to bare *amwu* and applies the same diagnostics and modelling strategy to its diachronic trajectory.

4.3 Bare *amwu* as a bare root

4.3.1 Raw frequencies over time

Figure 4.5 shows the normalized frequency (per million words, pmw) of bare *amwu* from the 1710s to the 1990s. The earliest clear examples appear in the 1710s and 1730s: there are 2 tokens (100.0 pmw) in the 1710s and 3 tokens (105.1 pmw) in the 1730s. Through the mid-18th century, bare *amwu* continues to occur regularly, with 8 tokens (127.8 pmw) in the 1740s, 8 tokens (142.5 pmw) in the 1750s, and 20 tokens (144.1 pmw) in the 1760s. Even at this early stage, *amwu* is a noticeable indefinite root rather than a rare form.

There is a local peak in the 1770s and especially the 1780s. In the 1770s, 30 tokens (444.8 pmw) are found, and in the 1780s, 16 tokens (648.3 pmw). These high pmw values reflect both small corpus sizes in those decades and the fact that *amwu* is frequent in some texts. After this, the early 19th century moves into a more moderate range: 23 tokens (242.7 pmw) in the 1790s, 4 tokens (254.7 pmw) in the 1840s, 11 tokens (121.3 pmw) in the 1850s, 14 tokens (192.9 pmw) in the 1860s, and 5 tokens (29.3 pmw) in the 1870s.

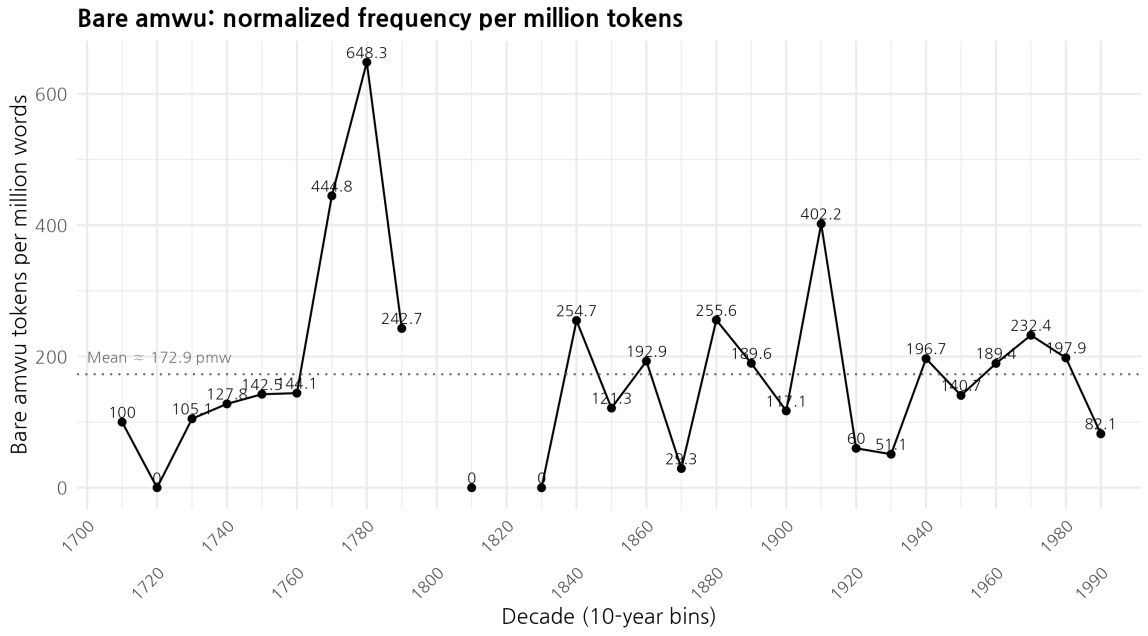


Figure 4.5 Normalized frequency (per million words) of bare *amwu* by decade.

From the late 19th century onward, the corpora become much larger, and both raw counts and normalized frequencies show that *amwu* continues to be used. In the 1880s, there are 57 tokens (255.6 pmw), and in the 1890s, 176 tokens (189.6 pmw). In the 20th century, raw counts rise sharply as the corpus grows, while pmw values stay in a similar band: 191 tokens (117.1 pmw) in the 1900s, 152 tokens (402.2 pmw) in the 1910s, 277 tokens (60.0 pmw) in the 1920s, 326 tokens (51.1 pmw) in the 1930s, and 392 tokens (196.7 pmw) in the 1940s. Later decades continue in the same way: 413 tokens (140.7 pmw) in the 1950s, 1693 tokens (189.4 pmw) in the 1960s, 2684 tokens (232.4 pmw) in the 1970s, and 4079 tokens (197.9 pmw) in the 1980s. Only the 1990s show a lower normalized value, with 6326 tokens (82.1 pmw), which is likely due to the very large modern corpus (especially newspapers and other expository prose) rather than a sudden drop in the use of *amwu* itself.

Overall, Figure 4.5 shows that bare *amwu* appears throughout the period covered here. As with *nwukwu*, there is no sign that the root is being lost. The main

diachronic question is how the *use* of *amwu* changes over time—more specifically, how it moves from a broad existential indefinite toward an item that is mainly associated with negative contexts. The next subsection looks at this in terms of PPI, NEUTRAL, NPI, and FREE-CHOICE environments.

4.3.2 Licensing environments by period

As with *nwukwu* (§4.2.2), each bare *amwu* token is classified into the four macro-contexts introduced in §A.12: PPI, NEUTRAL, NPI, and FREE-CHOICE. Off-map labels such as Q_WH, NO_MATTER, and EXPRESSION are excluded from the quantitative analysis here. I therefore use the same semantic-map zones for *amwu* as for *nwukwu*, but, as will become clear in §4.3.3 and §4.3.3, special attention will be paid to how often *amwu* appears in NPI contexts over time.

Table 4.3 shows how often bare *amwu* appears in each macro-context across three periods (Early 1700–1849, Mid 1850–1899, Late 1900–1999), and Figure 4.6 plots the same information by decade.

Table 4.3 Distribution of PPI, NEUTRAL, NPI, and FREE-CHOICE contexts for bare *amwu* by period.

Period	PPI	NEUTRAL	NPI	FREE-CHOICE
Early (1700–1849)	0.7620	0.0792	0.0594	0.0990
Mid (1850–1899)	0.4940	0.0776	0.2900	0.1390
Late (1900–1999)	0.0451	0.0013	0.9290	0.0245

Before turning to these proportions, it is useful to have concrete examples for each macro-context.

In PPI contexts, *amwu* behaves like an ordinary existential indefinite, referring to a specific but unnamed person or time. In (10), a 1765 text uses *amwu tal* ‘a certain month’ in a repayment agreement:

(10) Ku un-ul naynyen amwu tal
that silver-ACC next.year some month
nay-ey ireu-e kapum-ul eynak-ha-ye ⋯
inside reach-INF repay-ACC promise-do-CONN ⋯

'He promised to repay the silver by some month next year.' ⇒ a specific but not named month for repayment

[[PPI (SK), 1765, HISTORY (朴通事新釋諺解)]]

Similar uses appear with expressions such as *amwu taekwanin* 'a certain official', where the person is known in context but not named explicitly. These SK-type uses make up the core of the PPI band in the Early period.

NEUTRAL contexts include conditionals and other non-negative sentences that talk about possible situations. In (11), a 1737 text has *amwu salam* 'any person' in the antecedent of a conditional:

(11) Amwu salam-i na-lul chac-ke-tun
amwu person-NOM me-ACC find-CONN-COND
ce-lul nayil o-la ha-ko ⋯
3SG-ACC tomorrow come-IMP say-CONN ⋯

'If any person finds me, tell them to come tomorrow.' ⇒ conditional with a non-specific subject

[[NEUTRAL (CA), 1737, HISTORY (捷解蒙語)]]

Here the clause talks about any potential finder; it does not claim that anyone has already appeared. Such conditionals contribute to the NEUTRAL band.

In NPI contexts, *amwu* appears under the scope of negation and is interpreted as 'anyone/anything' in a negative statement. In (12), a 19th-century religious text describes someone as knowing nothing at all:

(12) Kyoman-ha-ye amwu kes-ul al-ci
arrogant-be-CONN amwu thing-ACC know-COMP
mos-ha-ko ⋯
cannot-do-CONN ⋯
'Being arrogant, he knows nothing.' ⇒ negated existential 'not know anything'
[[NPI (DN), 1887, HISTORY (예수셩교젼셔)]]

The string *amwu kes-ul al-ci mos-ha-* is naturally read as 'he does not know anything'. In AA-type NPI uses, *amwu* combines with expressions like *amwu il eps-i* 'without any trouble', which are coded as anti-additive negative environments and play a major role in *amwu* NPIs (see §4.3.3).

FREE-CHOICE contexts for *amwu* cover several subtypes in the fine-grained labelling (CO, GEN, IND). A concessive comparison ('than anything else') is illustrated in (13):

(13) Amwu il-to-kon aleumtap-key ye-ki-naida.
amwu thing-even-COMP beautiful-ADV think-DECL
'We consider this more beautiful than anything.' ⇒ comparison against an 'anything' standard
[[FREE_CHOICE (CO), 1748, HISTORY (改修捷解新語)]]

GEN and IND uses often involve generic or loosely defined sets. In (14), from a 1900 theology journal, *amwu paekseng* refers to people in general:

(14) Pubu-to-lul cung-ha-key kyenghi-ye-ki-m-un
marriage-bond-ACC important-ADV light-think-NOM-TOP
amwu paekseng-uy mwunhwa-lul tele-nay-mi-ni
amwu people-GEN culture-ACC reveal-NMLZ-DECL
'To value the marriage bond highly or lightly shows the culture of any people.' ⇒ generic statement about 'any people'
[[FREE_CHOICE (GEN), 1900, HISTORY (신학월보)]]

In (15), from a 1796 history text, *amwu kes* 'any things' is used for a random assortment of items:

(15) Kulssey ssui-n cong-i-lul kyengsekh-ci anh-a-hamye
writing be.on-REL paper-ACC waste-COMP not-do-CONN
amwu kes-ul ssa-mye chang pala-mye ⋯
amwu thing-ACC wrap-CONN window paste-CONN ⋯
'They did not waste paper with writing on it, but wrapped and pasted windows with any things.' ⇒ indiscriminative 'any/random things'
[[FREE_CHOICE (IND), 1796, HISTORY (N/A)]]

Here *amwu kes* does not point to a fixed set but to 'whatever things' are used for wrapping.

These examples show that, in the historical data, bare *amwu* appears in all four macro-contexts defined in §A.12: PPI, NEUTRAL, NPI, and several types of FREE-CHOICE uses (CO, GEN, IND). In the Early period, PPI and FREE-CHOICE uses are common, and NPI uses form only a small part of the overall distribution. Table 4.3 and Figure 4.6 now show how this balance changes over time.

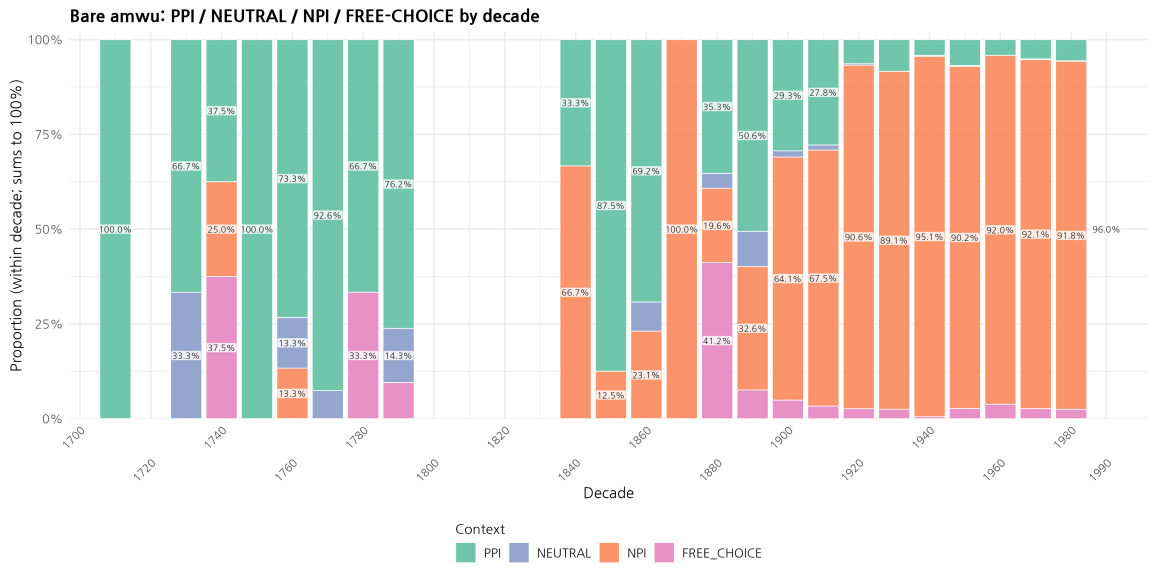


Figure 4.6 Stacked proportions of PPI, NEUTRAL, NPI, and FREE-CHOICE contexts for bare *amwu* by decade.

In the **Early period** (1700–1849), bare *amwu* behaves like a general existential indefinite. About 76% of tokens occur in PPI contexts, roughly 8% in NEUTRAL contexts, about 10% in FREE-CHOICE contexts, and only about 6% in NPI contexts (Table 4.3). The root appears in straightforward positive clauses ('someone/some N'), in questions and conditionals, and in free-choice environments ('any N'), with negative contexts making up only a small part of its use.

The **Mid period** (1850–1899) is a transition stage. PPI uses drop to about 49%, NPI uses rise to 29%, FREE-CHOICE uses are about 14%, and NEUTRAL uses are around 8% (Table 4.3). The stacked plot in Figure 4.6 shows that this shift builds up across the 19th century: existential and free-choice uses remain important, but negative contexts become a much more visible part of the picture. This is the kind of middle stage we would expect if the root is moving along Jäger's (2010) polarity cline from positive or neutral existentials toward an NPI.

In the **Late period** (1900–1999), the narrowing is essentially complete. NPI contexts account for approximately 93% of all bare *amwu* tokens, with PPI uses at 4.5%,

FREE-CHOICE uses at 2.5%, and NEUTRAL uses at about 0.1% (Table 4.3). The decade-wise distribution in Figure 4.6 makes this clear: from the 1920s onward, the NPI band covers almost the entire stack, and the other functions survive only as small remainders. The root's most usual reading has moved from the existential and free-choice parts of the map into the NPI node. The next subsection makes this polarity-driven narrowing more explicit using logistic and segmented regression, and examines which negative environments dominate the NPI uses of *amwu*.

4.3.3 Change-point analysis and modelling results

The descriptive patterns in §4.3.2 suggest that bare *amwu* moves from a broadly available existential indefinite to an item whose uses are mostly tied to negative contexts: NPI uses go from marginal to covering almost all tokens, while PPI and FREE-CHOICE uses shrink to small remainders. In this subsection I make that idea more precise. First, I model the rise of NPI uses over time with logistic and segmented regression. Second, I examine the internal make-up of *amwu* NPIs in terms of the NPI subtypes AA, AM, and DN (§A.11.1).

Rise of NPI usage along Jäger's polarity cline

To measure the shift toward NPI uses, I treat each bare *amwu* token as either NPI or non-NPI and fit a logistic regression to decade-aggregated data, with decade as the predictor. The fitted curve in Figure 4.7 shows the proportion of NPI tokens over time, together with the observed decade-wise proportions.

In the 18th century, NPI uses are almost absent: in the 1710s, 1730s, 1750s, 1770s, 1780s, and 1790s, NPI counts are zero, and even in the 1760s only 2 out of 15 tokens (about 13%) are NPI. The 19th century begins to show clear growth in NPI uses. In the 1880s, 10 out of 51 tokens are NPI (about 20%), and by the 1890s, 56 out of 172 tokens (about 33%) are NPI. In the early 20th century the NPI share rises quickly:

in the 1900s, 118 out of 184 tokens are NPI (about 64%), in the 1910s 102 out of 151 tokens (about 68%), and in the 1920s 242 out of 267 tokens (about 91%). By the 1930s, the change is effectively complete: 287 out of 322 tokens are NPI (about 89%), and non-NPI uses are already a small minority.

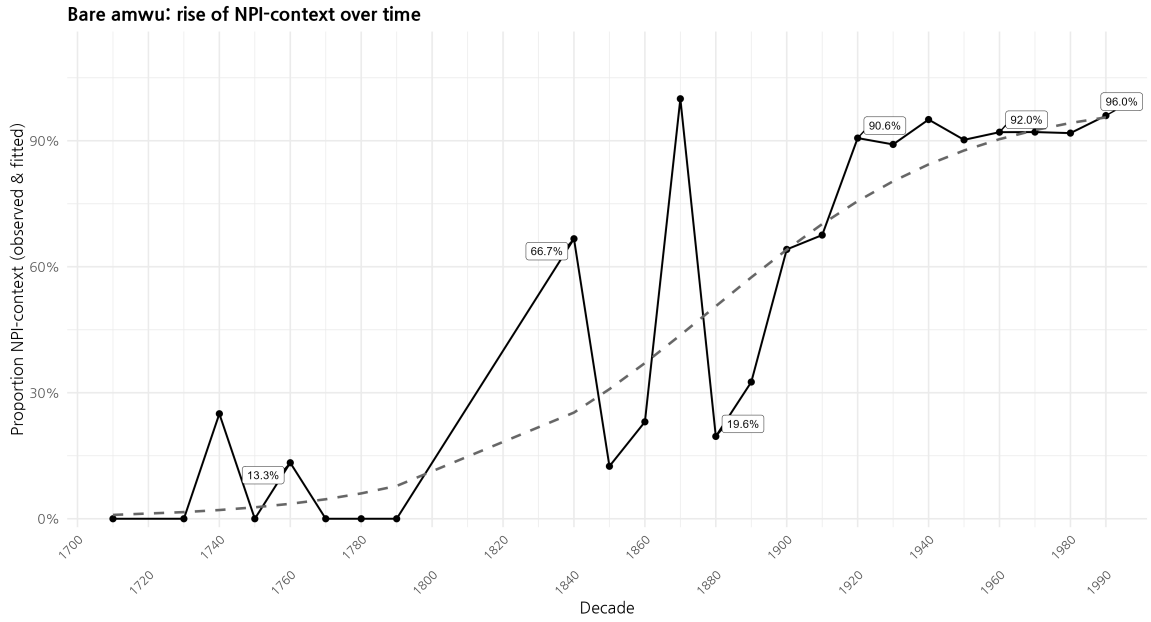


Figure 4.7 Observed and fitted proportions of NPI contexts for bare *amwu* by decade (logistic regression).

To identify when the steepest part of this change occurs, I also fit a segmented logistic regression with a single breakpoint (§3.3.3). The segmented model places the breakpoint in the late 1920s (estimate $\psi \approx 1928$, SE ≈ 4.7; Figure 4.8). Before this point, the slope on the decade predictor is relatively large (about 0.055), which matches the sharp rise in NPI shares between the late 19$^{\text{th}}$ century and the early 20$^{\text{th}}$ century. After the late 1920s, the slope drops to a much smaller positive value (about 0.015), which reflects the fact that NPI proportions are already close to ceiling and can only increase slightly.

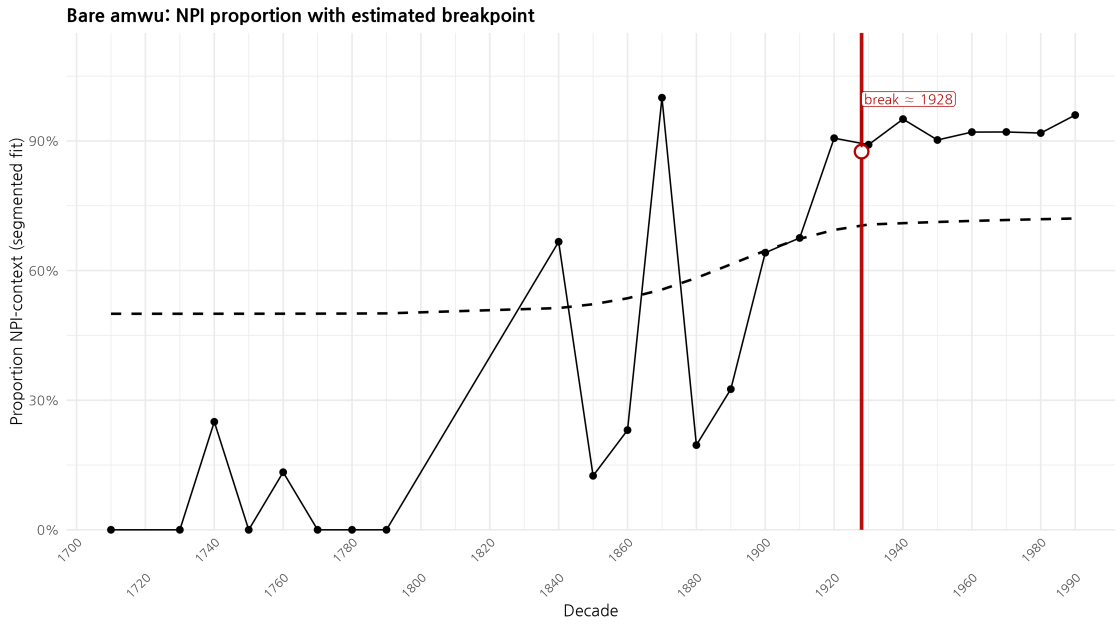


Figure 4.8 Segmented logistic regression for NPI vs. non-NPI uses of bare *amwu*, with estimated breakpoint (vertical line).

In short, the modelling shows a clear shift along Jäger's (2010) polarity cline from existential to negative uses. Until the mid-19th century, *amwu* is used mainly in positive and NEUTRAL contexts, with NPI uses forming at most a small minority. Between the late 19th and early 20th centuries, NPI uses quickly become the majority, and by around 1930 bare *amwu* is strongly linked to NPI contexts: about nine out of ten tokens are NPI, and this proportion stays at or above that level for the rest of the 20th century. Non-NPI uses survive only as small remainders.

More into negative contexts: AA vs. DN

Within the negative contexts (NPI category), there are three subtypes—AA, AM, and DN—ordered by licensing strength (§A.11.1). For bare *amwu*, AM is essentially absent in the historical data, and almost all NPI tokens are coded as either

AA or DN. AA is the broadest subtype (anti-additive environments, including lexical negation and similar set-subtraction contexts), and DN is the most restricted (clause-mate sentential negation).

Figure 4.9 plots the decade-wise per-million frequencies of AA- and DN-licensed *amwu* NPIs, and Figure 4.10 shows their proportions within the NPI set. Both AA and DN appear early. The earliest AA-type NPI uses date from the 1740s (2 AA tokens, 31.9 pmw), and the first DN example appears in the 1760s (1 DN token and 1 AA token, each about 7.2 pmw). After this point, both subtypes continue to be used, but AA quickly becomes the more frequent one. In the 1890s, for example, 41 out of 56 NPI tokens are AA (about 73%) and 15 are DN (about 27%); in the 1920s, 216 out of 242 NPI tokens are AA (about 89%) and 26 are DN (about 11%); and in the 1930s, 245 out of 287 NPI tokens are AA (about 85%) and 42 are DN (about 15%). The same pattern continues through the 20th century: in the 1950s, 330 AA vs. 39 DN (about 89% vs. 11%); in the 1970s, 2146 AA vs. 316 DN (about 87% vs. 13%); and in the 1990s, 4925 AA vs. 1133 DN (about 81% vs. 19%).

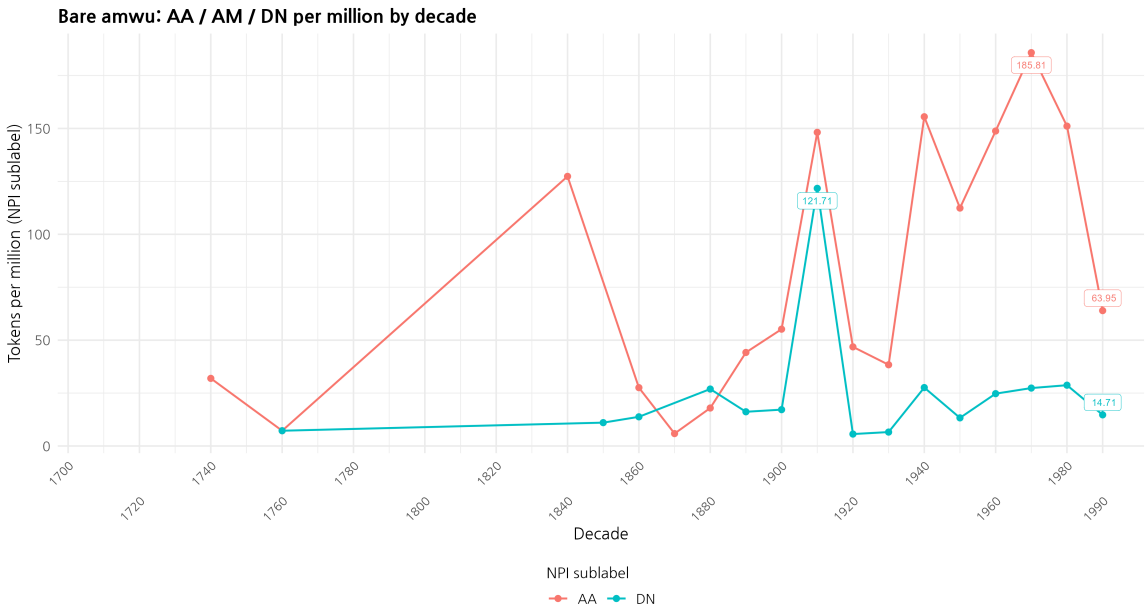


Figure 4.9 Per-million-word frequency of AA and DN NPI subtypes for bare *amwu* by decade.

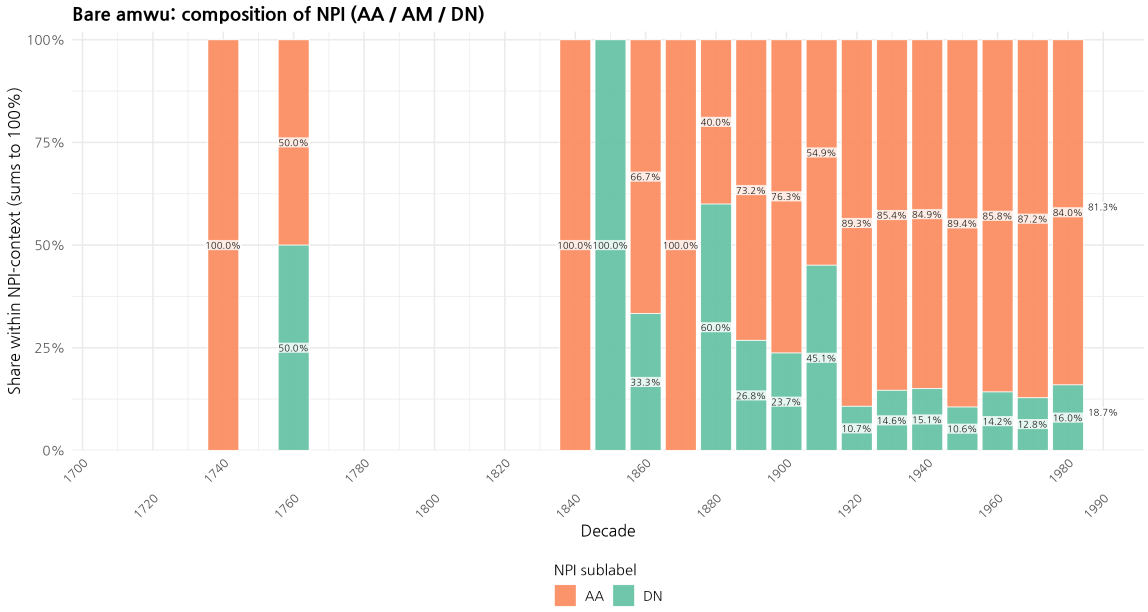


Figure 4.10 Stacked proportions of AA and DN within NPI contexts for bare *amwu* by decade.

In normalized terms, AA-licensed *amwu* reaches about 150–190 pmw in the mid-20th century (for instance, about 156 pmw in the 1940s and 186 pmw in the 1970s), whereas DN-licensed *amwu* rarely exceeds about 25–30 pmw (about 27.6 pmw in the 1940s and 27.4 pmw in the 1970s). Over the whole 1700–1999 dataset, AA licenses 12,816 NPI tokens (roughly 84% of all NPI uses of *amwu*), and DN licenses 2522 tokens (about 16%).

From the perspective of the semantic map, this means that when *amwu* narrows into the negative region, it does so mainly via the broad AA node: anti-additive environments, especially those involving lexical negation, form the centre of its NPI use. Clause-mate sentential negation (DN) provides a smaller area inside that space, and genuinely antimorphic contexts (AM) hardly figure in the data. Combined with the overall rise of NPI usage described in §4.3.3, these results suggest that *amwu* enters and fills the NPI node in an orderly way: first by building up in the most inclusive anti-additive environments and only secondarily by extending into the strongest, clause-local negative configurations.

4.3.4 Interim conclusion for bare *amwu*

The corpus results clarify the diachrony of bare *amwu* in two main respects. First, they show that *amwu* follows a narrowing path: it starts as a widely used existential indefinite and ends up as a strong NPI. Second, they show that this narrowing follows a particular route on the semantic map and is closely connected to the forms of negation used in Korean.

In the Early period (1700–1849), bare *amwu* behaves like a general existential indefinite. About 76% of tokens appear in PPI contexts and roughly 10% in FREE-CHOICE contexts, with NPI uses making up only about 6%. The root appears in positive statements, questions, and conditionals without an overt negative licensor, and simply means ‘someone / some N / any N’. On Haspelmath’s map, *amwu* is

therefore placed in the existential and free-choice area, with only a small connection to the negative-polarity node.

From the Mid period (1850–1899) onward, the distribution shifts steadily toward NPI contexts. NPI uses rise to about 29% of tokens and PPI drops to about 49%. By the Late period (1900–1999), NPI contexts account for about 93% of all *amwu* tokens, with PPI and FREE-CHOICE reduced to 4.5% and 2.5%, respectively. The logistic and segmented regressions in §4.3.3 show that this change is systematic, with a clear bend in the late 1920s and NPI shares staying near or above 90% from the 1930s onward. On Jäger's (2010) polarity cline, *amwu* thus moves from the positive or neutral end to the strong NPI end, and the shift is essentially complete by the early 20th century.

Within the NPI macro-category, bare *amwu* is not evenly spread across negative environments. Almost all NPI tokens are coded as either AA or DN, with AM essentially absent. AA (anti-additive environments, including lexical negation and related set-subtraction contexts) is the main subtype: across 1700–1999, AA licenses about 12,816 NPI tokens (roughly 84% of all NPI uses of *amwu*), while DN (clausemate negation) licenses about 2522 tokens (about 16%). In most decades with substantial data, AA accounts for 80–90% of NPI tokens, with DN making up the remaining 10–20%. Normalized frequencies show the same contrast: AA-licensed *amwu* reaches about 150–190 pmw in the mid-20th century, whereas DN-licensed *amwu* rarely exceeds about 25–30 pmw. From a map point of view, this means that when *amwu* enters the NPI node, it first becomes firmly established in the broad AA part of the negative space and only later extends into the more restricted DN subnode.

The same pattern appears if we look at types of negation. In the Early period, all NPI uses of bare *amwu* are licensed by lexical negation (LEX). In the Mid period, LEX still licenses about 83% of NPI tokens, with Long Form negation (LFN) at

around 16% and Short Form negation (SFN) at about 1%. In the Late period, about 90% of all NPI tokens occur under lexical negation, with SFN and LFN together accounting for only about 10%. Within LEX, a large share of tokens involve the predicate *-eps-* 'not exist, be without', so that many *amwu* NPIs in the 20th-century data have the structural shape [*-eps-* + *amwu* XP].[1]

In short, the diachronic picture is that bare *amwu* moves from the existential and free-choice area of Haspelmath's map into the NPI node. Within that node, its NPI uses concentrate in the broad AA space, especially under lexical negation with *-eps-*, with DN and clause-level negation forming a smaller, secondary niche. Earlier existential and FREE-CHOICE uses survive only as small traces at the edge of this negative cluster. In the next section, I compare this narrowed, lexically anchored NPI behaviour of *amwu* with the simultaneous but opposite development of bare *nwukwu*.

4.4 Comparing the two roots: narrowing vs. expansion

The separate case studies of bare *amwu* and bare *nwukwu* can now be brought together. Figure 4.11 visualizes their joint distribution on the semantic map across the three macro-periods, using a dot-density representation.

[1] LEX here refers to lexical negative predicates such as *-eps-* 'not exist, be without'. SFN (short negation) is preverbal *an/ani* (S-Neg), and LFN (long negation) is the auxiliary pattern *V-ci anh-/V-ci ani ha-* (L-Neg). Previous work (e.g. Nam 2021) has argued that S-Neg and L-Neg differ in distribution, information structure, and diachrony. The present results suggest that this functional split in the negation system also shapes how NPIs such as *amwu* are recruited and specialized in negative contexts, even though a full account of the interaction must be left to future research.

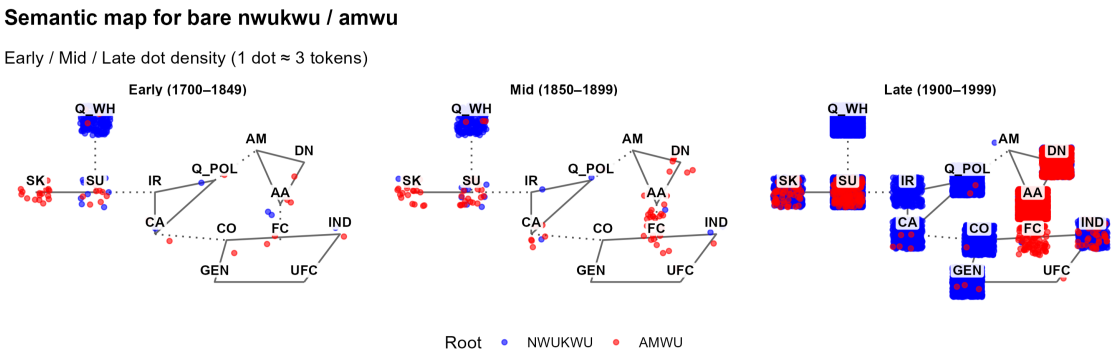


Figure 4.11 Semantic map for bare *nwukwu* and *amwu* across the Early, Mid, and Late periods. Red dots represent *amwu* tokens, blue dots represent *nwukwu* tokens; each dot corresponds to approximately three tokens.

In the **Early panel** (1700–1849), blue dots (*nwukwu*) are almost entirely concentrated at the Q_WH node, with a small spill-over into the adjacent SU and IR nodes. Red dots (*amwu*) cluster in the SK/SU area and spread into IR, FC, GEN, and IND, with only a few tokens in AA and DN. This matches the counts discussed above: *nwukwu* is essentially a question word, while *amwu* is an existential or free-choice indefinite with only a light negative edge.

In the **Mid panel** (1850–1899), we see the crossover. The blue area for *nwukwu* begins to spread from Q_WH into SU, IR, GEN, and IND, reflecting the growth of PPI and NEUTRAL uses. The red area for *amwu* thickens in AA and FC and starts to move away from the SK/SU region: more tokens appear in AA and DN, and fewer in the positive and neutral nodes. This is the overlap stage: *nwukwu* is starting to fill the existential and free-choice space just as *amwu* is moving out of it into the negative region.

In the **Late panel** (1900–1999), the specialization is clear. Blue dots are spread over Q_WH, SU, IR, GEN, IND, and FC, with only a thin band in AA and DN. Red dots are concentrated in AA, FC, and DN, with very few remaining in SK/SU or IR. Visually, *amwu* has moved into the NPI corner (AA/DN, with FC nearby), while

nwukwu occupies the interrogative node plus a broad, connected set of existential and free-choice nodes. The division of labour is visible on the map: one root is anchored in the negative cluster, the other in the interrogative–indefinite cluster.

This figure gives a compact semantic-map summary of the main points in the chapter. It shows that (i) change follows neighbouring nodes on the map (there are no jumps across distant nodes), (ii) *amwu* narrows from existential or free-choice to NPI, mostly through AA environments with lexical negation, and (iii) *nwukwu* expands from Q_WH into existential and free-choice territory without becoming a strong NPI. The next chapter asks how these root-level patterns are carried over and extended in the particle-bearing paradigms built on *nwukwu* and *amwu*.

4.5 Discussion: semantic map, polarity cline, and two-way specialization

The two roots show a clear split on Haspelmath's semantic map and on Jäger's polarity cline. The results support a strong local-pathway view of change, but they also invite a more precise answer to three broader questions raised by the reviewer comments: what kind of prediction the map actually supports, why skewed input does not simply produce stasis, and what the modern configuration suggests about likely future developments.

Local movement on the map, but not perfect symmetry

On the semantic map, the division of labour is straightforward at the level of dominant functions. Bare *amwu* starts out spread across the existential and free-choice area, with only a small foothold in the negative region, and then narrows into the NPI node. Inside that node, its NPI uses are concentrated in the AA part of the negative space, especially in environments with lexical negation such as *-eps-*, with DN-type clausal negation forming a smaller fringe. Bare *nwukwu*, by contrast, starts at

the interrogative node and expands into adjacent existential and free-choice nodes, while only ever making a modest and numerically minor move into NPI territory.

At the same time, the chapter also shows that the semantic map should not be understood too rigidly. The dominant trajectories are local and contiguous, but not every residual use fits a perfectly linear story. Bare *nwukwu* retains a thin negative edge in late-period data, and bare *amwu* retains a small free-choice residue even after NPI uses have become dominant. These are not violations of the map; rather, they show that the map predicts *plausible directions of movement*, not a perfectly clean elimination of all overlap or residue. The Korean evidence therefore supports a constrained but non-deterministic view: the map predicts which neighbourhoods are locally accessible, while usage, source constructions, and competition determine how fully those neighbourhoods are occupied.

Why skewed input does not simply produce stasis

One natural objection is that if each root is repeatedly encountered in one part of the map, then skewed input should simply reinforce the status quo. Why, then, does skewed input yield change rather than only stasis?

The answer is that skewed input can do both. When the distribution of evidence is stable, it reinforces an existing analysis and yields stasis. But when the distribution itself is being reshaped—for example, because a competing form expands into neighbouring functions or because one set of constructions becomes more common than another—learners are exposed to a changing balance of evidence. Under those conditions, skewed input no longer preserves the old analysis; it favors a new one.

This distinction is visible in both roots. For *amwu*, early learners see a form that is available in many kinds of non-negative contexts, but over the 19th and early 20th centuries they increasingly encounter it in strongly negative environments. The NPI statistics show that this is not just 'any negative environment': most NPI uses

are AA-type, and within those AA instances, most are licensed by lexical negation, very often *-eps-* 'not exist, be without'. In 20th-century input, a typical token of bare *amwu* is an *amwu* phrase in the scope of *-eps-* or a similar AA licenser, with clause-mate negation (S-Neg/L-Neg) contributing only a secondary layer of NPI uses. From a learner's point of view, it becomes increasingly natural to treat *amwu* as 'the element that goes with strong negation', and the narrowing into the NPI node on the map is the lexical reflection of that skewed input.

For *nwukwu*, the starting point is different but the logic is similar. The root is initially stored as an interrogative. Embedded questions, rhetorical questions, and conditionals already provide environments where a *who*-variable can be interpreted as 'someone'. As such uses accumulate, the learner gains evidence for an analysis in which *nwukwu* introduces a human variable whose interpretation (interrogative or indefinite) depends on the larger environment. The logistic change-point on the Q_WH vs. NON_Q split shows that this mixed usage stabilizes in the 19th century, and the licensing counts show that the new indefinite uses are linked to PPI and NEUTRAL contexts rather than to negative ones. In other words, skewed input first drives change by redistributing evidence, and then drives stasis again by entrenching the new allocation.

Directionality and the question of reversal

The trajectory of *amwu* also raises a further question: why does the shift appear effectively one-directional? If *amwu* can move from a broad indefinite toward a strong NPI, why should it not shift back?

The safest answer is historical rather than absolute. The present chapter does not show that a reverse shift is impossible in principle. What it does show is that, once *amwu* has become strongly entrenched in AA/DN environments, the corpus offers no evidence of renewed pressure toward broad existential use. A reverse

shift would require not only sporadic positive tokens, but a robust change in the distribution of evidence: new source constructions in which bare *amwu* regularly appears in positive or irrealis existential uses, and enough repeated exposure for learners to reanalyze that pattern as normal. Nothing in the present data suggests such a development. The apparent one-directionality of the shift is therefore best understood as a consequence of entrenchment plus absence of countervailing sources, not as a formal ban on reversal.

What the modern state predicts next

The modern distribution also allows a limited forward-looking inference. By the Late period, bare *amwu* is not simply 'negative' in the abstract; it is anchored above all in AA-type negative environments with lexical negation, while retaining only small residues in FREE-CHOICE territory. This means that the most plausible future continuation, if the same pressures remain in place, is not a return to broad existential use but further consolidation of the negative profile. More concretely, the current configuration suggests three plausible expectations:

1. residual bare free-choice uses of *amwu* may continue to erode, with free-choice work increasingly carried by particle-bearing paradigms rather than by the bare root;

2. the strong association between bare *amwu* and lexical negation, especially *-eps-*, may remain the most stable part of the system;

3. any renewed broadening of bare *amwu* would require a new source of positive or irrealis uses, not just isolated exceptions.

For *nwukwu*, the modern state suggests a different sort of stability. The root has already expanded well beyond the pure interrogative node, but it still remains anchored there. This makes further drift into the strong NPI sector unlikely on

present evidence. A more plausible continuation is continued stabilization of its mixed interrogative–indefinite profile, with modest activity in adjacent existential and free-choice zones but no deep move into AA/DN territory.

These forward-looking claims should be understood cautiously. The semantic map has predictive value at the level of *plausible local pathways*, not exact outcomes. The Korean evidence allows us to say that certain next steps are more consistent with the present organization than others, but it does not allow us to predict timing or certainty.

Predictive value and small irregularities

A final question is whether the root-level developments conform completely to the semantic map or whether there are small surprises. The answer is that the broad movements conform very strongly, but there are minor irregularities that are informative rather than problematic.

One is the persistence of a thin NPI fringe for *nwukwu*. This shows that an interrogative root can reach the edge of the negative space without becoming a true NPI root. Another is the persistence of a small free-choice residue for *amwu* even after the NPI profile becomes dominant. This suggests that older functions are not always instantly eliminated when a new specialization takes over; they can survive as low-frequency residue. A third is the internal shape of the negative region itself: *amwu* does not move evenly into 'negation', but specifically into AA first and DN second. This is precisely the kind of refinement that the extended map makes visible.

The broader conclusion is therefore not that 'everything literally conforms' in a mechanically perfect way, but that the dominant changes do follow the lines we would expect on the map. The map predicts local directions of extension and retreat; the Korean data confirm those directions while also showing that residual

uses, skewed frequencies, and competition among neighbouring functions matter for the realized outcome.

4.6 Conclusion

This chapter has traced how the two bare roots *nwukwu* and *amwu* change over time, using Haspelmath's (1997) semantic map and Jäger's (2010) polarity scale as guides. The main question was whether their modern behaviour—one mainly interrogative but also indefinite, the other a strong NPI—can be explained from actual usage patterns in the corpus instead of being fixed in the lexicon from the start. The results indicate that this is indeed possible. In short:

1. Bare *nwukwu* expands from a pure question word to a mixed interrogative–indefinite item, with most non-question uses in positive and neutral contexts.

2. Bare *amwu* narrows from a general indefinite to a strong NPI, mainly under lexical negation and especially with *-eps-*.

3. The two roots end up with a clear division of labour: *nwukwu* covers the interrogative and non-negative indefinite space, while *amwu* concentrates in the strongly negative space.

For bare *nwukwu* 'who', the main change is an expansion. In the Early period (1700–1849), more than 90% of its tokens are coded as Q_WH, and only a small number are non-questions. Over the 19th and 20th centuries, non-question uses grow. The logistic and segmented regressions show that the share of NON_Q tokens increases, with a noticeable change in the 19th century, and then settles into a Late-period pattern where about 40% of tokens are NON_Q. These non-question uses are not mainly negative: they are mostly PPI and NEUTRAL contexts, with a smaller but steady FREE-CHOICE band and only a minority of NPI tokens. On the

semantic map, *nwukwu* moves outward from the question node into nearby 'someone'-type and 'anyone / whoever'-type uses. It does reach some negative contexts, but those never become central. On Jäger's scale, *nwukwu* thus shifts from a pure question word to a mixed interrogative–indefinite item, without becoming an ordinary NPI.

Bare *amwu* follows the opposite pattern: it narrows into the negative region. In the Early period it behaves like a general indefinite: about 76% of tokens are PPI, around 10% are FREE-CHOICE, and only about 6% are NPI. In the Mid period, NPI uses rise to about 29%, and PPI uses drop to about 49%. In the Late period, NPI contexts make up about 93% of all tokens, with PPI and FREE-CHOICE reduced to small remaining shares. The logistic and segmented models place the sharpest rise in NPI use between the late 19th and early 20th centuries, with a change-point in the late 1920s. After that, NPI use is close to ceiling. Inside the NPI set, *amwu* is not evenly distributed: most NPI tokens are in AA-type contexts, especially under lexical negation with *-eps-* 'not exist, be without', while DN-type clausal negation licenses a much smaller part. On the map, *amwu* pulls back from the positive and free-choice area into the negative node, entering mainly through AA environments. On the polarity scale, it moves from the positive or neutral side to the strong-NPI end.

Viewed together, the two roots give a clear split. In the Early period, *nwukwu* is almost only at Q_WH, while *amwu* is spread over positive, neutral, and free-choice zones, with only a light presence in negative contexts. In the Mid period, *nwukwu* begins to appear more often in 'someone / anyone'-type uses, while *amwu* gains more negative uses and gradually leaves part of its earlier positive area. In the Late period, *nwukwu* covers the question node plus a connected band of PPI, NEUTRAL, and FREE-CHOICE uses, with only a thin negative edge, whereas *amwu* is concentrated in AA/DN negative contexts, with only a small remainder outside them. In

both cases, the movements on the map are local: each new function lies next to an old one, and there is no jump across distant nodes.

The chapter therefore supports all three hypotheses in §4.1. (H-ROOT i) is supported: *nwukwu* extends from a question word to a mixed interrogative–indefinite item, with most non-question uses in positive and neutral settings. (H-ROOT ii) is supported: *amwu* changes from a general indefinite to a strong NPI, mostly under lexical negation and especially with *-eps-*. (H-ROOT iii) is also supported: by the Late period, the two roots are functionally complementary. *Nwukwu* covers the question side and much of the non-negative indefinite space, and *amwu* covers the strongly negative space.

More broadly, the chapter supports a constrained but non-deterministic view of semantic maps. The Korean data show that the map is genuinely predictive in one important sense: change proceeds by locally plausible moves into adjacent regions. At the same time, the map does not by itself determine exact timing, rates, or winners. Those outcomes are shaped by the distribution of source constructions, by skewed input, and by paradigm-internal competition. In this respect, the root-level Korean evidence provides both a confirmation and a refinement of the semantic-map approach.

These root-level patterns set the stage for the next chapter. Once bare *nwukwu* has become a mixed interrogative–indefinite word, and bare *amwu* has become a strong NPI, the particle series built on them (wh+-(*i*)*nka*, wh+-*na*, *amwu* NPIs and FCIs) are not starting from nothing. They grow on top of this already split system. Chapter 5 will show how these series inherit, strengthen, or reshape the division of labour between *nwukwu* and *amwu*, and how the same map-based and polarity-based pressures continue to shape the development of the full indefinite paradigm in Korean.

Chapter 5

Grammaticalization of wh-based indefinites

Chapter 4 asked how the bare roots *nwukwu-* and *amwu-* changed their uses over time. This chapter turns from roots to the role of morphology in shaping the Korean indefinite system (Eckardt, 2006; Hopper & Traugott, 1993; Traugott & Dasher, 2002). The focus is on the question particle or complementizer -(*i*)*nka* and on how it is recruited into wh-based indefinites. This is where Puzzle 2 arises: wh-roots combine productively with -(*i*)*nka*, yielding forms such as *nwukwu-inka* or *nwukwunka* 'someone or other', while the parallel combination **amwu-inka* is not attested.

I argue that this asymmetry is not a random gap, but follows from the different diachronic paths of wh-roots and *amwu-*. Using the corpora and labelling introduced in Chapter 3, I show that wh-based indefinites develop out of clausal interrogative structures such as *nwukwu-i-nka?* 'Who is it?' (Kang, 2021; J.-M. Yoon, 2005). Over time, these clauses are reinterpreted as DP-like expressions, so that the whole sequence comes to behave as a nominal indefinite with an epistemic flavour ('someone or other; I do not know who') (Eckardt, 2006; Hopper & Traugott, 1993). On Haspelmath's (1997) map (Degano & Aloni, 2022; Haspelmath, 1997, 2003), this path runs from the QUESTION node into the neighbouring SPECIFIC UNKNOWN and irrealis zones.

A crucial point developed in this chapter is that the analytic and contracted forms do not contribute equally to this development. The analytic form *nwukwu-i-nka* remains predominantly clausal throughout the period covered here, even after some DP uses emerge. In that sense, -(*i*)*nka* continues to function as a clause-typing

element in part of the system. The contracted form *nwukwunka*, by contrast, becomes the main lexicalized DP indefinite. The Korean data therefore do not show a simple replacement of a question marker by a nominal marker. Rather, they show a layered system in which the older clausal function remains visible in the analytic string while a new, DP-level indefinite becomes stabilized in the contracted form.

For *amwu-*, I propose a different story. Chapter 4 showed that by the late 19th century *amwu-* has already given up any interrogative role and acts as a polarity- and free-choice-oriented base, via combinations such as *amwu-to* and *amwu-na*. If wh+-(*i*)*nka* indefinites are fed by interrogative source constructions, then a root that no longer participates in such constructions will not be on the input side of the -(*i*)*nka* path. On this view, the missing **amwu-inka* is the natural outcome of a timing mismatch between the rise of -(*i*)*nka* as an indefinite marker and the earlier specialization of *amwu-* as an NPI/FCI base (Hopper & Traugott, 1993; Traugott & Dasher, 2002).

The logic of this chapter is therefore twofold. First, it traces a concrete clause-to-DP pathway for the wh-series. Second, it explains why a formally parallel *amwu-inka* series fails to arise. Importantly, I argue that the absence of *amwu-inka* should not be reduced to 'overlap avoidance' alone. Transitional overlap is common in changing paradigms. The more precise claim is that by the time the wh+-(*i*)*nka* pathway becomes productive, *amwu* lacks both a viable source construction and a clear stable niche in which a new *amwu-inka* form could differentiate itself from the existing system.

The chapter proceeds as follows. Section 5.1 states three hypotheses about the origin, asymmetry, and functional range of -(*i*)*nka*-forms. Section 5.2 presents overall token counts, showing both the late appearance of wh+-(*i*)*nka* and the complete absence of **amwu-inka*. Sections 5.3–5.3.4 trace the development of the analytic form *nwukwu-i-nka*. Sections 5.4–5.4.2 then turn to the contracted form *nwukwunka*,

which becomes the main lexicalized indefinite. Section 5.5 places these developments on the semantic map. Section 5.7 returns to the absence of **amwu-inka* and argues that it follows from a source mismatch plus a lack of functional niche. Section 5.8 concludes.

5.1 Hypotheses for -(*i*)*nka*

To structure the discussion of -(*i*)*nka*, I adopt three hypotheses about its diachrony and its place in the indefinite system (Haspelmath, 1997; Jäger, 2010). Each one links grammaticalization ideas to Haspelmath's semantic map and to the *nwukwu-/amwu-* split established in Chapter 4.

(**H-INKA i**) **Stagewise interrogative origin.**

Combinations of a wh-root and -(*i*)*nka* (e.g. *nwukwu-i-nka* 'who is it?') begin as clausal interrogatives. Only later are some of these structures reinterpreted as DP-level indefinites of the 'someone or other' type. If this is correct, earlier historical tokens should be genuine questions, and uses with epistemic indefinite meanings should appear only later. Moreover, the analytic and contracted forms should diverge: the analytic form should remain closer to its clausal source, while the contracted form should become the main lexicalized indefinite. On Haspelmath's map, this predicts a path from QUESTION to the nearby SPECIFIC UNKNOWN and irrealis nodes, rather than a direct jump into distant polarity or free-choice functions.

(**H-INKA ii**) **The missing *_amwu-inka_ reflects source mismatch and niche failure, not mere overlap.**

The form **amwu-inka* is expected to be absent. Unlike the wh-series, *amwu-*based indefinites specialise early into negative-polarity and free-choice functions by combining with particles such as *-to* and *-na*. I hypothesise that by

the time the wh+-(*i*)*nka* pathway becomes productive, *amwu-* no longer participates in the interrogative or ignorance-type source constructions that feed the reanalysis. In the corpus, we therefore expect to find no, or virtually no, *amwu* + -(*i*)*nka* tokens. The key claim is not that overlap would in itself block such a form; rather, a new *amwu-inka* would lack both a viable source configuration and a clearly differentiable niche within a paradigm whose SU/IR region is already being filled by wh+-(*i*)*nka* and whose NPI/FC region is already occupied by the *amwu* series.

(**H-INKA iii**) **Functional restriction of the wh+-(*i*)*nka* series.**

The wh+-(*i*)*nka* series is predicted to remain separate from the core NPI and FCI systems. Unlike *amwu-to* (a strict NPI) or *amwu-na* (free choice), wh+-(*i*)*nka* indefinites are expected to behave like positive or neutral indefinites. Reflecting their interrogative origins, they should be most at home in questions, positive declaratives, and modal or irrealis contexts, and much less at home in negative environments. On the semantic map, their uses should remain within the connected region linking QUESTION, SPECIFIC UNKNOWN, and IRREALIS, with at most a thin negative edge and no systematic spread into free-choice or generic zones.

If these hypotheses are on the right track, the Korean indefinite system shows a clear division of labour: -(*i*)*nka* gives rise to wh-based indefinites specialised for ignorance and epistemic readings, while the *amwu*-series (via particles such as *-to*, *-na*, -(*i*)*lato*) covers negative and free-choice uses.

The rest of the chapter tests these predictions against corpus data. Section 5.2 checks the morphological asymmetry (H-INKA ii) using token counts. Sections 5.3.1 and 5.3.2 follow the development of *nwukwu-i-nka* from clause to DP

(H-INKA i). Section 5.5 places these stages on the semantic map. Finally, Section 5.7 returns to the system level and asks why the wh-series remains confined to a relatively narrow functional area (H-INKA iii), while the *amwu*-series is entrenched elsewhere on the map.

5.2 Overall distribution: the wh+-(*i*)*nka* series

Before turning to the internal development of the wh+-(*i*)*nka* series, it is useful to compare how often -(*i*)*nka* combines with wh-roots and with *amwu-* in the historical corpora. In the *nwukwu* material there are no wh+-(*i*)*nka* tokens before the 1890s. The first instance appears in the 1890s (one token), followed by 13 tokens in the 1900s. From 1910 through 1999, the corpus contains 15,808 tokens of *nwukwu*-based -(*i*)*nka* forms, including both the analytic *nwukwu-i-nka* and the contracted *nwukwunka*. A parallel search in the *amwu* material yields no instances of *amwu-inka* in any period.

For ease of presentation, Table 5.1 groups the *nwukwu*+-(*i*)*nka* tokens into three broad periods aligned with the corpus's decade structure.

Table 5.1 Tokens of [root-*inka*] forms by period (1700–1999).

Form	**1700–1889**	**1890–1909**	**1910–1999**
nwukwu+-(*i*)*nka*	0	14	15,808
amwu-inka (—)	0	0	0

Wh-root + -(*i*)*nka* is thus entirely absent up to 1889, rare in 1890–1909 (14 tokens), and overwhelmingly a 20th-century pattern (15,808 tokens after 1910). In other words, the attested history of wh+-(*i*)*nka* as an indefinite belongs almost entirely to the modern period. By contrast, *amwu* + -(*i*)*nka* is unattested at all times, despite the large number of *amwu*-based indefinites documented in Chapter 4. This corpus result matches contemporary judgements that **amwu-inka* is not a natural form.

These findings confirm the morphological asymmetry in H-INKA ii: in the period covered by the corpora, -(*i*)*nka* combines productively only with wh-roots. In what follows, I focus first on the analytic sequence *nwukwu-i-nka*, where *-i* is the copula and -(*i*)*nka* is a clause-typing element. Following earlier work, I treat *-i* as a copula throughout; the innovation lies in the reanalysis of the wh+copula+-(*i*)*nka* construction, not in a change to the copula itself. Having established the global asymmetry between wh- and *amwu*-bases, the next section examines how the *nwukwu-i-nka* construction changes over time.

5.3 *Nwukwu-i-nka*: clause first, DP later

I begin with the historical usage of the analytic form *nwukwu-i-nka*, which most clearly preserves the internal make-up wh-root + copula *-i* + *-nka*. Before turning to the contracted indefinite *nwukwunka*, it is useful to see how this analytic sequence behaves on its own.

Figure 5.1 plots the functional distribution of the *nwukwu-i-nka* group by decade. In the earlier part of the record, *nwukwu-i-nka* is used only in questions. Up through the 1910s, 1920s, and into the 1930s, every attested instance is annotated as either a direct question (Q_matrix) or an embedded question (Q_embedded), never as a nominal indefinite. From the turn of the 20th century through the first few decades of the century, then, *nwukwu-i-nka* functions as a pure interrogative expression: *-i* is a copula, *-nka* is a clause-typing element, and the combination sits in the C-domain of a clause.

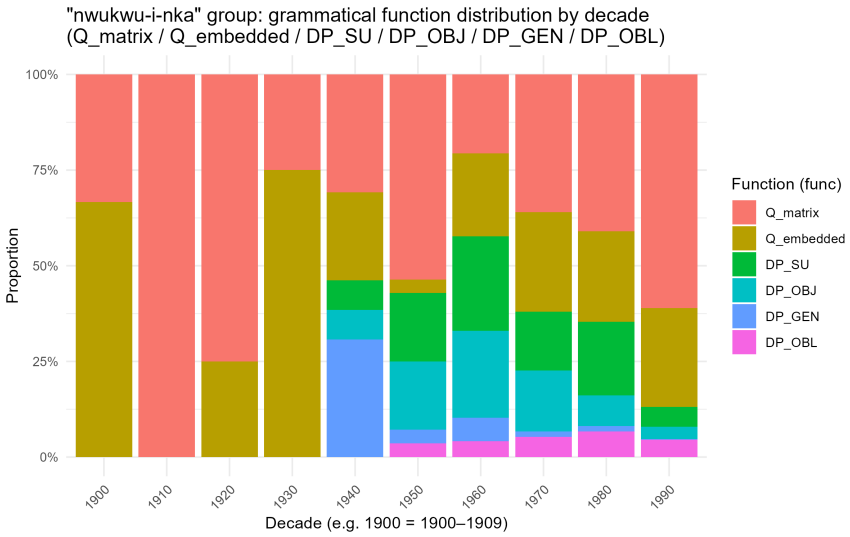


Figure 5.1 Functional distribution of *nwukwu-i-nka* by decade (Q_matrix, Q_embedded, DP_SU, DP_OBJ, DP_GEN, DP_OBL).

5.3.1 Clausal interrogative uses of *nwukwu-i-nka*

In the 1900s, for example, *nwukwu-i-nka* is used in direct questions about a person's identity. A 1900 newspaper text asks who the 'sun-thief' is, as in (1):

(1) Hayss-tocek-i nwukwu-i-nka?
sun-thief-NOM nwukwu-INKA
'Who is the (sun-)thief?'
[[Q (matrix), 1900, YONSEI press]]

Here *nwukwu-i-nka* is the predicate of a matrix interrogative clause; it does not behave like a DP 'someone', but simply marks a question about identity.

Embedded uses work in the same way. A 1930s literary text has a character who wants to know who the visitor is, using *nwukwu-i-nka* as the complement of *al-* 'know', as in (2):

(2) Sonnim-i nwukwu-i-nka al-ko siph-e-ha-yess-ta.
guest-NOM nwukwu-INKA know-CONN want-do-PST-DECL
'She wanted to know who the visitor was.'
[[Q (embedded), 1930, YONSEI fiction]]

In examples of this kind, *nwukwu-i-nka* introduces a wh-clause ('who it is') that serves as the internal argument of *al-* 'know'. There is no sign at this stage that *nwukwu-i-nka* is parsed as an indefinite DP 'someone'; it is a clausal interrogative expression. On Haspelmath's map, these uses are located at the QUESTION node and do not yet reach the SPECIFIC UNKNOWN or irrealis nodes.

This is also the point at which the reviewer question about clause marking can be answered most directly. Yes: in the analytic form, *-(i)nka* still serves as a clause-typing marker. The emergence of DP uses later in the record does not erase that earlier function; rather, it creates a layered system in which the same morphologically transparent sequence can still behave as a question clause in one part of the grammar while feeding lexicalized indefinite uses elsewhere.

5.3.2 Emerging DP uses in non-question roles

From the mid-20th century onward, *nwukwu-i-nka* begins to appear in clear nominal positions. Table 5.2 lists the earliest attestations of *nwukwu-i-nka* in each major DP role (subject, object, genitive, oblique) in the corpus.[1]

According to Table 5.2, the 1940s mark a turning point: this is the decade when *nwukwu-i-nka* first appears as a DP in subject, object, and genitive roles. A magazine story, for example, has *nwukwu-i-nka-ka* as the subject of a clause, as in (3):

[1]The earliest oblique example is *nwukwu-i-nka-ey* 'at/to someone', attested in the early 1950s.

Table 5.2 First attested use of *nwukwu-i-nka* as a DP in various grammatical roles.

Function (DP role)	Earliest example (year)	First decade observed
Subject (DP_SU)	1940	1940s
Object (DP_OBJ)	1940	1940s
Genitive (DP_GEN)	1940	1940s
Oblique (DP_OBL)	1950	1950s

(3) Nwukwu-i-nka-ka mitaci-lul pelketok yel-ko
nwukwu-INKA-NOM sliding.door-ACC suddenly open-CONN
tuleo-wa-ss-ta.
enter-come-PST-DECL
'She was startled when someone suddenly opened the door and came in.'
[[DP (subject), 1940, YONSEI magazine]]

Here *nwukwu-i-nka-ka* is understood as an indefinite subject 'someone' who performs the action. This is a genuine DP use, coded as DP_SU.

In the same decade, *nwukwu-i-nka* also appears in object position. A 1940 literary text contains a line where the protagonist wants to ask someone about his real parents, using *nwukwu-i-nka-lul* as a direct object, as in (4):

(4) Cinsaymi nwukwu-i-nka-lul nal-nal-i mwut-ko
real.parent-NOM nwukwu-INKA-ACC day-day-ADV ask-CONN
siph-ess-ta.
want-PST-DECL
'At times he wanted to ask every day whether someone might be his real parent.'
[[DP (object), 1940, YONSEI fiction]]

We also see *nwukwu-i-nka-uy* used genitively in the same period. A 1940 historical narrative describes a pro-foreign figure killed by 'the sword of someone' (*nwukwu-i-nka-uy khal*), as in (5):

(5) Cinhoy-pha nwukwu-i-nka-uy khal-ey cwuk-e
pro.foreign-faction nwukwu-INKA-GEN sword-DAT die-CONN
ppally-e-ss-ta.
fall-PST-DECL

'He was killed by the sword of someone from the pro-foreign faction.'

[[DP (genitive), 1940, YONSEI magazine]]

By the early 1950s, oblique or dative uses also appear. A 1950 reference text discusses how the form of the state depends on who holds the power, using *nwukwu-i-nka-ey* in a dative-like role, as in (6):

(6) Kukkapwollyek-i nwukwu-i-nka-ey kwisok-toy-nya-ey ttala
state.power-NOM nwukwu-INKA-DAT belong-NPST.Q-ADN-LOC follow
kukche-lul pwunlyu-ha-n-ta.
state.type-ACC classify-do-NPST-DECL

'The type of state is classified according to who holds the state power.' $\Rightarrow$ depending on which person x the state power belongs to, the type of state is classified

[[DP (oblique), 1950, YONSEI reference]]

These mid-century examples show that analytic *nwukwu-i-nka* is no longer confined to clausal environments: it can fill DP slots and bear subject, object, genitive, and oblique roles. On Haspelmath's map, this corresponds to an expansion from the QUESTION node into the specific-unknown and irrealis nodes: *nwukwu-i-nka* is

now used to refer to some person whose identity is not specified or not important to the speaker, while it still has its original clausal question use in other contexts.

5.3.3 Case marking on *nwukwu-i-nka*

A further diagnostic concerns the morphological side of this development: when did speakers begin attaching case particles directly to *nwukwu-i-nka*? This gives another way to detect reanalysis as a noun phrase and complements the distributional picture just described. Table 5.3 lists the first appearances of each major case particle in combination with *nwukwu-i-nka*.

Table 5.3 First occurrences of case particles attached to *nwukwu-i-nka*.

Case-marked form	Earliest example (year)	First decade observed
Object – *-lul* (*nwukwu-i-nka-lul*)	1930	1930s
Subject – *-ka* (*nwukwu-i-nka-ka*)	1940	1940s
Genitive – *-uy* (*nwukwu-i-nka-uy*)	1940	1940s
Oblique – *-ey* (*nwukwu-i-nka-ey*)	1950	1950s

The earliest case-marked form is an object-marked *nwukwu-i-nka-lul* in a 1930s text. The sentence is a rhetorical question in which *nwukwu-i-nka-lul* appears clause-finally. In context, the line essentially asks ‘Who is it that counts as the driving force?’, and the construction is still read as interrogative rather than as a straightforward object DP ‘someone (accusative)’, as in (7):

(7) Manju kaythak-uy kongloce-nun
Manju cultivation-GEN benefactor-TOP
nwukwu, cwungkuk kwukmin-eykey
who China people-DAT
mwut-nora manju kaypal-uy
ask-DECL.HON Manju development-GEN
yekkwun-un nwukwu-i-nka-lul
driving.force-TOP who-NKA-ACC

'Who are the benefactors of Manchurian reclamation? I ask the Chinese people: who is the driving force of Manchurian development?' ⇒ rhetorical question with object-marked *nwukwu-i-nka-lul*

[Q (matrix, rhetorical)]

Here *-lul* marks focus within a rhetorical question rather than an ordinary object noun phrase.

The 1940s provide clearer evidence of nominal reanalysis. This is when we first see *nwukwu-i-nka-ka* (subject) and *nwukwu-i-nka-uy* ('of someone') in examples such as (3)–(5). Once forms like *nwukwu-i-nka-uy* 'of someone unknown' appear, it is clear that *nwukwu-i-nka* is being treated as a noun-like unit that can take genitive case—a strong sign of DP status. The subject form *nwukwu-i-nka-ka* likewise behaves as a noun phrase with nominative marking, rather than as a purely clausal question marker. None of this requires us to reanalyse *-i* as a nominal suffix: following previous work, I continue to treat *-i* as a copula and *-nka* as a clause-typing element. What changes is that the complex *nwukwu-i-nka* begins to be parsed as a nominal unit that case can attach to.

Finally, oblique forms such as *nwukwu-i-nka-ey* are attested by the early 1950s, as in (6). This is the last core case form to appear, completing the basic case paradigm

for *nwukwu-i-nka* as a DP. The relative lateness of the oblique is probably not accidental. Subject, object, and genitive uses provide especially central environments for nominal parsing, whereas oblique uses remain compatible for longer with clausal-source analyses and may simply be less frequent in the relevant discourse types. The present chapter cannot decide between frequency and structure as the primary cause, but the chronology is at least consistent with the idea that oblique conventionalizes slightly later than the more central DP roles.

To sum up, the chronological sequence looks as follows. First, *nwukwu-i-nka* is used only as an unmarked clausal interrogative (matrix and embedded Q). Second, case marking appears for the first time in a rhetorical question (*nwukwu-i-nka-lul*), where the form still has a clausal interpretation. Third, starting in the 1940s, we see nominative, accusative, and genitive *nwukwu-i-nka* in clear DP positions, followed by oblique uses in the 1950s. If overt case on *nwukwu-i-nka* is taken as evidence for a Q→DP reanalysis, then this timeline suggests that case marking is initially a rhetorical extension of clausal Q, and only later becomes a regular part of a nominalized *nwukwu-i-nka* used as an indefinite.

5.3.4 Statistical summary: Q vs. DP uses of *nwukwu-i-nka*

The qualitative picture so far is that analytic *nwukwu-i-nka* begins as a pure clausal interrogative and later acquires a limited range of DP uses. To make this contrast clearer, Figure 5.2 plots, by decade, the proportion of *nwukwu-i-nka* tokens tagged as questions (Q_matrix + Q_embedded) versus DPs (DP_SU/OBJ/GEN/OBL), together with a logistic regression fit (see Chapter 3 for modelling details).

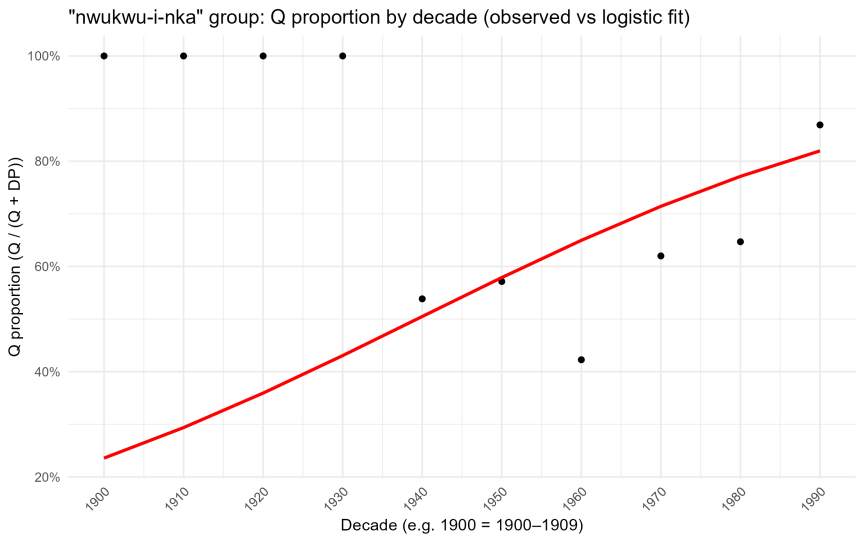


Figure 5.2 Proportion of question vs. DP uses of *nwukwu-i-nka* by decade, with logistic regression fit (Q / (Q+DP)).

The early decades (1900–1930) show only question uses (Q/(Q+DP) = 1.0), in line with a purely clausal analysis. In the 1940s and 1950s the proportion of Q uses drops to roughly 0.54–0.57, reflecting the emergence of subject, object, and genitive DPs documented above. In the 1960s the Q share dips further (0.42), corresponding to a decade in which DP uses of *nwukwu-i-nka* are actually more frequent than Q uses. From the 1970s onward, however, the Q proportion rises again (0.62 in the 1970s, 0.65 in the 1980s), reaching around 0.87 in the 1990s. Over the whole period 1900–1999, about 78% of *nwukwu-i-nka* tokens are annotated as Q and 22% as DP (1557 vs. 433 tokens).

A simple logistic regression of Q vs. DP uses on decade summarises this pattern with a positive slope for decade ($\hat{\beta}_1 = 0.0299$, SE = 0.0039, $z = 7.67$, $p < 0.0001$), which means that, overall, later decades are more likely to contain Q uses than earlier ones. The fitted curve in Figure 5.2 is monotonic and smooths over decade-to-decade fluctuations, but it highlights two points already visible in the raw proportions. First, DP uses of analytic *nwukwu-i-nka* are concentrated in the middle of

the 20th century: they are absent in the early data, become substantial in the 1940s–1960s, and then recede. Second, there is no long-term movement toward a DP-only indefinite: across the century, *nwukwu-i-nka* remains used mainly as an interrogative form, and its DP uses sit alongside, rather than replace, the original clausal function.

In other words, the analytic form remains structurally conservative. It participates in the clause-to-DP development, but it does not itself become the main lexical indefinite. That role is taken over by the contracted form discussed in the next section.

5.4 *Nwukwunka*: a contracted wh+-(*i*)*nka* indefinite

The DP uses of analytic *nwukwu-i-nka* described in the previous section provide a bridge to a further stage of grammaticalization, where wh + -(*i*)*nka* is no longer transparently segmentable. In the contracted form *nwukwunka*, the copular vowel *-i-* is reduced and the whole sequence behaves as a single lexical item. In this section I turn to *nwukwunka* 'someone', which represents a later stage of development, where wh+-(*i*)*nka* has become a lexical indefinite.

The functional tagging in the corpus shows that *nwukwunka* appears as an indefinite much earlier and much more widely than analytic *nwukwu-i-nka*. In total, the corpus contains 13,634 orthographic tokens tagged NWUKWUNKA. Of the tokens with a clear functional label, 9,554 are annotated as subject DPs, 1,159 as object DPs, 1,135 as genitives, and 987 as obliques. Only 488 tokens are matrix questions and 157 are embedded questions. Over 95% of the tagged *nwukwunka* tokens are thus used as DPs, and fewer than 5% occur as interrogatives.

Figure 5.3 summarises the distribution of *nwukwunka* by decade and function, in parallel to Figure 5.1 for analytic *nwukwu-i-nka*.

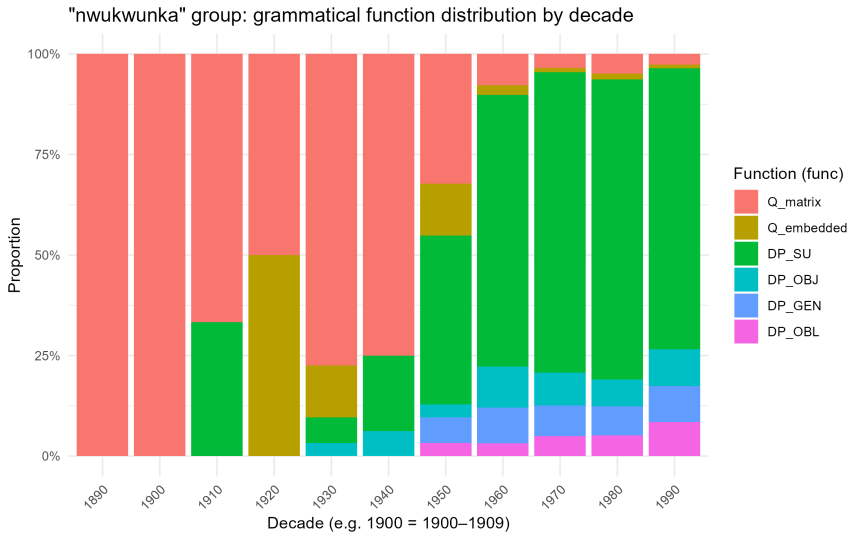


Figure 5.3 Functional distribution of *nwukwunka* by decade (Q_matrix, Q_embedded, DP_SU, DP_OBJ, DP_GEN, DP_OBL).

The earliest corpus instances of *nwukwunka* include embedded questions, which reflect its origin in wh-questions. Example (8) illustrates one such early case:

(8) Kidokto-ka nwukwunka casyehi
Christian-NOM who.it.is closely
salphyepo-mye Kisinhyang-ey sal-nun
examine-while Apostasy.town-LOC live-REL
Payto-i-la-n kes-ul al-kyey toy-ess-ta.
Turn.away-COP-REL thing-ACC realise-CONN become-PST-DECL

‘When Christian looked closely to see who it was, he realized it was Turn-away, who lives in the Town of Apostasy.’ ⇒ embedded-question use of early *nwukwunka*

[Q_embedded]

What matters is not that the contracted form is never interrogative, but that it very quickly becomes overwhelmingly nominal. In contrast to the analytic form, its clausal uses survive only as a small residue.

5.4.1 Emergence and early DP functions

The form *nwukwunka* is best analysed as a fusion of *nwukwu* + -(*i*)*nka* that behaves as a single noun. It is attested in the corpus from the 1890s onward. Table 5.4 lists the first attested DP roles.

Table 5.4 First attested use of *nwukwunka* in various DP roles.

Function (DP role)	Earliest example (year)	First decade observed
Subject (DP_SU)	1910	1910s
Object (DP_OBJ)	1930	1930s
Genitive (DP_GEN)	1950	1950s
Oblique (DP_OBL)	1950	1950s

The earliest subject use of *nwukwunka* as a DP is recorded in a 1910 text, where it is used as a subject referring to some unspecified person. A schematic example is given in (9):

(9) Nwukwunka-ka tochakha-ss-ta.
nwukwunka-NOM arrive-PST-DECL
'Someone arrived.'
[[DP (subject)]]

By the 1930s, *nwukwunka* is also clearly used as an object DP. A schematic example is given in (10):

(10) Hoju-ka nwukwunka-lul po-ss-ta.
head.of.household-NOM nwukwunka-ACC see-PST-DECL
'He saw that the head of household was someone (he did not know).'
[[DP (object)]]

Genitive and oblique uses appear somewhat later in the written record. In the early 1950s we find instances of *nwukwunka-uy* 'someone's X' and *nwukwunka-eykey*

'to someone', showing that *nwukwunka* participates in the full range of NP case frames:

(11) Nwukwunka-uy chayk-ul pilly-ess-ta.
nwukwunka-GEN book-ACC borrow-PST-DECL
'I borrowed someone's book.'
[[DP (genitive)]]

(12) Phyenci-lul nwukwunka-eykey cwue-ss-ta.
letter-ACC nwukwunka-DAT give-PST-DECL
'She handed the letter to someone.'
[[DP (oblique/dative)]]

So *nwukwunka* is used as a DP 'someone' from early on: subject uses are attested by the 1910s, object uses by the 1930s, and genitive and oblique uses by the early 1950s. There is little to suggest that the contracted form was ever a common clausal interrogative; its uses are overwhelmingly nominal, in line with the idea that *nwukwunka* is a lexical Stage 3 indefinite derived from the wh+-(*i*)*nka* construction.

5.4.2 Statistical summary: Q vs. DP uses of *nwukwunka*

The descriptive discussion so far suggests that *nwukwunka* begins with some interrogative uses but soon comes to be used mainly as a DP 'someone'. To make this pattern clearer, Figure 5.4 plots, by decade, the proportion of *nwukwunka* tokens tagged as questions (Q_matrix + Q_embedded) versus DPs (DP_SU/OBJ/GEN/OBL), together with a logistic regression fit (see Chapter 3 for modelling details).

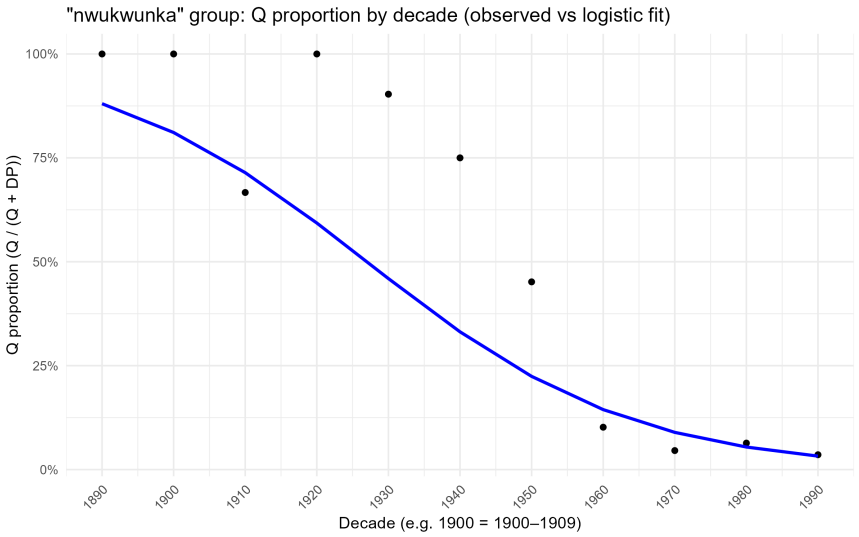


Figure 5.4 Proportion of question vs. DP uses of *nwukwunka* by decade, with logistic regression fit (Q / (Q+DP)).

The earliest decades in which *nwukwunka* appears (1890s and 1900s) show only question uses (Q/(Q+DP) = 1.0). In these decades, *nwukwunka* behaves much like a reduced version of the older interrogative string. In the 1910s, the Q proportion drops to about 0.67, so roughly one third of the tokens are now tagged as DPs, but this is followed by a return to 1.0 in the 1920s and a still high Q share of about 0.90 in the 1930s. Up to this point, interrogative uses still dominate.

From the 1940s onward, the balance starts to shift. In the 1940s, the Q proportion is about 0.75, with a quarter of the tokens now used as DPs. In the 1950s, the Q share falls to about 0.45, so DPs become more frequent than questions in that decade. The 1960s show the strongest change: only about 10% of *nwukwunka* tokens are questions (Q/(Q+DP) $\approx$ 0.10), and around 90% are DPs. The 1970s, 1980s, and 1990s follow the same pattern, with Q proportions of about 0.046, 0.064, and 0.036, respectively: DP readings are now plainly the normal use, and question readings are reduced to a small residue.

A logistic regression of Q vs. DP on decade summarises this trend with a negative slope for decade ($\hat{\beta}_1 = -0.05396$, SE = 0.00320, $z = -16.84$, $p < 0.0001$). The fitted curve in Figure 5.4 smooths over decade-by-decade variation, but it highlights two simple points. First, the earliest decades are dominated by interrogative uses. Second, there is a clear movement toward DP uses: from the mid-20$^{\text{th}}$ century onward, most tokens are tagged as DPs rather than as questions, and from the 1960s onwards DP uses far outnumber Q uses.

The contracted form therefore behaves very differently from the analytic one. Analytic *nwukwu-i-nka* retains its clausal core; contracted *nwukwunka* becomes the main lexicalized epistemic indefinite.

5.5 Haspelmath-map development of *nwukwu-i-nka* and *nwukwunka*

In the previous sections, we saw that *nwukwu-i-nka* and its contracted variant *nwukwunka* shift from purely clausal Q_WH functions to a wider set of DP functions, as shown by the emergence of case marking. I now describe this development in terms of the semantic map. I look first at three non-overlapping periods (1900–1930, 1930–1960, 1960–1990), which show how the two forms spread across nearby map nodes. I then turn to an entry-time analysis that relates the decade in which a function becomes frequent to its distance from Q_WH. This allows us to see more explicitly how far the development respects Haspelmath's continuity constraint and what kind of predictive value the map has in this case.

5.5.1 1900–1930: Q_WH only vs. Q_WH+SU

Figure 5.5 shows the semantic map for 1900–1930. Each dot represents roughly three tokens in this period; blue dots represent *nwukwu-i-nka*, red dots represent *nwukwunka*.

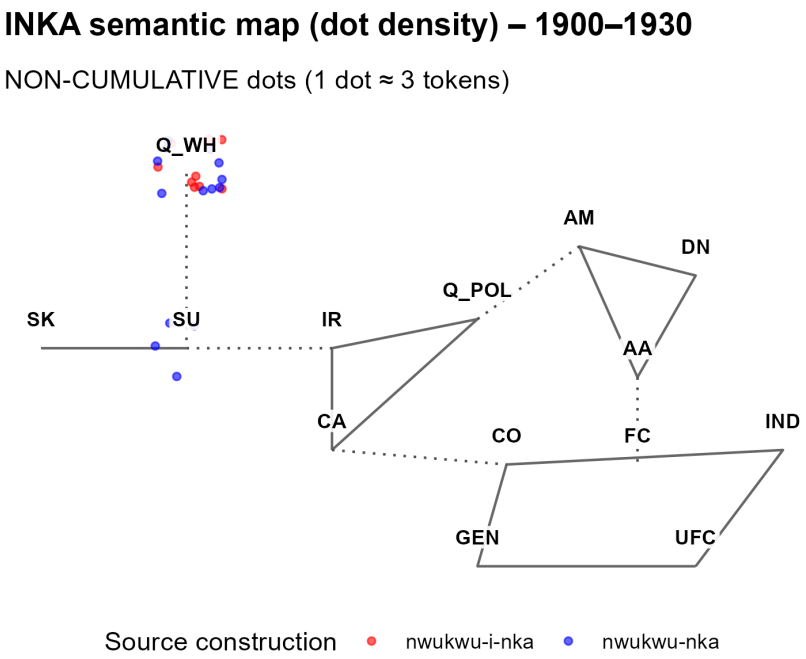


Figure 5.5 Semantic-map locations of *nwukwu-i-nka* (red) and *nwukwunka* (blue), 1900–1930 (non-cumulative). One dot ≈ 3 tokens.

In this earliest slice, *nwukwu-i-nka* appears only at the Q_WH node (26 tokens) and nowhere else. This matches the purely interrogative uses described in §5.3.1. On the map, analytic *nwukwu-i-nka* occupies a single node.

By contrast, *nwukwunka* already spans two nodes. In 1900–1930 it has Q_WH uses (29 tokens) and SU uses (12 tokens). That is, the contracted form is used both to ask 'who is it?' and as a specific-unknown DP, of the sort later realized as 'someone or other'. There are no tokens of either form in IR, CA, DN, AA, or Q_POL in this period. On the map, analytic *nwukwu-i-nka* is still confined to the interrogative node, while *nwukwunka* has already moved one step outward into the adjacent SU node.

5.5.2 1930–1960: SHARED SU AND FIRST IR/CA

Figure 5.6 plots the same information for 1930–1960 (non-cumulative).

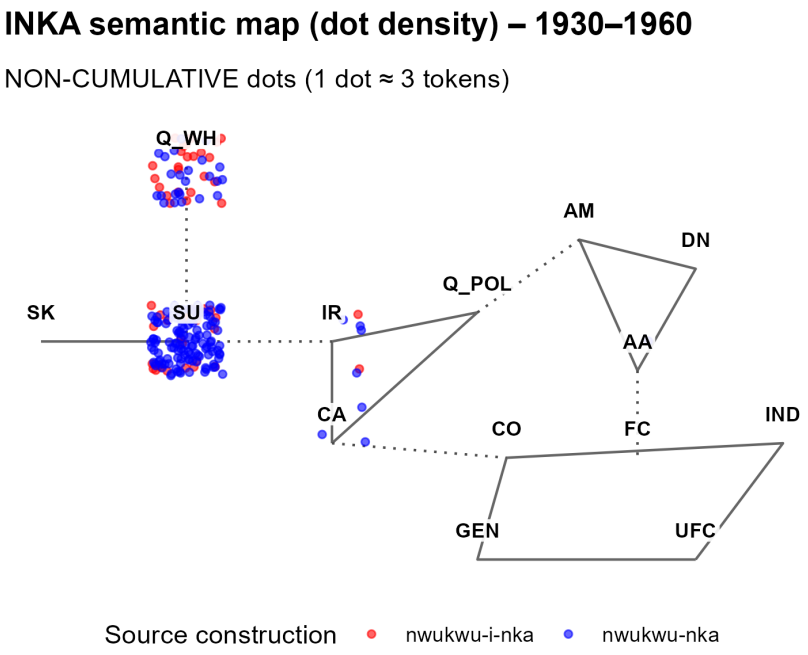


Figure 5.6 Semantic-map locations of *nwukwu-i-nka* (red) and *nwukwunka* (blue), 1930–1960 (non-cumulative). One dot ≈ 3 tokens.

In this middle period, analytic *nwukwu-i-nka* begins to move beyond Q_WH. It now appears in Q_WH (71 tokens), in SU (70 tokens), and in IR (4 tokens), with no tokens in CA, DN, AA, Q_POL, or any of the free-choice and generic nodes. In map terms, analytic *nwukwu-i-nka* has spread from Q_WH one step into SU and two steps into IR, but nowhere else.

The contracted *nwukwunka* continues to occupy the same corner, but with far more tokens. In 1930–1960 it appears in Q_WH (70 tokens), SU (364 tokens), IR (10 tokens), and CA (7 tokens), with a single AA token at the negative edge and no tokens in DN, Q_POL, or the free-choice or generic nodes. So in this period, both

forms live in the Q_WH–SU–IR–CA neighbourhood, but *nwukwunka* is the main realization of SU and is also the only form that clearly reaches CA and AA.

5.5.3 1960–1990: SU/IR/CA filled and a small negative edge

The last slice, 1960–1990, is shown in Figure 5.7.

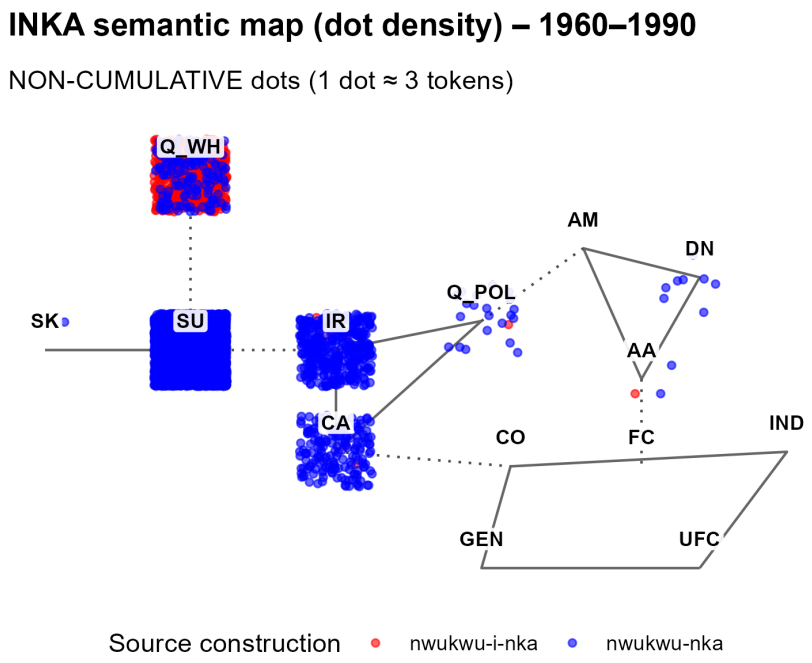


Figure 5.7 Semantic-map locations of *nwukwu-i-nka* (red) and *nwukwunka* (blue), 1960–1990 (non-cumulative). One dot ≈ 3 tokens.

For *nwukwu-i-nka*, the distribution in this period is still centred on Q_WH, with additional points in nearby nodes. It appears at Q_WH (1406 tokens), at SU (463 tokens), at IR (28 tokens), and at CA (6 tokens), plus one token in Q_POL and one in AA. There are no analytic tokens in DN or in any free-choice or generic node. Analytic *nwukwu-i-nka* therefore remains heavily concentrated on the interrogative

node, with a relatively narrow band of non-Q_WH uses along SU, IR, CA, Q_POL, and AA.

For *nwukwunka*, 1960–1990 is the period in which SU-like functions are most frequent. The contracted form appears at SU (11,301 tokens), at IR (1116 tokens), at CA (516 tokens), at Q_WH (434 tokens), and at Q_POL (65 tokens). It also has a small number of SK tokens (2), AA tokens (6), and DN tokens (23), and no tokens in FC, UFC, GEN, or IND. In this period, then, SU, IR, and CA are overwhelmingly realized by *nwukwunka*. The contracted form is also present in Q_WH and Q_POL, and it has a thin negative edge in AA and DN. These negative uses are still glued to the same corner of the map; they do not form a separate free-choice or pure-NPI area.

5.5.4 Examples for SK, IR/Q_POL, AA, and DN

Within the SU–IR–CA cluster, four nodes are especially informative for the semantics of *nwukwunka*: SK, IR, Q_POL, and the negative nodes AA and DN. This subsection gives examples for each, to make the functional labels more concrete. In the interlinear glosses, *nwukwunka* is glossed as *nwukwunka* (with case markers), in a theory-neutral way; English translations use ‘someone’ where appropriate.

The SK node includes cases where the indefinite is specific and anchored to a particular individual known to the speaker (or recoverable in context), even if their identity is not spelled out for the addressee. Example (13) illustrates an SK-tagged *nwukwunka*:

(13)

Wuli-tul	cwung	nwukwunka-uy
we-PL	among	nwukwunka-GEN
mal-ey	ttalumyen	kopsatengi-nun
word-LOC	follow-if	hunchback-TOP
ponti	nayil-ul	mek-ci
originally	age-ACC	eat-CONN
anh-nun	pep-i-la-ko	ha-yess-ta.
not-REL	law-COP-COMP	say-PST-DECL

'According to what one of us said, hunchbacks are originally such that they never age.' ⇒ specific-unknown speaker *nwukwunka-uy* 'someone's'

[SK (specific known)]

Here *nwukwunka-uy mal* 'the words of someone' refers to a particular member of 'us' whose identity the narrator has in mind but does not name. On the map, SK sits next to SU, and both analytic *nwukwu-i-nka* and contracted *nwukwunka* have SK tokens, though *nwukwunka* provides almost all of them in the later period.

IR and Q_POL bring out the irrealis, non-free-choice character of *nwukwunka*. In (14), *nwukwunka-lul* appears in the antecedent of an irrealis conditional:

(14)

Nwukwunka-lul	cwuk-eci	anh-umyen
nwukwunka-ACC	kill-CAUS.CONN	not.do-if
ttappwunhay-se	kyent-eci	mos-ha-nun
be.bored-CONN	endure-CONN	cannot-do-ADN
cewang-i	iss-ess-tamyen	ku
emperor-NOM	exist-PST-IF	that
kwancwung-tul-un	motwu-ka	ku-uy
spectators-PL-TOP	all-NOM	he-GEN
cason-ilnun-ci-to	molu-n-ta.	
descendant-be-MAYBE-ALSO	not.know-NPST-DECL	

'If there had been an emperor who could not endure boredom without killing someone, those spectators might all have been his descendants.' ⇒ irrealis non-specific *nwukwunka-lul*

[IR (irrealis)]

The identity of any potential victim is not relevant; the clause ranges over possible victims. On the map, such tokens live in IR, adjoining SU and CA.

A similar non-specific reading appears in polite yes/no questions (Q_POL), as in (15):

(15) Sensayngnim hoksi i
teacher by.chance these
kuncey-ey nwukwunka-eykey myengham-ul
recent.time-LOC nwukwunka-DAT business.card-ACC
cwusin il-i eps-usi-pni-kka?
give-HON.ADN matter-NOM not.exist-HON-INT.Q

'Professor, have you by any chance recently given your business card to someone?' ⇒ non-specific *nwukwunka-eykey* in a polite yes/no question

[Q_POL]

The question presupposes an unknown addressee of a possible card-giving event. This fits the SU/IR/Q_POL region rather than any free-choice or pure-NPI node.

For negative contexts, it is useful to distinguish AA (anti-additive) from DN (direct negation), as in §A.11.1. In the present data, cases with *eps-ta* 'not exist' and similar lexical negatives are labelled AA. Example (16) shows *nwukwunka-eykey* inside a nominalized predicate *phil-yo* 'need' under *eps-ta* 'not exist':

(16) Cinaka-nun nwukwunka-eykey mwul-ul
pass.by-REL nwukwunka-DAT ask-ACC
phil-yo-ka eps-ta.
need-NMLZ-NOM not.exist-DECL

'When you don't know which way to go, there is no need to ask someone passing by.' ⇒ *nwukwunka-eykey* under lexical negation *eps-ta* 'not exist'

[AA (anti-additive)]

The interpretation is ‘for any passer-by, there is no need to ask that person’. Because the negative operator is a lexical predicate of the anti-additive type, these tokens are coded as AA on the map.

By contrast, DN covers cases where *nwukwunka* appears under clausemate sentential negation (short or long *-anh-*), sometimes with the additive particle *-to*. Example (17) shows *nwukwunka-eykey-to* under *-anh-*:

(17) Kuleh-tako nay-ka pon
like.that-QUOT I-NOM see-ADN
kes-ul nwukwunka-eykey-to nulu-noch-ci-n
thing-ACC nwukwunka-DAT-EVEN tell.out-CONN-REL
anh-ass-ta.
not.do-PST-DECL
‘Even so, I did not tell what I had seen to anyone.’ ⇒ *nwukwunka-eykey-to* under clausal negation *-anh-*
[DN (direct negation)]

Here clausal *-anh-* and *-to* yield a ‘not tell to any nwukwunka’ reading. These DN examples, along with the few AA examples, correspond to the small negative edge on the map in the latest period. Their number is small compared to the mass of SU/IR/CA tokens, so they are best seen as a narrow extension along the edge of the same cluster rather than as a separate negative-indefinite system.

5.5.5 Entry times and distance from Q_WH

The examples above show how SK, SU, IR/Q_POL, AA, and DN differ in meaning. We can now ask when each of these functions becomes frequent enough to count as a stable part of the *nwukwu-i-nka* / *nwukwunka* system, and how this timing relates to their distance from Q_WH. Following §3.3.5, I take the ‘entry decade’ of a label (e.g. SU, IR) to be the first decade in which a form has at least five tokens with that

label. I also assign each label a distance from Q_WH on the semantic map. The core entry decades and distances are summarized in Table 5.5.

Table 5.5 Entry decades (threshold ≥ 5 tokens per decade) and distance from Q_WH.

Form	Label	First decade	Context	Dist. from Q_WH
nwukwu-i-nka	Q_WH	1900	Q_WH	0
	SU	1940	SU	1
	IR	1990	IR	2
nwukwunka	Q_WH	1930	Q_WH	0
	SU	1940	SU	1
	IR	1960	IR	2
	DN	1990	DN	(adjacent to IR)

If we restrict attention to labels whose distance from Q_WH is defined in the semantic map (0 for Q_WH, 1 for SU, 2 for IR), the entry decades display a clear ordering. Labels that are further from Q_WH on the map reach the frequency threshold later in time. This relationship is summarized in Figure 5.8 and quantified by the linear model in (18), where entry decade is predicted by map distance.

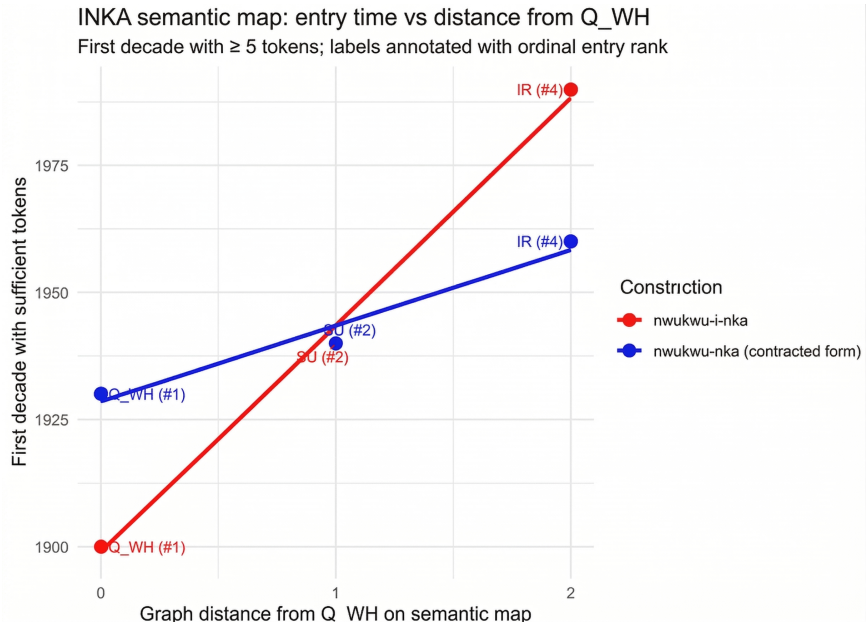


Figure 5.8 Entry decade of labels (threshold ≥ 5 tokens per decade) as a function of distance from Q_WH.

(18)

$$\text{first\_decade}_\ell \approx 1913.33 + 30.00 \times \text{dist\_from\_qwh}_\ell,$$

The estimated slope is positive, about 30 years per additional map step (standard error 7.64, $t \approx 3.93$, $p \approx 0.017$). The model explains a large share of the variation in entry decades ($R^2 \approx 0.79$), indicating that semantic-map distance is a strong predictor of when a function becomes robustly attested. Substantively, the pattern is that SU, one step from Q_WH, becomes frequent roughly one generation after Q_WH, and IR, two steps away, becomes frequent roughly another generation later. The figure therefore suggests a stepwise expansion in which new uses emerge first in nodes that are closest to the interrogative node and only later extend to more distant nodes, rather than appearing in an order-free way.

This is also where the predictive value of the map becomes clearest. The map does not predict the exact decade at which a new function must appear, but it does predict which developments are locally plausible. The entry-time results support

that weaker but important claim: the attested order of expansion tracks distance from the source node rather than skipping to unrelated parts of the space.

5.5.6 Cumulative map and summary

Finally, Figure 5.9 shows the cumulative distribution of *nwukwu-i-nka* and *nwukwunka* on the semantic map for the entire 1900–1990 period.

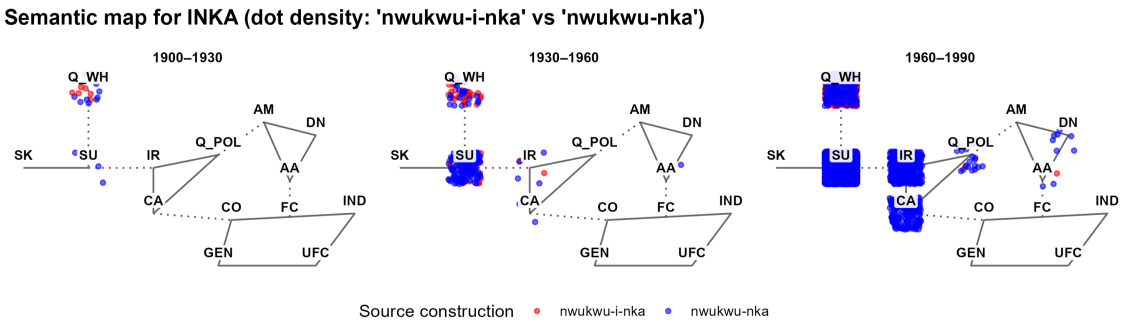


Figure 5.9 Cumulative semantic-map locations of *nwukwu-i-nka* (red) and *nwukwunka* (blue), 1900–1990. One dot ≈ 3 tokens.

Across the century, both forms remain in a single connected region around Q_WH, SU, IR, and CA, with only a narrow AA/DN edge in the later decades. Analytic *nwukwu-i-nka* continues to serve mainly as a clause-level interrogative, with limited extension into nearby SU/IR/CA functions. Contracted *nwukwunka* is the lexical wh+-(*i*)*nka* form that realizes 'someone (or other)' and related irrealis readings in the same region, and it is this form that fills the specific-unknown and conditional nodes in actual usage. The entry-time analysis shows that the way these functions appear over time mirrors the structure of the map: new functions emerge first in nodes that are closest to Q_WH, in line with H-INKA i and H-INKA iii.

The semantic-map results line up neatly with the distributional and case-marking facts from the earlier sections. In time, *nwukwu-i-nka* first appears as a purely clausal interrogative ('who is it?'), used only in matrix and embedded question contexts (Figure 5.1). In the 1930s we see the first case-marked form *nwukwu-i-nka-lul* in a rhetorical question, still interpreted as a wh-clause. From the 1940s onward, nominative, accusative, and genitive *nwukwu-i-nka* occur in clear DP positions, and in the 1950s oblique *nwukwu-i-nka-ey* also appears (§5.3.2–5.3.3). If overt case on *nwukwu-i-nka* is taken as a diagnostic for a Q→DP reanalysis, then this timeline suggests that case marking first attaches to a still-clausal interrogative in rhetorical questions and only later becomes part of a nominalized *nwukwu-i-nka* that can function as an indefinite.

The Q vs. DP statistics show that this nominal reanalysis is real but does not eradicate the interrogative use. Analytic *nwukwu-i-nka* gains DP uses mainly in the mid-20th century (1940s–1960s), yet question uses remain dominant across the century, and there is no shift toward a purely DP-based indefinite (Figure 5.2). On the semantic map (§5.5), this is reflected in the fact that analytic *nwukwu-i-nka* stays close to its interrogative origin: its core is Q_WH, with moderate extensions into SU (and some SK) and a small number of IR and CA tokens. It does not develop free-choice or generic uses.

The picture for *nwukwunka*, the contracted *nwukwu-nka* form, is quite different. As soon as it appears in the corpus, it is used as a DP 'someone': subject uses are attested by the 1910s, object uses by the 1930s, and genitive and oblique uses by the early 1950s (§5.4.1). The functional distribution over time (Figure 5.4) shows a fast change: interrogative uses are common in the early decades but become marginal after the 1950s, and by the 1970s–1990s *nwukwunka* is used almost exclusively as an indefinite. In the map-based view, *nwukwunka* occupies the same corner as analytic

nwukwu-i-nka but fills it very differently: it is densely attested in SU and IR, has substantial CA and some SK, Q_POL, AA, and DN uses, while its Q_WH uses have shrunk to a residue.

The period-by-period maps (Figures 5.5–5.7) and the entry-time analysis (§5.5.5, Figure 5.8) support a simple two-step scenario. First, analytic *nwukwu-i-nka* moves from the Q_WH node into the adjacent SU (and later IR and CA) nodes: clausal 'who is it?' strings are reinterpreted in some contexts as 'someone (of unknown identity)' and begin to accept case marking. Second, the contracted form *nwuk-wunka* stabilizes and generalizes this DP use, becoming the main wh+-(*i*)*nka* indefinite in the modern language. At the same time, the two forms share out the same connected region of the map rather than competing for identical functions. Analytic *nwukwu-i-nka* continues to carry most of the interrogative load, while *nwuk-wunka* bears the main weight of specific-unknown (SU and SK) and irrealis (IR/CA) uses and only marginally participates in AA and DN contexts. This division of labour is exactly what we expect if -(*i*)*nka* has moved from a clause-typing element in 'who is it?' strings to a DP-internal marker of epistemic indefiniteness, and if the lexicalized contracted form is the one that realizes the resulting indefinite series.

5.7 Why is **amwu-inka* missing? Source mismatch, timing, and paradigm structure

We can now return to the question of why **amwu-inka* is missing from the paradigm. Several natural explanations suggest themselves; the corpus data and the timing of changes let us narrow these down.

A first possibility would be that -(*i*)*nka* is lexically restricted to wh-roots. On this view, -(*i*)*nka* would simply select wh-forms and never combine with *amwu*.

This statement matches the modern outcome, but it cannot be the whole explanation. Historically, -(*i*)*nka* is a general interrogative marker, not an inherent indefinite suffix. And in earlier Korean, bare *amo*/*amwu* does appear in configurations that are close to embedded ignorance questions.

All clear embedded interrogative-like uses of *amo*/*amwu* in the corpus are ignorance constructions with *mol-* 'not know'. The earliest example comes from 1713 and already shows *amo* in an embedded question-like environment:

(19) Amo-cen cwul mol-ola.
amo-KIND fact not.know-PST-DECL
'He did not know at all what kind it was.'
[[Q (embedded), 1713, HISTORY]]

A much later example from 1876 has essentially the same shape:

(20) Amo-kin cwul mol-ola.
amo-WHICH fact not.know-PST-DECL
'He did not know at all which one it was.'
[[Q (embedded), 1876, HISTORY]]

After 1876 there are no further embedded-question uses of *amo*/*amwu* in the corpus. Later *amwu* tokens occur as NPI or FREE-CHOICE forms such as *amwu-to* and *amwu-na*. So *amwu* was once compatible with interrogative-like syntax, but that use disappears well before wh+-(*i*)*nka* begins to be used productively on *nwukwu*. A simple 'wh-only selection' statement about -(*i*)*nka* therefore describes the outcome; it does not explain how the system reached that point.

A second possibility is that *amwu-inka* existed but happened not to be attested. Given the size and time depth of the data, this looks implausible. From 1700–1999 the corpus contains more than 46,000 tokens of *amo*/*amwu* and not a single *amwu-inka*, despite targeted searches on the INKA field. By contrast, wh+-(*i*)*nka* takes off

rapidly once it appears, and reaches tens of thousands of tokens in the 20th century. If *amwu-inka* existed at all, it was so rare that it left no trace in three centuries of written material and does not look like a productive member of the paradigm.

The more plausible explanation links the absence of **amwu-inka* to the relative timing of two independent developments. On the wh-side, the first wh+-(*i*)*nka* token with *nwukwu* in the data is from 1895, where *nwukwunka* appears in a context of careful identification. This is exactly the sort of interrogative or ignorance-related environment that later feeds epistemic DP uses. On the *amwu* side, the embedded ignorance uses of *amo*/*amwu* in (19)–(20) disappear by 1876. Chapter 4 (§4.3.3) showed that over the 19th century *amwu* moves far along the polarity cline: earlier positive and interrogative-like uses are lost, and the root becomes heavily concentrated in NPI and emerging FREE-CHOICE contexts.

In other words, when wh+-(*i*)*nka* first appears with *nwukwu* (from 1895 onward) and starts its own clause-to-DP path, *nwukwu* still has a strong interrogative life, but *amwu* no longer does. The only wh-like uses of *amo*/*amwu* that are visible in the corpus are the embedded ignorance configurations with *mol-*, and these vanish a few decades before wh+-(*i*)*nka* is recruited into the *nwukwu* series. By the time -(*i*)*nka* begins to attach productively to wh-forms, there is no productive *amwu*-based interrogative template from which an *amwu-inka* DP could plausibly be reanalysed.

At the same time, the division of labour inside the indefinite paradigm is tightening. Wh+-(*i*)*nka* moves into the specific-unknown and irrealis corner of the semantic map (SU, IR, CA), while *amwu-to* and *amwu-na* anchor the negative and free-choice sectors (AA/DN, FC/IND). Once this pattern is in place, there is little functional space left for a new *amwu-inka* series to occupy.

It is important to formulate this point carefully. Overlap by itself is not the problem. The history described in the earlier chapters already shows that transitional overlap is normal: wh-based forms expand into regions once occupied more

broadly by *amwu*, and for a time the system contains both residue and competition. So the argument cannot be that paradigms simply ban overlap. The more precise point is that a new *amwu-inka* series would need both (i) a viable source configuration and (ii) a stable functional contribution distinct enough to survive. By the time the relevant wh-based source becomes productive, *amwu* no longer participates in that source, and the SU/IR niche is already being filled by the wh+-(*i*)*nka* series. In that sense, the issue is not overlap as such, but overlap *without* either source support or differentiating function.

The corpus evidence therefore points to a simple conjecture. **Amwu-inka* is missing not because *amwu* was inherently incompatible with -(*i*)*nka*, but because by the time the wh+-(*i*)*nka* path opened up for *nwukwu* around the turn of the 20th century, *amwu* had already abandoned its earlier interrogative-like uses and settled into the NPI/FC part of the map. In that state of the system, nothing in the input would favour building a parallel *amwu-inka* series, and the gap in the modern paradigm is exactly what we would expect.

This analysis also suggests where parallel developments should and should not be looked for. If further clause-to-DP lexicalizations of this type are found in Korean, the most plausible candidates are other wh-root + -(*i*)*nka* series, not *amwu*. The relevant generalization is therefore source-type based rather than root-neutral: the Q→DP pathway belongs to the wh-side of the paradigm.

5.8 Conclusion

This chapter has examined how the clause-typing element -(*i*)*nka*, originally a question particle or complementizer, enters the nominal domain to form wh-based indefinites, and how this development relates to the older *amwu*-series. The main points can be organized as follows:

1. The corpus shows that *nwukwu-i-nka* has in fact undergone a Q→DP shift: it starts as a purely clausal interrogative, and only later appears in DP roles with case marking, even though its clausal uses remain dominant.

2. Once the contracted form *nwukwunka* is in place, the two forms specialize in different ways: *nwukwu-i-nka* continues to be used mainly in questions, while *nwukwunka* is widely used as an epistemic indefinite across neighbouring nodes on the semantic map. Neither form develops into a core NPI or FCI item.

3. The absence of **amwu-inka* is best understood in terms of timing, source mismatch, and lack of niche. By the time the wh+-(*i*)*nka* path opens productively for *nwukwu*, *amwu* no longer participates in the relevant interrogative-type sources and is already entrenched elsewhere in the paradigm.

The first point corresponds to H-INKA i and is supported by several independent diagnostics. Analytic *nwukwu-i-nka* appears in the early 20th century only as a clausal expression 'who is it?', in matrix and embedded questions (§5.3); every token from 1900–1930 is tagged Q_WH. There is no early DP use and no affinity for negative or free-choice environments. From the 1940s onward, however, *nwukwu-i-nka* begins to occur in subject, object, genitive, and oblique positions (§5.3.2), and the first case-marked forms appear in the 1930s–1950s (§5.3.3). If overt case on *nwukwu-i-nka* is treated as a sign of nominal reanalysis, the observed sequence is clear: first a purely clausal interrogative, then rhetorical questions with focal case, and finally a limited but stable set of NP-level epistemic indefinites. The modelling in §5.3.4 shows that this shift does not eliminate the original function: across 1900–1999 about four fifths of analytic *nwukwu-i-nka* tokens are still questions, with DP uses peaking around the mid-20th century and then declining. In this respect, the chapter answers one reviewer question directly: yes, -(*i*)*nka* or its analytic reflex is

still serving as a clause marker in the grammar, even after DP uses emerge. On the semantic map (§5.5), analytic *nwukwu-i-nka* remains anchored at Q_WH and extends only into nearby SU/IR/CA, with no free-choice or generic uses.

The second point concerns the contracted form *nwukwunka*. Phonologically it fuses *nwukwu-i-nka* into a single word; functionally it behaves like a lexical epistemic indefinite. Subject uses are attested by the 1910s, object uses by the 1930s, and genitive and oblique uses by the early 1950s (§5.4.1). In the tagged data more than 95% of *nwukwunka* tokens are DPs rather than questions (§5.4.2). On the semantic map, *nwukwunka* occupies the same corner as analytic *nwukwu-i-nka* but fills it very differently: it is heavily attested in SU, SK, IR, and CA, with some Q_POL and a thin AA/DN edge, while its Q_WH uses are few (Figures 5.5–5.9). In this sense, *nwukwunka* realizes in a systematic way what analytic *nwukwu-i-nka* only begins to do in the mid-20th century: it stabilizes the epistemic 'someone' reading without developing into a core NPI or FCI. The analytic form therefore remains closer to the clausal source, while the contracted form lexicalizes the DP indefinite.

The third point, about **amwu-inka*, ties the wh+-(*i*)*nka* development back to the polarity-based changes in Chapter 4 and to H-INKA ii. Earlier Korean does have *amo*/*amwu* in embedded ignorance constructions with *mol-* 'not know' (19)–(20), so in principle *amwu* could have been part of the same Q→DP path. But these uses disappear by the late 19th century. When *nwukwunka* appears in 1895 and wh+-(*i*)*nka* starts to grammaticalize with *nwukwu*, *nwukwu* is still a live Q_WH root, whereas *amwu* has already moved far along the polarity cline into the NPI/FC corner (as shown in Chapter 4, especially §4.3.3). There is no productive *amwu-i-nka?* template left to feed an *amwu-inka* DP, and the functional space around SU/IR is already being filled by wh+-(*i*)*nka*, while *amwu-to* and *amwu-na* occupy the AA/DN and FC/IND sectors. Crucially, overlap alone is not the reason. Overlap is common

in transitional systems. What is missing here is a combination of viable source and viable niche.

Overall, this chapter adds a morphological layer to the root-level story of Chapter 4. At the level of roots, *nwukwu-* and *amwu-* already occupy different parts of Haspelmath's map and of Jäger's polarity scale: *nwukwu-* supports interrogative and epistemic-specific uses, while *amwu-* is specialized for NPI and free-choice uses. The wh+-(*i*)*nka* development reinforces this split. Analytic *nwukwu-i-nka* and contracted *nwukwunka* form a wh-based series of ignorance indefinites that stay in the interrogative/epistemic corner, while *amwu*-based combinations with *-to*, *-na*, and -(*i*)*lato* continue to realize negative and FREE-CHOICE functions. The broader lesson is therefore not merely that one form is missing, but that source constructions and paradigm structure jointly determine which morphological possibilities actually stabilize.

This conclusion also suggests a natural direction for future work. If the Korean Q→DP pathway is genuinely tied to wh+copula+-(*i*)*nka* structures, then other wh-roots are the obvious place to look for parallel histories. The present chapter has established that *amwu* is not such a case. In the next chapter we return to these series in their fully developed modern form and ask how *nwukwu-* and *amwu*-based particle pairs—notably *amwu-to*/*nwukwu-to* and *amwu-na*/*nwukwu-na*—now compete and specialize within the shared semantic space that the preceding chapters have mapped out.

Chapter 6

Competition in shared particle series on Haspelmath's map

This chapter develops the competition part of the dissertation. Chapters 4 and 5 showed that the two roots *nwukwu-* and *amwu-* no longer behave alike: *nwukwu-* covers interrogative and non-negative indefinite uses, while *amwu-* has narrowed into strongly negative and free-choice uses. Chapter 3 formulated this as H3: when both roots share a particle, *amwu*-series should anchor the most negative and indiscriminative corners of the map, and *nwukwu*-series should stay closer to neutral, extensional, or epistemic uses.

The present chapter tests that expectation for the two particle series that are both well represented in the corpus and central in the synchronic literature:

- the NPI series with *-to*, where *amwu-to* has been analysed as a very strong NPI ('not even anyone') and *nwukwu-to* as a polarity-sensitive form that still allows somewhat wider use; and
- the free-choice series with *-na*, where *nwukwu-na* is described as a domain-based free-choice item ('whoever / anyone in this set'), and *amwu-na* as an indiscriminative free-choice item ('just any person', often with evaluative flavour).

What has been missing so far is a semantic-map based corpus view of how these pairs divide up the negative and free-choice zones across the 20th century. Using the annotation and modelling framework from Chapter 3, I ask:

- how *amwu-to* and *nwukwu-to* distribute across DN, AA, and nearby nodes on the extended semantic map, and whether the data support the idea that *amwu-to* is concentrated in the strongest negative environments while *nwukwu-to* extends into neighbouring, less strongly negative contexts; and
- how *amwu-na* and *nwukwu-na* distribute across FC, GEN, IND, and related nodes, and whether the data support the idea that *nwukwu-na* anchors the extensional FC/GEN/UFC corridor while *amwu-na* specializes in indiscriminative 'just any X' uses near IND.

Throughout, I focus on the 20th-century part of the Yonsei corpus, where all four forms are frequent. For each series I give period-based distributions, assign semantic-map labels (DN, AA, FC, GEN, IND, etc. as in §2.1.2), and visualise the results on dot-density semantic maps. For *-to*, the maps are centred on the NPI cluster (DN/AA), and for *-na* they are centred on the FC cluster (FC/GEN). In addition, I use the entry-time diagnostics from Chapter 3 to ask when new functions become frequent enough to count as part of a series.

The broader claim is that shared-particle competition does not produce two interchangeable 'anyone' systems. Instead, it produces *coexistence with specialization*. This chapter therefore also clarifies what predictive value the semantic map has in a competition setting. The map predicts where competition is likely to be resolved —within contiguous neighbourhoods—but it does not by itself predict which form will dominate which node. That outcome depends on the root histories established in the earlier chapters.

To turn these questions into something measurable, I restate H3 from Chapter 3 in a form tailored to *-to* and *-na*, using the extended semantic map from §2.1.2 as the underlying space.

Following Haspelmath's continuity idea, I assume that each morphological series occupies a contiguous region on the map and that changes progress via neighbouring nodes rather than by skipping functions. For the shared particle series, this gives three working hypotheses. Importantly, the hypotheses are about *local specialization*, not total blocking: the two roots may overlap transitionally or at the edges, but they are expected to develop different centres of gravity.

(**H-COMP to**) **Specialization in the negative sector** (***-to***).

In the *-to* series, *amwu-to* is expected to concentrate in the strongest negative environments. On the extended map this means that almost all *amwu-to* tokens should fall under DN and AA (direct clausal negation and anti-additive contexts), with very few tokens in questions, conditionals, or other weaker environments. *Nwukwu-to* is expected to share the DN/AA node but to appear more often than *amwu-to* in neighbouring contexts such as comparatives, weak free-choice, and related edge environments. In short, *amwu-to* should anchor the most negative core of the map, while *nwukwu-to* should occupy a wider strip around it.

(**H-COMP na**) **Specialization in the free-choice sector** (***-na***).

In the *-na* series, *nwukwu-na* is expected to align with the FC/UFC/GEN part of the map, where free-choice uses approach extensional, universal, and generic readings. *Amwu-na*, by contrast, is expected to align more strongly with FC/IND, where the relevant meaning is closer to 'just any X' or evaluative indiscriminacy. This reflects the synchronic descriptions in

§2.8.2: *nwukwu-na* as extensional free choice, *amwu-na* as indiscriminative free choice.

(**H-COMP global**) **Complementarity across shared particles.**

Across *-to* and *-na*, the two roots are expected to end up in different parts of the map. *Amwu*-based forms should dominate the strong negative and strong indiscriminative corners (DN/AA and FC/IND), while wh-based forms should dominate the extensional FC/GEN/UFC corridor and the weaker or more neutral edge of negative space. When a particle is available for both roots, the two series should therefore split up the shared space rather than duplicate each other, in a way that respects contiguity on the map and matches the root-level tendencies established in Chapter 4.

The following sections test these hypotheses on the labelled corpus data. Section 6.2 turns first to the negative domain and examines how *amwu-to* and *nwukwu-to* are distributed over DN, AA, and nearby nodes. Section 6.3 then turns to the free-choice domain and compares *amwu-na* and *nwukwu-na* across FC, GEN, IND, and related labels. Section 6.4 draws the two case studies together and asks what kind of competition the Korean data support: replacement, coexistence with specialization, or something in between. Section 6.5 closes the chapter by reassessing H-COMP to, H-COMP na, and H-COMP global.

6.2 Negative competition: *amwu-to* and *nwukwu-to*

This section focuses on the best-established NPI pair in Korean, *amwu-to* and *nwukwu-to*. The goal is to see, in semantic-map terms, how far they share the same negative territory and how far they specialize. Throughout I ignore tokens labelled no_matter, which usually involve additional concessive morphology and are treated elsewhere.

6.2.1 Overall frequencies and map locations

To sum up, the 20th-century distributions show that *amwu-to* is overwhelmingly concentrated in the NPI core, whereas *nwukwu-to* occupies a broader band that includes both the NPI sector and its comparative or free-choice edge. To make this visible, I divide the data into three broad thirty-year slices (1900–1930, 1930–1960, 1960–1990) and plot token counts at the corresponding map nodes.

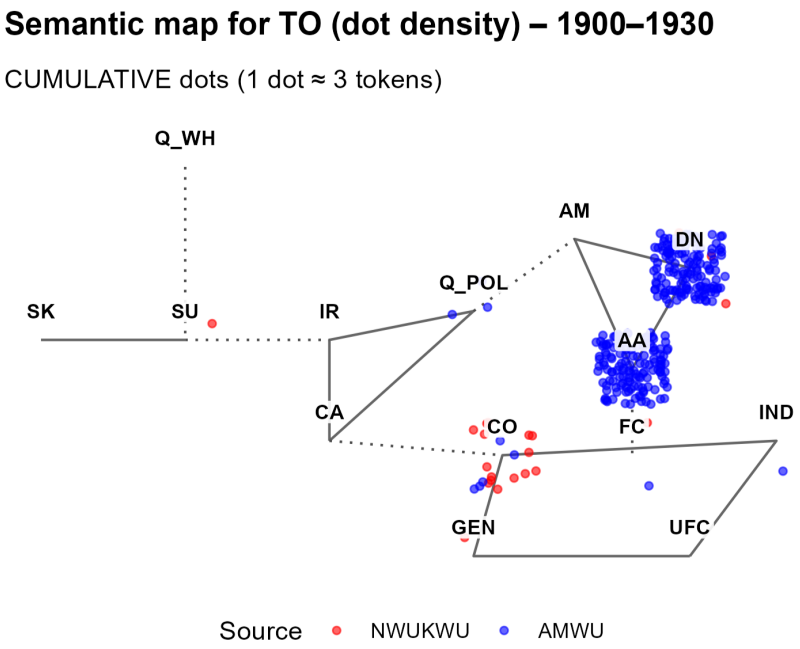


Figure 6.1 Semantic-map locations of *amwu-to* and *nwukwu-to*, 1900–1930 (ignoring NO_MATTER).

In the earliest slice (Figure 6.1), *amwu-to* already anchors the NPI core. It has 403 AA tokens and 414 DN tokens, plus only a few side uses (13 CO, 3 FC, 2 IND, 9 Q_POL). In contrast, *nwukwu-to* is concentrated almost entirely in the CO and GEN nodes (44 CO, 3 GEN, 1 SU, and only 9 DN tokens). Thus, in this earliest period, *amwu-to* is firmly in the NPI node (AA/DN), while *nwukwu-to* enters from the comparative or free-choice side, mostly as a CO-type item with only a small negative strand.

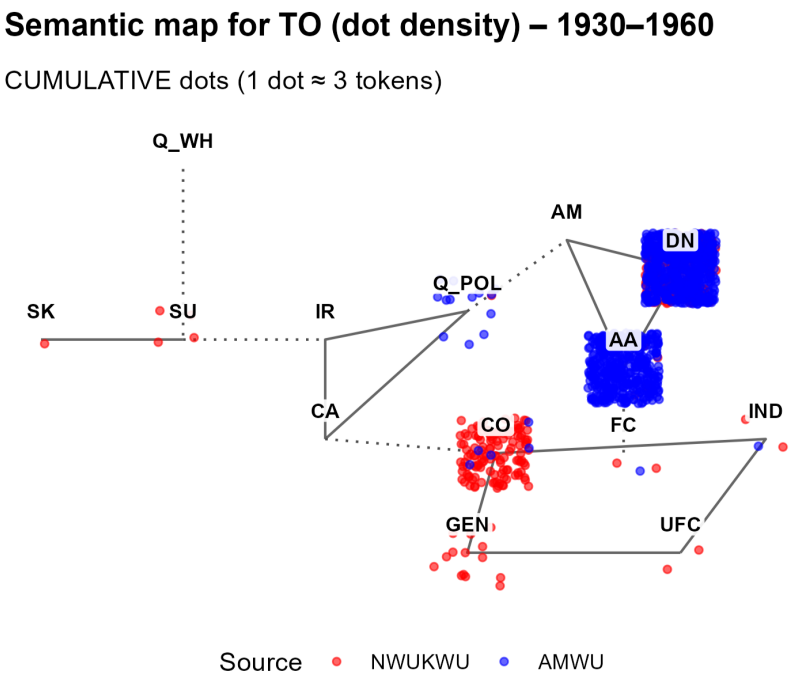


Figure 6.2 Semantic-map locations of *amwu-to* and *nwukwu-to*, 1930–1960 (ignoring NO_MATTER).

In the middle slice (Figure 6.2), both series increase in number, and the overlap in the NPI sector becomes clearer. *Amwu-to* has 1,208 AA and 1,615 DN tokens, with only a handful of other labels (13 CO, 3 FC, 3 IND, 37 Q_POL). *Nwukwu-to* now has 12 AA and 422 DN tokens, but also 375 CO tokens and 41 GEN tokens, plus a few IND, Q_POL, SK, SU, and UFC tokens. So by mid-century, both roots are firmly present in the NPI sector, but the division of labour remains: the densest AA/DN dots belong to *amwu-to*, while *nwukwu-to* has many more dots in CO and GEN.

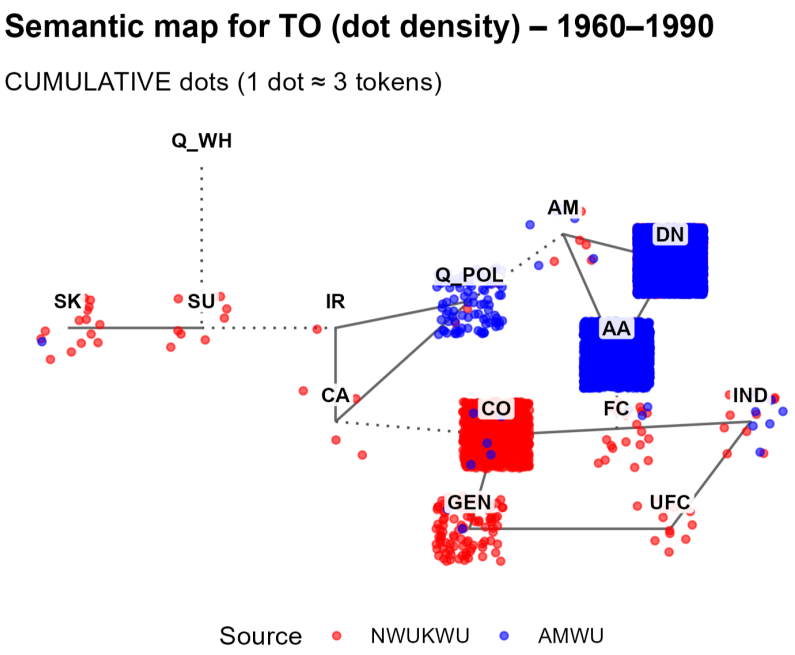


Figure 6.3 Semantic-map locations of *amwu-to* and *nwukwu-to*, 1960–1990 (ignoring NO_MATTER).

In the late slice (Figure 6.3), the contrast is sharpest. *Amwu-to* has 8,046 AA and 13,481 DN tokens and only a small number of uses elsewhere (14 CO, 9 FC, 4 GEN, 20 IND, 20 AM). *Nwukwu-to* has 6,519 DN and 32 AA tokens, but also 2,056 CO tokens, 60 FC tokens, 280 GEN tokens, 35 UFC tokens, and a scattering of CA, SU, SK, IR, and Q_POL tokens. In the late 20th century, then, *amwu-to* is almost entirely confined to the NPI nodes (AA/DN/AM), while *nwukwu-to* spans both the NPI cluster and the nearby free-choice and neutral nodes (CO, FC, GEN, UFC, CA).

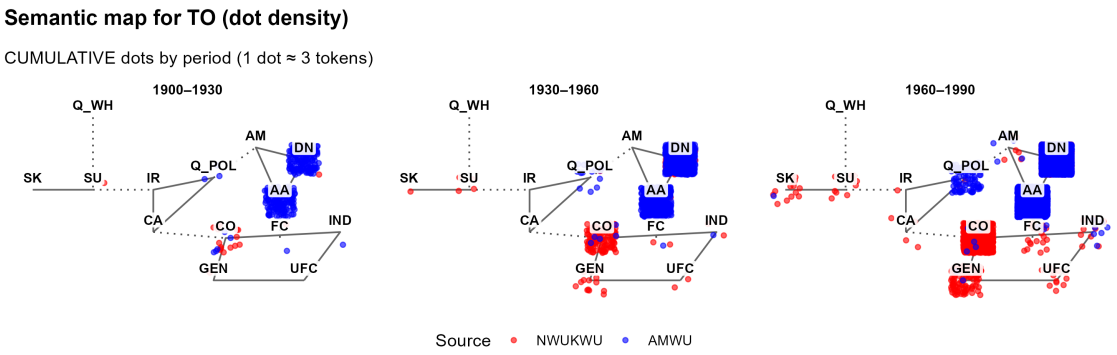


Figure 6.4 Semantic-map locations of *amwu-to* and *nwukwu-to* across all three periods (1900–1990, ignoring NO_MATTER).

Figure 6.4 overlays the three slices. *Amwu-to* stays at the NPI core throughout: its main labels are AA and DN, with only small satellites in CO, FC, and GEN. *Nwukwu-to* starts from the CO side, grows into DN/AA, and ends up occupying a band that straddles the boundary between the NPI sector and the comparative/free-choice sector. This is already enough to support the main descriptive claim of H-COMP$_{to}$: the strongest negative corner is anchored by *amwu-to*, while *nwukwu-to* also covers a wider comparative and weakly negative band around it.

6.2.2 Entry times relative to the NPI core

The period maps show that *amwu-to* and *nwukwu-to* enter the map from different directions: *amwu-to* from the NPI core, *nwukwu-to* from the CO node. To connect this more directly to the structure of the map, I now look at when each semantic function becomes established for each series and how this relates to distance from the NPI core.

For this, I take the NPI sector (DN/AA/AM) as the reference point and assign each label a graph distance from that sector (`dist_from_npi`). DN, AA, and AM

have distance 0; FC is at distance 1; CO at distance 2; CA, GEN, and UFC at distance 3; Q_POL at distance 4; IR at distance 5; and SU at distance 6. IND and SK are treated as off-map for this calculation. For each label ℓ and each root, I then identify the first decade in which ℓ meets the entry threshold (at least 5 tokens and at least 5% of all tokens of that series in that decade), as in §3.3.5. The resulting entry times are plotted against distance in Figure 6.5.

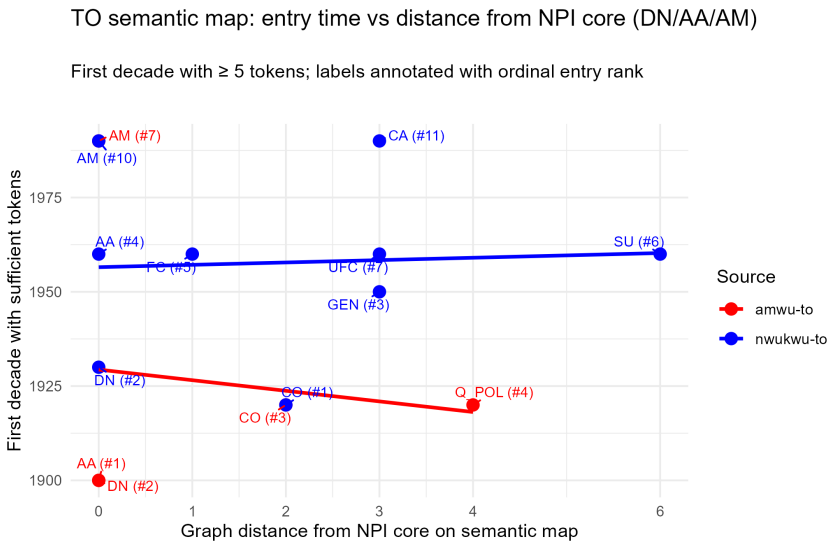


Figure 6.5 Entry decade vs. distance from NPI core for *-to* labels (both roots).

For *amwu-to*, AA and DN enter immediately at distance 0 (1900s), CO enters in the 1920s at distance 2, Q_POL enters in the 1920s at distance 4, and AM enters only in the 1990s, again at distance 0. IND and NO_MATTER never cross the threshold and are ignored here. So the first firmly established functions of *amwu-to* lie squarely in the NPI core; only later is there a small extension into CO and Q_POL, and even those labels remain numerically minor.

For *nwukwu-to*, the first established label is CO in the 1920s at distance 2, followed by DN in the 1930s at distance 0. GEN enters in the 1950s at distance 3, AA

in the 1960s at distance 0, and FC, SU, and UFC all in the 1960s at distances 1, 6, and 3 respectively. IND, SK, CA, AM, and US enter only in the 1970s–1990s or never reach the threshold. So *nwukwu-to* first becomes entrenched in a comparative/free-choice node (CO), then acquires a strong NPI use (DN), and only later adds nearby FC, GEN, and UFC functions.

A simple linear model schematizing this relation is given in (1):

(1)

$$\text{first\_decade}_\ell \sim \text{dist\_from\_npi}_\ell$$

Fitted to the pooled labels for both roots, this model does not show a strong linear trend. The estimated slope is small and not significant.[1] This is expected, since the two roots start from different parts of the map: *amwu-to* enters at distance 0, *nwukwu-to* at distance 2. What matters for the present chapter is the direction of movement for each root. *Amwu-to* has its first and strongest foothold in the NPI core and remains overwhelmingly concentrated there. *Nwukwu-to* begins at CO and then moves inward to DN/AA while also spreading into nearby FC/GEN/UFC and a few neutral nodes.

In short, the entry-time analysis matches the period maps. For *amwu-to*, the development is 'NPI first and NPI throughout', with only small later extensions into neighbouring functions. For *nwukwu-to*, the earliest solid use is comparative/free-choice (CO), and later uses add DN/AA and other nearby nodes. This is exactly the pattern anticipated by H-COMP to: *amwu-to* occupies the strongest negative corner, while *nwukwu-to* also covers a band of comparative and weakly negative functions surrounding it.

[1]For the pooled *-to* data, the fitted model is first_decade $= 1944.0 + 1.40 \cdot$ dist_from_npi, with $R^2 \approx 0.01$. This does not support a simple claim that labels further from the NPI core always enter later.

6.2.3 Interim summary: competition in the *-to* series

The *-to* data confirm and refine the synchronic picture in §2.8.1. Both *amwu-to* and *nwukwu-to* are clearly polarity-sensitive and strongly drawn to the negative corner of the semantic map, but they do not fill that corner in the same way.

On the *amwu* side, the corpus shows exactly what the descriptive literature suggests. From the earliest decades, *amwu-to* enters the map at the NPI core (AA and DN) and stays there. Across all three period slices, its overwhelmingly dominant labels are AA and DN, and later additions such as CO and Q_POL never come close in frequency. This matches the 'maximally strong NPI' view in §50 and 54: *amwu-to* is tied to direct negation and strong negative predicates, and other uses are marginal.

On the wh side, the map and the timing look quite different. *Nwukwu-to* first becomes established in CO and only then develops a solid DN use, with AA following later. By the late 20th century it is clearly a negative item, but it also keeps a substantial comparative/free-choice neighbourhood that *amwu-to* lacks. This fits the intuitions in §50: *nwukwu-to* is a strong NPI under negation, but it also keeps comparative and weak free-choice readings that *amwu-to* does not.

Seen against the map-based hypotheses in §6.1, the *-to* series behaves exactly as H-COMP to predicts. *Amwu-to* is the fixed point at the negative centre: it enters as AA/DN, stays overwhelmingly in AA/DN, and contributes almost nothing to the surrounding free-choice or neutral nodes. *Nwukwu-to* enters from the CO side, then moves into DN/AA, and ends up covering both the NPI nodes and a belt of comparative and weakly free-choice uses around them. The shared particle *-to* therefore does not produce two identical 'anyone' series. Instead, it produces an asymmetrical overlap in which *amwu-to* anchors the core and *nwukwu-to* bridges that core to a neighbouring band of functions.

6.3 Free-choice competition: *amwu-na* and *nwukwu-na*

I now turn to the free-choice particle *-na*, and ask how *amwu-na* and *nwukwu-na* share the free-choice area of the semantic map. As in the *-to* section, I first describe three broad period slices (1900–1930, 1930–1960, 1960–1990), and then use an entry-time analysis to relate the decade in which a function becomes frequent to its distance from the FC node. The focus is on FC-related labels (FC, GEN, UFC, IND, DN, Q_POL, Q_WH); no_matter is rare for *-na* and will not be central here.

6.3.1 Overall frequencies and map locations

Across the 20th century, *nwukwu-na* is much more frequent than *amwu-na*, and the two forms settle in slightly different parts of the free-choice space.

In the earliest slice, 1900–1930, *amwu-na* has 2 FC tokens, 14 GEN tokens, and 15 IND tokens, with no DN tokens. *Nwukwu-na* has 15 FC tokens, 265 GEN tokens, 2 IND tokens, and 3 DN tokens, plus a single Q_POL token. So even in the earliest period, wh+*-na* supplies most FC and GEN uses, while *amwu-na* already contributes many more IND tokens (‘just any X’).

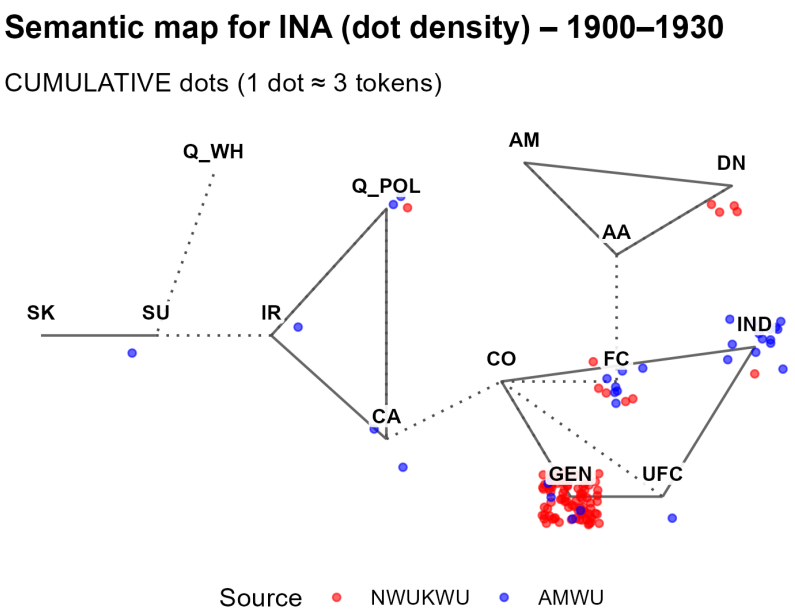


Figure 6.6 Semantic-map locations of *amwu-na* and *nwukwu-na*, 1900–1930.

On the map in Figure 6.6, *nwukwu-na* appears at FC and GEN (with a few DN dots), and almost not at IND. *Amwu-na* places most of its dots at IND, with only a thin trace at FC and GEN. That is, wh+-*na* enters through the central free-choice nodes, while *amwu-na* enters slightly off-centre in more indiscriminative contexts.

In the middle slice, 1930–1960, both forms become more frequent. *Amwu-na* has 24 FC tokens, 35 GEN tokens, 51 IND tokens, and 1 UFC token. *Nwukwu-na* has 481 FC tokens, 1,761 GEN tokens, 7 IND tokens, 9 DN tokens, 5 Q_POL tokens, 10 Q_WH tokens, 1 SK token, 1 SU token, and 2 UFC tokens. So by mid-century *nwukwu-na* contributes around twenty times as many FC tokens and fifty times as many GEN tokens as *amwu-na*, while *amwu-na* still has many more IND tokens.

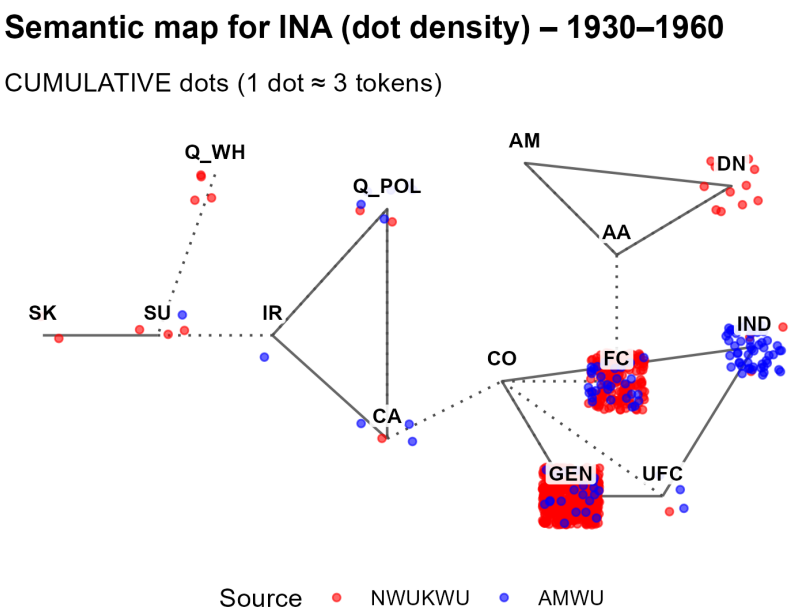


Figure 6.7 Semantic-map locations of *amwu-na* and *nwukwu-na*, 1930–1960.

Figure 6.7 shows the mid-century map. The FC and GEN nodes are now filled mainly by wh-based dots; *amwu-na* dots are tightly clustered at IND, with a smaller presence at FC and GEN. DN remains a fringe node in this period and is overwhelmingly associated with the wh-series, not with *amwu-na*.

In the late slice, 1960–1990, both series are very frequent. *Amwu-na* has 244 FC tokens, 155 GEN tokens, and 827 IND tokens, plus 1 SK token and 2 UFC tokens. *Nwukwu-na* has 4,630 FC tokens, 9,860 GEN tokens, 50 IND tokens, 35 DN tokens, 31 Q_POL tokens, 51 Q_WH tokens, 1 SK token, 2 SU tokens, and 11 UFC tokens. Thus in the late 20th century *nwukwu-na* supplies nearly fifty times as many GEN tokens and almost twenty times as many FC tokens as *amwu-na*, while *amwu-na* supplies more than sixteen times as many IND tokens (827 vs. 50).

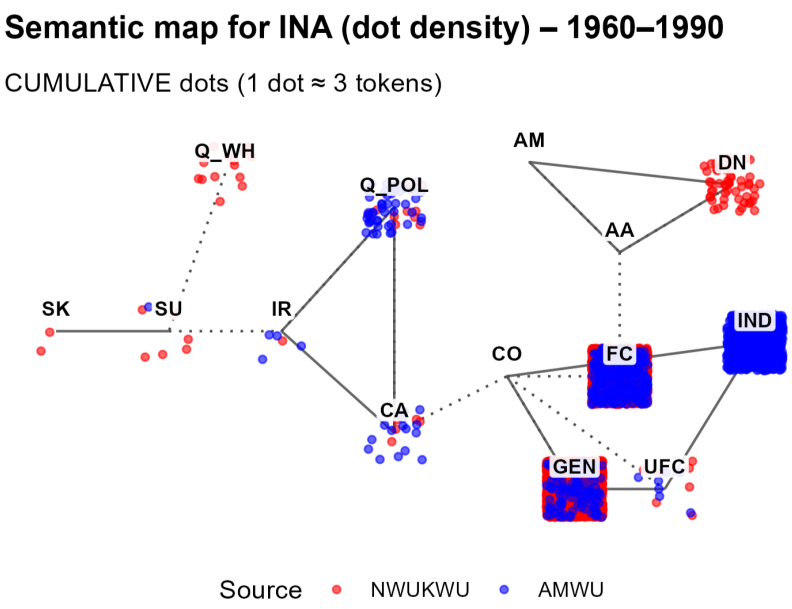


Figure 6.8 Semantic-map locations of *amwu-na* and *nwukwu-na*, 1960–1990.

Figure 6.8 makes the split very clear. *Nwukwu-na* now covers the FC and GEN nodes almost entirely, with extra dots at UFC and a small band at DN and Q_POL/Q_WH. *Amwu-na* remains closest to IND, with a smaller but visible trail at FC and GEN. This matches the contrast reviewed earlier in Chapter 2: *nwukwu-na* behaves like an extensional universal in the FC/GEN corridor ('whoever / anyone in this set'), while *amwu-na* tends toward 'just any X' readings anchored in IND.

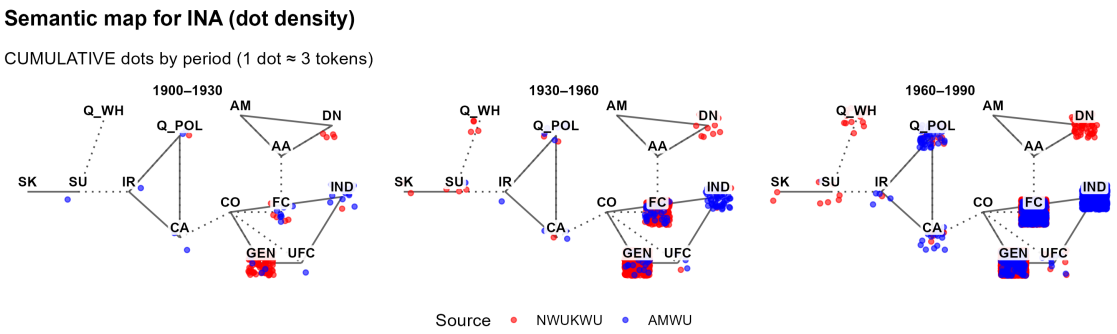


Figure 6.9 Semantic-map locations of *amwu-na* and *nwukwu-na* across all three periods (1900–1990).

The overlay in Figure 6.9 summarizes the full century. *Nwukwu-na* fills the FC and GEN nodes and sends a thin band of dots into DN, UFC, Q_POL, and Q_WH. *Amwu-na* is concentrated around IND, with a modest FC/GEN trail. On the extended free-choice map this can be stated as follows: *nwukwu-na* dominates the FC/UFC/GEN corridor, and *amwu-na* keeps its centre of gravity near FC/IND.

6.3.2 Entry times relative to the FC core

To relate these counts more directly to the structure of the map, I now ask when each function becomes established for each root and how this correlates with distance from the FC node.

I treat the FC node as centre and assign each label a graph distance from FC (`dist_from_fc`), with FC at distance 0, AA and CO at distance 1, AM/CA/DN/GEN/UFC at distance 2, Q_POL at distance 3, IR at distance 4, SU at distance 5, and Q_WH at distance 6. IND and SK are not given a numeric distance here because they sit at the edge of the map in the present representation. For each label ℓ and each root, I then identify the first decade in which ℓ meets the entry threshold (at least 5 *-na* tokens and at least 5% of that root's *-na* tokens in that decade). This yields the entry pattern summarized in Figure 6.10.

For *amwu-na*, the earliest entrenched function is GEN in the 1920s (distance 2), followed by IND in the 1930s (distance not used in the regression) and FC only in the 1960s (distance 0). So *amwu-na* first settles as a GEN/IND-type indefinite in free-choice-like contexts and only later develops a stable FC use in the narrow sense.

For *nwukwu-na*, FC reaches the threshold already in the 1910s (distance 0), GEN in the 1920s (distance 2), and then, much later, Q_WH in the 1960s (distance 6), DN and IND in the 1970s (distance 2 and off-map), and Q_POL and UFC in the 1990s (distances 3 and 2). So *nwukwu-na* enters at the FC node itself, quickly adds GEN, and only later picks up small bands of DN, Q_POL, and UFC.

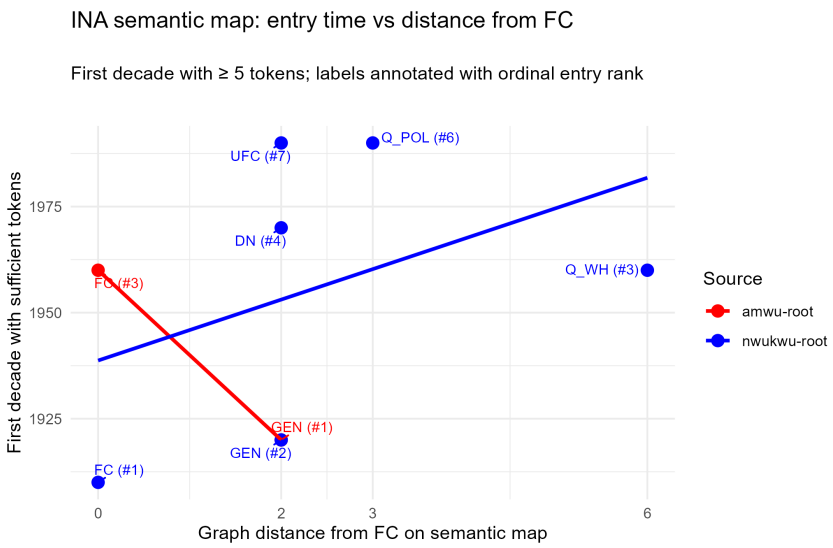


Figure 6.10 Entry decade vs. distance from FC for *-na* labels.

A linear model schematizing this relation is given in (2):

(2)

$$\text{first\_decade}_{\ell} \sim \text{dist\_from\_fc}_{\ell}$$

The estimates do not show a strong linear trend. The fitted slope is small and not significant.[2] This is in line with the fact that *nwukwu-na* enters directly at FC (distance 0), whereas *amwu-na* does not have FC as an early centre. What matters here is the direction of movement for each root.

For *nwukwu-na*, the development runs 'from the centre outwards': FC is first, GEN follows, and later functions such as UFC, DN, and Q_POL fill in nearby nodes. For *amwu-na*, the development runs 'from the side inwards': GEN and IND are first, and FC only becomes firmly established decades later. From a competition point of view, this matches the descriptive asymmetry from Chapter 2: *nwukwu-na* is the main extensional FCI, while *amwu-na* is the main source of indiscriminative readings around IND, with a smaller FC strand that emerges later.

6.3.3 Interim summary: competition in the *-na* series

The *-na* results confirm and refine the synchronic picture sketched in §2.8.2 and match H-COMP na.

First, the corpus supports the idea that *nwukwu-na* is the central free-choice marker. In every period, wh+*-na* supplies the vast majority of FC and GEN uses. Already in 1900–1930, *nwukwu-na* has many more FC and GEN tokens than *amwu-na*; by 1960–1990 it has 4,630 FC and 9,860 GEN tokens, whereas *amwu-na* has 244 FC and 155 GEN tokens but 827 IND tokens. On the map, *nwukwu-na* is firmly centred on the FC and GEN nodes, with some spread into UFC and a thin band into DN and Q_POL/Q_WH.

Second, the corpus confirms that *amwu-na* is heavily drawn to indiscriminative uses around IND. In every period *amwu-na* has many more IND tokens than *nwukwu-na*: 15 vs. 2 in 1900–1930, 51 vs. 7 in 1930–1960, and 827 vs. 50 in 1960–1990.

[2] For the pooled *-na* data, the fitted model is first_decade $= 1940.75 + 5.53 \cdot$ dist_from_fc, with $R^2 \approx 0.11$. This is expected, since the two roots start from different map locations.

Its FC and GEN counts do grow over time, but always remain small compared to the wh-series. This fits the evaluative 'just any person' flavour noted in the literature: the core habitat of *amwu-na* on the map is the FC/IND corner, with GEN and FC trailing behind rather than leading.

Third, the entry-time analysis shows that the two roots enter the *-na* space at different points. *Nwukwu-na* is entrenched at FC already in the 1910s (`dist_from_fc` = 0) and at GEN in the 1920s (distance 2), and only later picks up small DN, Q_POL, and UFC uses. *Amwu-na*, by contrast, reaches the threshold first at GEN in the 1920s (distance 2), then at IND in the 1930s, and only reaches FC in the 1960s (distance 0). So the wh-based series grows 'from the centre outwards' (FC $\rightarrow$ GEN $\rightarrow$ nearby nodes), while the *amwu*-series grows 'from the side inwards' (GEN/IND $\rightarrow$ FC). This pattern supports H-COMP na: the main extensional FC uses are carried by *nwukwu-na*, while *amwu-na* specializes in indiscriminative readings and adds a narrower FC strand later.

A small but important point is that DN uses of *nwukwu-na* are attested, whereas DN uses of *amwu-na* almost never meet the entry threshold. This is in line with the observation that *nwukwu-na* can take wide scope over negation and yield 'nobody' readings, while *amwu-na* tends to stay under negation and favour 'not everyone' readings. On the map, DN remains a fringe node for *-na*, but the fact that it is almost entirely wh-based rather than *amwu*-based fits the broader division of labour: strong free-choice under negation is handled by wh-based forms, whereas *amwu*-based forms reserve their strongest negative work for *-to* and related negative particles.

Overall, then, the *-na* competition results align well with the synchronic descriptions and with H-COMP na. The shared particle *-na* does not yield two interchangeable 'anyone' forms. Instead, *nwukwu-na* occupies the FC/UFC/GEN corridor on

the extended map, and *amwu-na* settles near FC/IND with a strong indiscriminative pull. This is exactly the kind of split we expect if the modern *-na* series sits on top of the root-level specialization established in Chapter 4.

6.4 Discussion: coexistence, specialization, and predictive value

In sum, the *-to* and *-na* case studies clarify what kind of competition the Korean paradigm actually exhibits. The pattern is not simple replacement. Nor is it free alternation. Instead, the best description is *coexistence with specialization*: both roots remain available with the same particle, but they settle into different subregions of the same semantic neighbourhood.

This point is clearest in the *-to* series. Here the competition is asymmetrical overlap around the NPI core. *Amwu-to* is the fixed centre: its distribution is overwhelmingly AA/DN from the beginning, and that centre of gravity remains stable throughout the century. *Nwukwu-to* overlaps with that core, especially in DN, but it also retains a substantial comparative and weak free-choice fringe in CO, FC, GEN, and UFC. The result is not that one series eliminates the other; rather, one series anchors the strongest negative space and the other connects that space to neighbouring functions.

The *-na* series shows a different but equally structured kind of competition. Here the issue is not 'negative core vs. wider fringe', but an internal partitioning of the free-choice sector itself. *Nwukwu-na* anchors the extensional FC/UFC/GEN corridor: this is the main series for 'whoever' and generic 'anyone' readings. *Amwu-na*, by contrast, is pulled toward IND and related indiscriminative uses: this is the main series for 'just any X' readings with a more evaluative flavour. Again, the two series coexist, but they do not duplicate each other.

This chapter also helps answer a broader committee-style question: are there any surprises or irregularities, or does the system simply conform perfectly to the

semantic map? The answer is that the dominant changes conform strongly, but the details are informative. One small surprise is that *nwukwu-to* becomes firm first in CO rather than in DN/AA, even though it later acquires a strong DN profile. Another is that *amwu-na* reaches stable GEN/IND use before it reaches FC in the narrow sense. These are not violations of the map. They are better understood as reminders that the map predicts neighbourhoods and local routes, not a single fixed sequence that all forms must follow.

This leads to the question of predictive value. The Korean competition data suggest that the semantic map has predictive force at the level of *where* competition is likely to be resolved. The map correctly predicts that the shared particle series will differentiate themselves within contiguous regions rather than by jumping to unrelated functions. It does not, by itself, predict which root will dominate which node. That outcome depends on earlier root-level histories and on the functional associations already established elsewhere in the paradigm.

The chapter therefore strengthens one of the dissertation's broader claims. Particle meaning alone does not determine the final organization of the system. The same particle yields different functional futures depending on which root it combines with and which part of the map that root already occupies. In this sense, the Korean evidence supports a paradigm-level view of morphosyntax: competition operates not only over forms, but over structured semantic space.

6.5 Conclusion

This chapter has examined how *amwu-* and wh-based indefinites behave when they share the same particle, using *-to* and *-na* as test cases. The main points can be organized as follows:

1. The *-to* series confirms a split in the negative sector: *amwu-to* is concentrated in the strongest NPI environments, while *nwukwu-to* shares those environments but also spreads into nearby comparative and weakly negative uses.

2. The *-na* series confirms a split in the free-choice sector: *nwukwu-na* carries most of the extensional FC/GEN load, while *amwu-na* is centred around indiscriminative IND uses with a smaller FC component that emerges later.

3. Across both particles, the shared space is not duplicated: each root–particle combination settles in a different neighbourhood on the semantic map, in line with the broader competition hypothesis H3 from Chapter 3.

The first point concerns the negative series with *-to*. Earlier work summarized in §2.8.1 had already argued, on synchronic grounds, that *amwu-to* is a maximally strong NPI ('not even anyone'), while *nwukwu-to* is polarity-sensitive but more flexible. The map-based corpus results sharpen this picture. In the earliest period (1900–1930), *amwu-to* already sits at the NPI core, with several hundred AA and DN tokens and only a few side uses; *nwukwu-to* appears mainly in CO and GEN, with very few DN tokens. In the mid-century period (1930–1960) both series enter DN/AA, but the counts show that *amwu-to* remains dominant in those nodes, while *nwukwu-to* has many more CO and GEN tokens. In the late period (1960–1990), *amwu-to* has thousands of AA and DN tokens and very little outside the NPI nodes, whereas *nwukwu-to* has a comparable number of DN tokens but also thousands of CO tokens and a visible cluster in FC/GEN/UFC. The entry-time analysis confirms that *amwu-to* is entrenched first and foremost at distance 0 from the NPI core (AA/DN), while *nwukwu-to* first reaches the CO node and then moves inward. This is exactly what H-COMP to predicts: *amwu-to* anchors the strongest AA/DN corner, *nwukwu-to* links that corner to neighbouring comparative and weakly negative uses.

The second point concerns the free-choice series with *-na*. The synchronic literature reviewed in §2.8.2 distinguished an extensional, 'whoever/anyone-in-this-set' *nwukwu-na* from a more indiscriminative *amwu-na*. The semantic-map counts in §6.3.1 support and refine this. In every period, wh+*-na* supplies the bulk of FC and GEN tokens. Already in 1900–1930, *nwukwu-na* has many more FC and GEN tokens than *amwu-na*; by 1960–1990 it has 4,630 FC and 9,860 GEN tokens, whereas *amwu-na* has 244 FC and 155 GEN tokens but 827 IND tokens. On the map, *nwukwu-na* fills the FC and GEN nodes and extends slightly into UFC and DN/Q_POL/Q_WH, while *amwu-na* clusters closer to IND with a smaller FC/GEN band. The entry-time analysis shows that *nwukwu-na* stabilizes first at FC and GEN, and only later touches DN, Q_POL, and UFC, whereas *amwu-na* stabilizes first at GEN and IND and only acquires a stable FC strand later. This matches H-COMP na: *nwukwu-na* is the main extensional FCI, *amwu-na* is the main source of 'just any X' readings.

The third point ties these findings back to the global competition hypothesis H3 and to the root-level story in Chapter 4. At the bare level, *nwukwu-* had expanded from Q_WH into non-negative indefinites, while *amwu-* had narrowed into the NPI/FC corner. The present chapter shows that, once particles are added, this split is preserved rather than erased. In the *-to* series, *amwu-* keeps the strongest negative corner, and *nwukwu-* takes over the CO/FC band adjacent to it. In the *-na* series, *nwukwu-* keeps the central FC/GEN corridor, and *amwu-* settles at the IND edge with a later FC strand. When a particle is available for both roots, the series do not collapse into two interchangeable 'anyone' forms; instead, they map onto different regions of the same semantic space.

The broader implication is that the Korean system is not best understood as a set of isolated lexical facts. It is a paradigm whose members compete locally over a structured semantic space. The map is predictive in that it constrains where that competition can be resolved, but the winner in each niche is shaped by root history

and prior specialization. Accordingly, Chapter 6 completes the competition piece: root asymmetries, grammaticalization paths, and particle choices conspire to yield a fine-grained, contiguous division of labour on Haspelmath's map.

Chapter 7

Conclusion

This dissertation has argued that Korean indefinites are best understood as a paradigm assembled through competition over a structured semantic space, rather than as a set of forms that change one by one in isolation. The central tools have been Haspelmath' s (1997) semantic map for indefinites and polarity-oriented approaches to diachrony in the spirit of Jäger (2010). Within this framework, I have addressed three linked puzzles, each corresponding to one of the three empirical chapters.

More broadly, the dissertation has argued that Korean is especially revealing because expressions comparable to English *someone* and *anyone* are often assembled transparently from wh-roots and particles whose contributions remain semantically visible. This makes Korean a useful testing ground for a more general question: how indefinite paradigms are built, how morphosyntax is organized over semantic space, and how clause-level morphology can be recruited into the nominal domain without erasing its earlier function. A general lesson of the study is therefore that forms comparable to 'someone' need not reduce to existential quantification alone: they may also encode epistemic, modal, discourse, concessive, or alternative-based structure, and these components help determine both their diachronic pathways and their synchronic distribution.

Empirically, the answers to these questions rest on a diachronic corpus study covering roughly three centuries of written Korean. The History corpus provides material from 1700 to 1890, and the Yonsei corpus extends the record from 1900 to

1990. From these corpora, I extracted instances of *nwukwu-* and *amo-/amwu-* (bare and with particles such as *-to*, *-na*, *-(i)nka*, *-tunci*, and *-(i)lato*) and assigned each token a functional label on an extended Haspelmath-style map (SK, SU, IR, Q, CA, CO, DN, AA, AM, FC, GEN, UFC, IND). For some analyses, these fine-grained labels were grouped into five macro-environments (Q_WH, PPI, NEUTRAL, NPI, FREE-CHOICE) to track broad shifts in interrogative, existential, negative, and free-choice uses over time. Logistic and segmented regressions, together with simple entry-threshold criteria, were used to relate these labels to diachronic trends.

The remainder of the chapter proceeds as follows. Section 7.1 returns to the three puzzles and states the answers directly, drawing together the results of Chapters 4 through 6. Section 7.2 draws out the broader theoretical implications for paradigms, semantic maps, polarity, grammaticalization, and language change. Section 7.3 highlights the methodological contribution of the corpus-and-map approach adopted here. Section 7.4 discusses limitations and open issues. Section 7.5 sketches directions for future research, and Section 7.6 closes.

7.1 ANSWERS TO THE THREE PUZZLES

7.1.1 PUZZLE 1 (ROOTS): *NWUKWU-* VS. *AMO-/AMWU-*

The first puzzle asked how Korean moved from an earlier system in which *amo* functioned as a broad 'someone' indefinite and *nwukwu* functioned as 'who?' , to the modern system in which *amwu-* is a polarity-bound 'anyone' -root and *nwukwu-* is the main 'who/someone' -root. The direct answer is that the two roots followed opposite but locally orderly trajectories on the semantic map.

For *nwukwu-*, the main change is expansion in function rather than loss or replacement. In the 18th century, around 94% of bare *nwukwu-* tokens are tagged as Q_WH questions, with only about 6% as non-questions. Over the 19th and 20th centuries, the share of non-question uses rises to roughly 40%, while Q_WH uses

remain the majority. Within the non-question set, most tokens fall into the PPI and NEUTRAL macro-environments (SK/SU and IR/Q), with smaller but steady bands of FREE-CHOICE (FC/GEN/UFC) and only a modest NPI strand (DN/AA). On Haspelmath' s map, this corresponds to a path from the interrogative node into the specific-unknown and irrealis nodes (SU/SK, IR, CA), and then into nearby free-choice nodes, without a strong drift into the negative corner. Bare *nwukwu-* thus changes from 'only Q_WH' to a mixed interrogative–indefinite root that lives around Q/SU/SK and the FC/GEN fringe, but not at the heart of the NPI sector.

Bare *amwu-* moves in the opposite direction. In the Early period (1700–1849), about three quarters of bare *amwu-* tokens are coded as PPI (positive, usually SK/SU), about 10% as FREE-CHOICE (FC/GEN/IND), and only about 6% as NPI (DN/AA). By the Mid period (1850–1899), NPI uses have risen to around 29% and PPI has dropped to about 49%. In the Late period (1900–1999), around 93% of bare *amwu-* tokens are NPI and only small remnants are PPI or FREE-CHOICE. Inside the NPI domain, the distribution is very skewed: almost all negative tokens are licensed in AA-type environments, and within AA the predicate *-eps-* 'not exist, be without' is dominant; DN with clausemate negation plays a secondary role. On the semantic map, this is a clear movement away from the existential/free-choice region into the NPI node, especially into AA with *-eps-*. On Jäger' s polarity cline, the same pattern can be phrased as a shift from positive/neutral uses into strong NPI status.

The root-level answer to Puzzle 1 is therefore clear. Bare *nwukwu-* expands along adjacent nodes from Q_WH into SU/SK and some FC/GEN uses, while bare *amwu-* retreats from earlier existential and free-choice uses into AA/DN-heavy NPI uses centered around lexical negation. The system ends up with *nwukwu-* covering the interrogative and non-negative indefinite space, and *amwu-* covering the strongly negative space. The two trajectories are complementary, but not perfectly mirror-symmetric: bare *nwukwu-* keeps a thin negative fringe, and bare *amwu-* keeps small

free-choice residues. What the data show is not perfect categorical partitioning, but strong specialization with limited residual overlap.

7.1.2 Puzzle 2 (-(*i*)*nka*): *nwukwu-i-nka*, *nwukwunka*, and missing **amwu-inka*

Q→DP shift for *nwukwu-i-nka*

The second puzzle asked why the Q-morpheme -(*i*)*nka* yields a productive wh-based epistemic series such as *nwukwunka* 'someone or other' , but no matching **amwu-inka* series. The first part of the answer is that analytic *nwukwu-i-nka* does in fact show a Q→DP shift, but only partially.

A useful early datum strengthens this interpretation. In *Akaksŭpnyŏng* (1713), the form *mwul-ppwun-i-nka* 'is it only water?' shows -(*i*)*nka* attaching to nominal material in a clearly clausal interrogative use. What this demonstrates most clearly is not an early indefinite function, but the opposite: early -(*i*)*nka* can attach to nominal material while still functioning as clause-typing Q morphology. The later indefinite use of *nwukwu*-(*i*)*nka* is therefore best understood as a subsequent reanalysis from a clausal source rather than as the straightforward continuation of an already nominal indefinite marker.

The corpus confirms that analytic *nwukwu-i-nka* begins as a clausal interrogative and only later develops DP uses. In the 1900–1930 data, every token of *nwukwu-i-nka* is annotated as a matrix or embedded question (Q_WH); there are no DP uses and no case-marked forms. From the 1930s to the 1950s, case-marked forms begin to appear: first *nwukwu-i-nka-lul* in rhetorical questions, and then clear subject, object, genitive, and oblique DP uses (*nwukwu-i-nka-ka*, *nwukwu-i-nka-lul*, *nwukwu-i-nka-uy*, *nwukwu-i-nka-ey*) in the 1940s and 1950s. If overt case on *nwukwu-i-nka* is taken as a sign of nominal reanalysis, the observed sequence is clear: first a purely clausal interrogative, then rhetorical questions with focal case, and finally a limited but stable set of epistemic indefinites ('someone, I do not say who') in DP roles.

Importantly, analytic *nwukwu-i-nka* never becomes a purely nominal form. Across 1900–1999, about 78% of its tokens are still tagged as questions and only 22% as DPs. Logistic modeling in §5.3.4 shows that question uses remain dominant even in late decades, and semantic-map analysis in §5.5 shows that analytic *nwukwu-i-nka* stays anchored at Q_WH, with only a modest extension into SU/SK/IR/CA and no FC or GEN uses. The answer to the reviewer' s question about clause marking is therefore straightforward: yes, in the analytic construction, -(*i*)*nka* still functions predominantly as a clause-typing element. The DP indefinite use emerges from that clausal source, but it does not erase it.

However, the pathway should not be overstated. What the historical record shows most clearly is the clause-level status of early -(*i*)*nka*; the intermediate stages that lead from such clause-typing uses to a productive epistemic indefinite are reconstructed from the later distribution rather than documented as a fully continuous chain. The INKA pathway is therefore well supported, but it should be treated as a strongly supported historical inference rather than as an exhaustively observed sequence at every intermediate point.

Specialization of analytic vs. contracted forms

Once the Q→DP path is open, the analytic and contracted forms split the work. The contracted form *nwukwunka* (historically *nwukwu-nka*) appears in the corpus from the late 19th century and behaves like a DP almost from the start. Subject uses are attested by the 1910s, object uses by the 1930s, and genitive and oblique uses by the early 1950s. In the tagged data, more than 95% of *nwukwunka* tokens are DPs rather than questions, and early DP examples already have a 'someone-or-other' flavor, often in recognition and identification contexts where a speaker first does not know who the referent is and then finds out.

On the semantic map, analytic *nwukwu-i-nka* and contracted *nwukwunka* occupy the same corner but fill it differently. Analytic *nwukwu-i-nka* sits at Q_WH, with a small halo of SU/SK/IR/CA uses. Contracted *nwukwunka* fills SU, SK, IR, and CA with large numbers of tokens, has some Q_POL and a thin AA/DN edge, and only a small fraction of Q_WH uses. This matches the idea that a clausal Q-element can enter the nominal domain without losing its clausal use, and that a phonologically reduced form can stabilize the DP interpretation and become the default epistemic indefinite. Analytic *nwukwu-i-nka* remains the ordinary 'who is it?' form, while contracted *nwukwunka* becomes the ordinary 'someone (I don' t know who)' form; neither develops into a core NPI or FC item.

The chronology of case-marking also raises a narrower syntactic question. Subject, object, and genitive DP uses of analytic *nwukwu-i-nka* are attested before regular oblique uses. This may partly reflect frequency, but it may also indicate that the earliest stable DP-like environments are those in which the sequence is most easily reanalyzed as a nominal argument or possessor rather than as a reduced clause. If so, the lag of the oblique is not wholly accidental. At the same time, since embedded clauses are commonly found in complement positions before nominalization, this asymmetry remains an open problem rather than a settled syntactic fact.

A related question concerns how far this kind of development extends beyond the specific case of *nwukwu-*(*i*)*nka*. Within Korean, the most obvious place to look is other wh-roots that combine with -(*i*)*nka*; cross-linguistically, Japanese wh+Q indefinites and 'dunno' -type indefinites provide further comparisons. The present dissertation has established the pathway in detail only for *nwukwu*, but the broader construction type may well prove to be more general.

The second part of Puzzle 2 asked why there is no **amwu-inka* series, despite the fact that *amwu*-based forms freely combine with other particles (*amwu-to*, *amwu-na*, *amwu-(i)lato*, etc.). The corpus suggests a principled explanation in terms of timing, source availability, and paradigm structure.

Earlier Korean does show *amo/amwu* in embedded ignorance constructions with *mol-* 'not know' ; examples from 1713 and 1876 have *amo-cen cwul mol-ola* and *amo-kin cwul mol-ola*. In principle, such configurations could have fed a Q→DP path for *amwu-inka*, parallel to the one for *nwukwu-i-nka*. But those *amo/amwu* + *mol-* uses disappear from the corpus by the late 19th century. By the time *nwukwunka* appears in 1895 and wh+-*(i)nka* starts to grammaticalize with *nwukwu*, *nwukwu* is still a live Q_WH root, whereas *amwu* has already moved far along the polarity cline into the NPI/FC corner (Chapter 4, §4.3.3).

Two consequences follow. First, when wh+-*(i)nka* becomes available as a Q→DP construction, there is no productive *amwu-i-nka?* template left that could serve as its clausal source. Second, the relevant space on the semantic map is already being divided up: SU/IR and neighbouring nodes are being filled by wh+-*(i)nka* forms such as *nwukwunka*, while AA/DN and FC/IND are being filled by *amwu*-series such as *amwu-to* and *amwu-na*. Overlap by itself is therefore not the issue. Transitional overlap is in fact part of the history: as wh-based forms expand, they do encroach on territory earlier associated with *amwu*. The crucial point is narrower. For a new *amwu-inka* series to stabilize, overlap would have had to be compensated for either by a live source construction or by some differentiable contribution within the existing paradigm. By the time a putative *amwu-inka* could have arisen, neither condition appears to be met.

The absence of *amwu-inka* is therefore not an arbitrary lexical gap. It follows from the order in which the Q→DP pathway opened for wh+-*(i)nka* and the way

polarity changes had already relocated *amwu-* to the negative and free-choice edges of the map. This also helps explain why the shift appears effectively unidirectional. A reverse development is not impossible in principle, but it would require renewed source constructions that place *amwu* back into interrogative-like or epistemic-indefinite environments, together with some functional payoff distinct from the already entrenched wh+-(*i*)*nka* series. The corpus offers no evidence for either condition.

7.1.3 Puzzle 3 (competition): shared particles *-to* and *-na*

-to: negative competition

The third puzzle asked how *amwu-* and *nwukwu-* divide the map when they share the same particle. The *-to* series shows that the outcome is not replacement but coexistence with specialization.

Earlier work reviewed in §2.8.1 argued, on synchronic grounds, that *amwu-to* is a maximally strong NPI ('not even anyone'), while *nwukwu-to* is polarity-sensitive but more flexible. The map-based corpus results sharpen this picture. In the earliest period (1900–1930), *amwu-to* already sits at the NPI core, with several hundred AA and DN tokens and only a few side uses; *nwukwu-to* appears mainly in CO and GEN, with very few DN tokens. In the mid-century period (1930–1960), both series enter DN/AA, but the counts show that *amwu-to* remains dominant in those nodes, while *nwukwu-to* has many more CO and GEN tokens. In the late period (1960–1990), *amwu-to* has thousands of AA and DN tokens and very little outside the NPI nodes, whereas *nwukwu-to* has a comparable number of DN tokens but also thousands of CO tokens and a visible cluster in FC/GEN/UFC.

The entry-time analysis confirms the same split. *Amwu-to* is entrenched first and foremost at distance 0 from the NPI core (AA/DN), whereas *nwukwu-to* first reaches the CO node and only later moves inward to DN/AA. The result is therefore

an asymmetrical overlap around the NPI core: *amwu-to* anchors the strongest AA/DN corner, while *nwukwu-to* links that corner to neighbouring comparative and weakly free-choice uses.

-*na*: free-choice competition

The *-na* series shows a different but equally structured partitioning. The synchronic literature reviewed in §2.8.2 distinguished an extensional, 'whoever/anyone-in-this-set' *nwukwu-na* from a more indiscriminative *amwu-na*. The semantic-map counts in Chapter 6 support and refine this.

In every period, wh+*-na* supplies the bulk of FC and GEN tokens. Already in 1900–1930, *nwukwu-na* has many more FC and GEN tokens than *amwu-na*; by 1960–1990 it has 4,630 FC and 9,860 GEN tokens, whereas *amwu-na* has 244 FC and 155 GEN tokens but 827 IND tokens. On the map, *nwukwu-na* fills the FC and GEN nodes and extends slightly into UFC and DN/Q_POL/Q_WH, while *amwu-na* clusters closer to IND with a smaller FC/GEN band. The entry-time analysis shows that *nwukwu-na* stabilizes first at FC and GEN, and only later touches DN, Q_POL, and UFC, whereas *amwu-na* stabilizes first at GEN and IND and only acquires a stable FC strand later. This matches H-COMP$_{na}$: *nwukwu-na* is the main extensional FCI, and *amwu-na* is the main source of 'just any X' readings.

Global answer to Puzzle 3

Puzzle 3 asked how shared particles interact with the earlier root-level and -(*i*)*nka* developments. The global answer is that shared-particle competition does not produce two interchangeable 'anyone' systems. It produces local coexistence with specialization.

Across *-to*, *amwu-to* dominates the AA/DN/AM core and is almost absent from FC/GEN/UFC, while *nwukwu-to* covers both DN/AA and a wide band of CO/

GEN/FC/UFC. Across *-na*, *nwukwu-na* fills FC and GEN and sends a small tail into DN/UFC/Q_POL/Q_WH, while *amwu-na* sits close to IND with a thinner FC/GEN trail and later entry into FC. In both series, the two roots do not simply duplicate each other. Instead, with the same particle, *amwu-* concentrates at the strong NPI and strong FC/IND edges, and *nwukwu-* occupies the central Q/SU/SK/IR and FC/GEN/UFC regions.

Viewed against the root-level and *-(i)nka* developments, this answer is coherent. The distribution of *amwu-to*, *amwu-na*, *nwukwu-to*, and *nwukwu-na* in the 20th century reflects both where the roots themselves ended up (Chapter 4) and how the clausal *-(i)nka* pathway carved out an epistemic corner for wh-based forms (Chapter 5). The shared particle series inherit this split and refine it: *amwu*-based series mark the strongest negative and indiscriminative corners, wh-based series mark the interrogative, specific-unknown, and extensional free-choice regions.

7.2 Theoretical implications

7.2.1 Paradigms as units of change

The broadest payoff of the dissertation lies in what it suggests about indefinites and, more generally, about how morphosyntax is organized over semantic space. The Korean evidence suggests that indefinite systems are best modelled not as inventories of independent form–meaning pairings, but as paradigms assembled through competition over a structured functional space. The important generalization is therefore not simply that one form means 'someone' and another means 'anyone'. It is that recurring source constructions, usage skew, and paradigm-internal differentiation jointly determine which forms stabilize where.

Korean is informative here not only because it has a rich indefinite system, but because the morphology makes the assembly of indefinite meaning unusually visible. Where English forms such as *someone* and *anyone* are often treated as lexically

given indefinites, Korean frequently builds comparable meanings by combining wh-bases with particles that contribute epistemic, concessive, disjunctive, scalar, or alternative-structural content. This makes Korean a particularly useful case for asking how indefinite meaning is compositionally assembled and why formally similar expressions can end up occupying sharply different regions of semantic space.

From this perspective, the Korean case contributes two broader claims. First, diachronic explanation may need to be stated at the level of the paradigm rather than the individual item. The modern distribution of *nwukwu-*, *amwu-*, *nwukwunka*, *amwu-to*, and *nwukwu-na* cannot be derived by following each item in isolation, because each series changes partly in response to the expansion, retreat, or entrenchment of the others. Second, the case suggests a general way in which morphosyntax can be built on semantic space. A language may repeatedly recruit clause-typing, concessive, scalar, and additive morphology into the nominal domain, but the outcome is not random: new series settle where they can inherit a viable source construction and maintain a stable niche relative to already existing competitors.

What is new here is therefore not simply another description of Korean indefinites, but a concrete demonstration that semantic space, source availability, and competition can be integrated into a single explanatory account of how a morphological system differentiates over time.

7.2.2 Semantic maps, continuity, and predictive value

One of the guiding ideas in Chapter 2 was Haspelmath' s continuity claim: when an indefinite gains or loses functions, it should do so by moving along neighbouring nodes on the map, rather than by 'jumping' across unrelated uses. The Korean evidence supports this structural claim very strongly.

At the root level (Chapter 4), bare *nwukwu-* moves from Q_WH into SU/SK and then into nearby irrealis and free-choice nodes (IR/CA, FC/GEN/UFC), but never acquires a dense band of DN/AA uses. Bare *amwu-* starts in the SK/SU/FC area and retreats into AA/DN under lexical negation, especially with *-eps-* 'not exist' . In the wh+-(*i*)*nka* domain (Chapter 5), analytic *nwukwu-i-nka* extends from Q_WH into SU/SK/IR/CA, and contracted *nwukwunka* fills the same cluster much more densely, with only a very thin DN edge. In the particle series (Chapter 6), *amwu-to* enters and stays in AA/DN, while *nwukwu-to* grows inward from CO; *nwukwu-na* enters at FC and expands into GEN/UFC, while *amwu-na* starts just off-centre in GEN/IND and only later develops a stable FC strand. In all of these cases, new uses are adjacent, in the graph sense, to older ones.

At the same time, the Korean evidence also clarifies the kind of predictive value that semantic maps do and do not have. The map predicts neighbourhoods, plausible pathways, and the kinds of local extensions and retreats that should be possible. It does not by itself predict exact timing, frequency, or which competitor will ultimately dominate a given niche. In that sense, its predictive force is constrained but real. It tells us where change can plausibly go next, but not when or under which social and paradigmatic conditions it must go there.

The entry-time analyses for *-to* and *-na* make this point clearly. If we regress first-attested decade for each label against its graph distance from a core node (NPI for *-to*, FC for *-na*), the slopes are small and not significant. Labels further from the core do not always enter later in calendar time. This is not a counterexample to continuity, but a reminder that starting points matter. For *-to*, *amwu-to* starts in the NPI core, while *nwukwu-to* starts at CO. For *-na*, *nwukwu-na* starts at FC, while *amwu-na* starts in GEN/IND. Once these different starting points are taken into account, the movements themselves are local: series do not skip nodes, but they

do begin from different corners, and their entry decades are shaped by that fact as well as by frequency, genre, and source availability.

The Korean case also helps answer a committee question about whether all developments conform perfectly to the lines of the semantic map or whether there are small surprises. The answer is that the dominant changes conform strongly, but there are informative irregularities. Analytic *nwukwu-i-nka* retains its clausal function longer than a simple one-way grammaticalization story might suggest. *Nwukwu-to* becomes robust first in CO before DN, and *amwu-na* stabilizes in GEN/IND before a narrower FC use becomes frequent. Bare *nwukwu-* keeps a thin negative fringe, and bare *amwu-* keeps small free-choice residues. These are not violations of the map, but reminders that the map is not a mechanical script. What it predicts are local possibilities and constraints, not a single obligatory sequence.

This perspective also helps frame the question of whether adjacency reflects learner behavior or a deeper representational organization. The present evidence bears most directly on the learner-facing side of the issue: the corpus shows forms being encountered repeatedly in neighbouring environments and then becoming entrenched there. That supports a usage-based explanation in which learners generalize locally from the distributions available to them. At the same time, the persistent clustering of adjacent functions is also compatible with a more representational view, on which neighbouring nodes share part of their featural composition and therefore resist simultaneous jumps across multiple functions. The dissertation does not resolve that larger theoretical question, but it shows that the Korean data are at least consistent with both views and directly support the distributional one.

7.2.3 Polarity clines, learning, and competition

A second set of implications concerns polarity. Jäger' s feature-based cline (from plain existential through increasingly restricted NPIs to negative indefinites) was used in Chapter 4 as a way of describing how *amwu-* and *nwukwu-* might change. The Korean evidence shows both directions of movement and shows how the cline interacts with the map.

Bare *amwu-* is a clear case of drift toward the NPI end. In the Early period, it is a fairly ordinary existential, with many SK/SU and some FC uses. By the Late period, more than 90% of its tokens are NPI, and almost all of those NPI tokens are in AA/DN environments, particularly under lexical negation with *-eps-*. Morphology then reinforces this drift. The *-to* series yields *amwu-to*, which behaves as a maximally strong negative form ('not even anyone') and sits almost entirely in AA/DN. The concessive and free-choice particles *-na*, *-(i)lato*, and *-tunci* yield *amwu-na*, *amwu-(i)lato*, and *amwu-tunci*, which cluster at the strong FC/IND and NO_MATTER edges, often with indiscriminative 'no matter who/what' flavor.

Nwukwu-, by contrast, shows extension away from pure Q_WH into non-negative existential and free-choice territory without crossing into strong NPI territory. At the bare-root level, *nwukwu-* acquires a substantial SU/SK/IR/FC halo but keeps Q_WH as its main use and never develops a dominant NPI band. In the wh+-*(i)nka* pathway, *nwukwu-i-nka* remains mostly clausal, and *nwukwunka* specializes in epistemic 'someone' uses, again in SU/SK/IR/CA with only a thin DN edge. Among the particle series, *nwukwu-to* is a strong NPI, but less extreme than *amwu-to*, and also carries many CO/GEN/FC/UFC uses; *nwukwu-na* is the main extensional FC item, living in FC/GEN/UFC with only a small number of DN/Q_POL/Q_WH uses.

Seen through the map, the polarity cline is not a separate axis but a re-description of where tokens accumulate. Polarity strengthening corresponds to

the movement of tokens into the NPI nodes (DN/AA/AM), and weakening to their reduction there. This supports viewing 'polarity behavior' as a summary of distributions on the map rather than as a primitive lexical feature: the categories 'strong NPI' , 'weak NPI' , and 'non-polar' can be captured by where and how often a form occurs, not only by stipulating a [NEG] value in the lexicon.

At the same time, skewed input should not be understood as automatically leading only to stasis. When the input distribution is stable, repeated association between a form and one region of the map will indeed reinforce the existing allocation. But when the distribution itself is being reshaped—because a competing form expands into neighbouring functions or because one set of constructions becomes more common than another—learners are exposed to changing evidence. Under those conditions, skewed input no longer preserves the old analysis; it favours a new one. Transitional overlap is therefore not a contradiction of the learning account, but part of the mechanism by which reallocation becomes possible. Once the new allocation is repeatedly reinforced, stasis returns at a new point in the system.

This perspective also bears on the question of what might happen next. Given the modern state of affairs, one plausible expectation is further consolidation: bare *amwu-* is unlikely to recover a broad existential profile unless new source constructions place it again in non-negative epistemic environments, and the current system provides no sign of that. More likely, the split between negative and free-choice uses will continue to be maintained through the particle-bearing paradigms rather than through a renewed expansion of the bare root. This remains speculative, but the present model predicts continued specialization more readily than reversal.

7.2.4 Grammaticalization paths

A third set of implications concerns grammaticalization sources. Two main paths were in focus:

- interrogative → epistemic indefinite, via wh-clauses with -(*i*)*nka*; and
- concessive / scalar → NPI / strong FC, via *amwu-* combined with particles such as *-to*, *-na*, -(*i*)*lato*, and *-tunci*.

For the interrogative path, the starting point is clausal: copular wh-questions with -(*i*)*nka* ('who is it?') and related embedded ignorance clauses ('I do not know who it is'). The Korean data show that this clausal -(*i*)*nka* moves into the nominal domain by first allowing case marking on *nwukwu-i-nka*, and then by contracting to *nwukwunka*. Crucially, neither analytic *nwukwu-i-nka* nor contracted *nwukwunka* becomes a central NPI or FC item. Their distribution stays in the interrogative and epistemic region of the map: Q_WH plus SU/SK/IR/CA, with only a thin DN edge. The path is therefore best summarized as interrogative → epistemic indefinite, not interrogative → NPI or interrogative → free choice.

For the concessive/scalar path, the starting environment is different. The *amwu-* series is attested early in contexts that combine a minimal element with a scalar or concessive flavor: *amwu-to* behaves like an *even+minimizer* construction in An' s (2007) terms, and *amwu-*(*i*)*lato* and *amwu-tunci* appear in 'even if it is any X' / 'no matter who X' contexts. These source constructions are already close to anti-additive (AA) and strong FC/IND uses. Diachronically, the corpus shows that *amwu-to* is entrenched from the start in AA/DN, that *amwu-*(*i*)*lato* and *amwu-tunci* have large numbers of NO_MATTER and IND tokens, and that *amwu-na* accumulates many IND tokens alongside a more modest FC strand. In other words, concessive and scalar sources feed directly into the strong NPI and strong FC corners of the map.

These two paths cross-cut the typology in a useful way. Typological work often contrasts Q-based indefinites and concessive-based indefinites. The Korean material shows both in one system and makes their division of labor clear. Wh-based -(*i*)*nka* yields epistemic 'someone-or-other' indefinites that stay in the interrogative/epistemic zone. Concessive/scalar combinations with *amwu-* yield NPIs and strong FC items that sit at AA/DN and FC/IND. The fact that these series coexist and partition the map, rather than collapsing into a single 'any' -like family, suggests that grammaticalization paths are shaped not just by local reanalyses but also by how existing series already occupy the map. A Q-based path need not continue all the way into the NPI corner if a strong NPI series is already present; a concessive-based path need not yield epistemic 'someone' forms if that corner is already taken by wh+-(*i*)*nka*.

At the same time, the Korean evidence is not equally strong for every particle. The interrogative → epistemic-indefinite pathway is clearest for -(*i*)*nka*. For *-na*, by contrast, the grammaticalization pathway remains open. Two possibilities remain plausible on current evidence: either *amwu*-(*i*)*na* was already established as a DP-level free-choice / 'anyone' pattern, and later wh-based forms such as *nwukwu*-(*i*)*na* spread by analogy from that schema; or *-na* reflects an independent clause→DP reanalysis from alternative-conditional or 'whether' structures. These scenarios are not equivalent, and the current corpus evidence does not yet allow a confident choice between them.

7.2.5 Socio-historical setting of the changes

The timing of the grammatical developments traced in this dissertation is not random. They line up strikingly with well-documented shifts in the written and social life of Korean from the late nineteenth century onward, and these external changes help explain why certain indefinite series stabilize precisely when they do.

The sharp rise in NPI uses of bare *amwu-* between roughly 1880 and 1930, and the corresponding decline of its earlier existential and free-choice uses, coincide with the decades in which Korean-language newspapers, mixed-script expository prose, and *hangul*-only fiction become stable, high-volume genres. The gradual expansion of bare *nwukwu-* from almost exclusively Q_WH into SU/SK/NEUTRAL and some FC contexts is likewise in place by the time vernacular fiction and orthographic standardization fix a more stable written norm. The Late-period distributions for both roots are therefore measured on a textual landscape where interrogatives and indefinites are regularly written for a mass readership, and where school-based literacy spreads those forms widely.

The same timing matters for -(*i*)*nka*. The corpus shows a clear mismatch between the age of -(*i*)*nka* as a clause-typing element and the much later rise of wh+-(*i*)*nka* indefinites. Early texts contain many instances of -(*i*)*nka* as Q morphology, but not wh+-(*i*)*nka* strings used as simple NPs 'someone' in declarative clauses. The first clear instance of a wh+-(*i*)*nka* form used as a noun phrase appears much later, in 1895, as *nwukwunka*. This raises a natural 'why now?' question. A full socio-historical answer is beyond the scope of the dissertation, but it is at least plausible that the late rise of wh+-(*i*)*nka* as an indefinite reflects the convergence of two conditions: intensified use of embedded ignorance constructions in vernacular prose, and a textual environment in which a reduced 'someone-or-other' form could become frequent enough to stabilize.

These correlations do not prove a direct causal link, but they do place the semantic-map reorganizations traced in Chapters 4 through 6 within a coherent socio-historical setting. The drift of *amwu-* toward the NPI/FC end of the polarity cline, the expansion of *nwukwu-* beyond pure questions, and the rise of wh+-(*i*)*nka* epistemic indefinites all unfold during the broader shift to a vernacular, school-mediated, and increasingly standardized Korean in print.

Alongside its substantive claims about Korean indefinites, the dissertation also puts a specific empirical toolkit on the table. The core ingredients are simple, but using them together allows reasonably fine-grained claims about diachrony without leaving the corpus.

First, the study combines two large diachronic resources—the History corpus (roughly 1700–1890) and the Yonsei corpus (1900–1990)—and builds a hand-checked token set for the series that matter most (*nwukwu-*, *amo-/amwu-*, wh+-(*i*)*nka*, *amwu-to/nwukwu-to*, *amwu-na/nwukwu-na*, and related forms). All automatic hits are manually screened to remove lexicalized expressions and homographs, so the analysis is grounded in curated rather than raw string counts.

Second, the dissertation develops a token-level annotation procedure based on an extended semantic map. Each token receives a fine-grained label from the set SK, SU, IR, Q, CA, CO, DN, AA, AM, FC, GEN, UFC, IND, plus a small number of auxiliary labels where necessary. These labels are not stipulated case by case, but derived by a common decision procedure that combines tests for specificity, force, clause type, and type of negative or modal environment. For many purposes, the fine-grained labels are then grouped into the five macro-environments Q_WH, PPI, Neutral, NPI, and Free-choice, so that both detailed and coarse-grained patterns can be tracked in a single system.

Third, the dissertation combines these annotations with deliberately modest temporal models: decade-wise pmw rates, logistic and segmented logistic regressions, and entry-threshold analyses. The value of this method is not that it provides a new statistical technique, but that it lets token-level semantic labels and semantic-map structure be related transparently to diachronic change. Nothing in this setup is Korean-specific. Any language with reasonably deep historical corpora and a typologically informed map of indefinite functions could, in principle, be examined

with the same combination of hand-checked annotation and lightweight temporal modelling.

7.4 Limitations and open issues

The picture that emerges in this dissertation is constrained by the data, the annotation scheme, and the analytic choices that were made. In this section, I spell out some of the main limitations and point to questions that remain open.

7.4.1 Data limitations

Although the combined History and Yonsei corpora give a long time span (roughly 1700–1990), coverage is uneven. Several early decades in History are represented by very small amounts of text, and some decades have no material at all. Normalized per-million values in these decades are therefore based on very few tokens and must be interpreted with care. Whenever a change seems to start in such a sparsely attested decade, it is possible that the actual onset was earlier or later than the corpus suggests.

The genre balance is also skewed. Both corpora mainly contain written and edited prose (fiction, religious writing, official documents, newspapers, magazines, and some reference material). Spontaneous spoken data and informal dialogue are under-represented, especially in the earlier period. This means that the patterns traced here are mainly patterns of written usage.

7.4.2 Annotation limits

The labelling scheme is deliberately simple: each token receives a single semantic-map label such as SK, SU, IR, Q, CA, CO, DN, AA, AM, FC, GEN, UFC, or IND. This keeps the analysis tractable, but it inevitably flattens some distinctions. There are borderline cases between IR and CA, between FC, GEN, and IND within the

free-choice region, and between DN and AA in the negative region. The decision procedure in Appendix A tries to handle these cases consistently, but some tokens could reasonably be classified in more than one way.

In addition, the annotations are based entirely on written text. Important dimensions such as prosody, intonational focus, and real-time processing are not directly observable. Likewise, scope information is inferred from syntax and context rather than measured experimentally. As a result, the labels capture broad functional roles, but they do not fully encode all aspects of interpretation that might matter for a complete theory of indefinites and polarity.

7.4.3 Analytic limits

Analytically, the dissertation focuses on the two roots *nwukwu-* and *amo-/amwu-* and on a small set of particles (*-(i)nka*, *-to*, *-na*, *-(i)lato*, *-tunci*). Other indefinites in Korean—for example, expressions based on generic nouns, numerals, or different particles—are mentioned only briefly or not at all. The claim is not that these other forms are unimportant, but that *nwukwu-* and *amwu-* plus their main particles already illustrate a rich system of polarity and competition that can be explored in detail.

On the semantic side, I have not attempted to build a complete compositional system for all series. For key items such as *amwu-to*, *amwu-na*, *nwukwunka*, and wh+*-(i)nka*, I align with existing proposals and use them as a guide to interpret the corpus facts. A fully explicit semantic account would require spelling out the denotations of all roots, particles, and combinations, and tracing how these interact with the syntax and discourse structure of Korean. That task lies beyond the present study.

Finally, the statistical models are intentionally modest. Logistic and segmented regressions, macro-environment contrasts, and entry-threshold analyses provide a

compact way to describe changes, but they do not exhaust all the structure in the data. More elaborate models incorporating genre as a predictor, allowing for non-monotonic change, or modelling individual constructions could refine some of the claims made here.

7.4.4 Open issues in particle pathways

One open issue concerns how strongly the dissertation should generalize beyond INKA. For INKA, the evidence is strong enough to defend a clause-to-DP pathway: early data show clear clause-level Q uses of *-(i)nka*, and the indefinite use of *nwukwu-(i)nka* emerges only later, with robust DP-like behavior becoming visible in the 20th century. Even here, however, the conclusion should be stated with some caution. The pathway is best treated as strongly supported rather than exhaustively demonstrated at every intermediate stage.

For INA, the present evidence does not justify an equally strong parallel grammaticalization claim. Two possibilities remain live. On one view, *amwu-(i)na* was already established as a DP-level free-choice / 'anyone' pattern, and later wh-based forms such as *nwukwu-(i)na* spread by analogy from that schema. On another view, INA may reflect an independent clause-to-DP pathway, perhaps from alternative-conditional or 'whether' constructions in which *-na* quantifies over alternatives. The current corpus evidence does not yet allow a confident choice between these scenarios.

7.5 Future directions

The dissertation has concentrated on two roots, a small set of particles, and a specific semantic-map / polarity framework applied to written Korean up to 1990. This section sketches some ways in which that work can be extended and deepened.

7.5.1 Prosody, processing, and map labels

In the present study, map labels such as SK, SU, IR, FC, IND, DN, and AA are assigned on the basis of written context alone. A natural next step is to link these labels to prosodic and processing evidence. Spoken Korean has rich intonational resources, and recent work suggests that wh-phrases and indefinites show systematic differences in pitch range, prominence, and phrasing even when segmentally identical. Future work could combine semantic-map annotation with prosodic measurements from spoken corpora or laboratory recordings, asking whether SU-, IR-, FC-, or IND-like uses have distinct prosodic profiles. Processing experiments could likewise test whether different map functions lead to different expectations in real time.

7.5.2 Extending beyond *nwukwu* and *amwu*

The present analysis has focused on human-denoting *nwukwu*- and *amwu*-series because these are frequent and central in the system. Other indefinites in Korean deserve similar treatment. On the wh side, roots such as *mwues* 'what' , *eti* 'where' , and *etten* 'which' combine with the same particles and may show parallel but not identical trajectories. On the *amwu* side, non-human series such as *amwu kes* 'anything' also appear with *-to*, *-na*, *-(i)lato*, and *-tunci*. Systematically annotating such forms would show whether the same narrowing into AA/DN and FC/IND occurs more generally.

Particles that were treated only briefly in this dissertation, such as *-(i)lato* and *-tunci*, also deserve fuller study. The present chapters only sketch their map positions; a more detailed study could follow their entry into new functions decade by decade, as was done here for *-to* and *-na*.

7.5.3 Cross-linguistic comparison

The Korean system raises cross-linguistic questions. Many languages have both interrogative-based indefinites and dedicated indefinite roots; many also have particles that resemble Korean *-to* and *-na* in meaning and distribution. A future strand of work would place the Korean results alongside similar systems. Japanese is an obvious comparison point. It has wh-based series such as *dare-ka*, *dare-mo*, and *dare-demo*, and a semantic-map and polarity-based analysis parallel to the present one could ask whether Japanese shows a similar split between epistemic, negative, and free-choice regions. Greek offers another useful comparison, with its *kapjos*, *kanenas*, and *opjosdhipote*. More generally, one could ask whether the combination of a wh-root and a dedicated indefinite root, both interacting with a small set of particles, tends to produce similar competition patterns across languages.

7.5.4 Beyond the 1990s

Finally, the present study stops at 1990 because that is the upper limit of the Yonsei corpus. In the decades since then, Korean has seen rapid changes in media, literacy, and style, and new forms may have emerged or shifted in frequency. Extending the corpus window to roughly 2000–2020, using web corpora, recent newspapers, online forums, and spoken material, would make it possible to test whether the specializations described here have stabilized.

Given the current model, one plausible expectation is further consolidation: bare *amwu-* remains strongly negative, while the particle-bearing paradigms maintain the split between negative and indiscriminative/free-choice uses. Any renewed broad existential use of bare *amwu-* would require new source constructions that place it again in non-negative epistemic environments, and the present system provides no sign of that. The post-1990 record is precisely where such predictions can begin to be tested.

7.6 Closing remarks

The chapters of this dissertation have approached Korean indefinites from several angles—roots, particles, polarity behavior, and grammaticalization—but a single theme has run through all of them. Korean does not have one undifferentiated 'indefinite system' . Instead, it has a set of constructional subsystems organized on a semantic map, with roots and particles jointly determining which regions each series occupies. The interrogative root *nwukwu-* and the dedicated root *amwu-* start in different parts of the space and move in different directions; clause-typing morphology like *-(i)nka* enters the nominal domain along one path; concessive and scalar particles such as *-to* and *-na* enter along others. The result is a paradigm in which *nwukwunka*, *amwu-to*, *nwukwu-na*, *amwu-na*, and related forms each carve out contiguous, partially overlapping regions on Haspelmath' s map rather than competing as interchangeable 'any' -words.

Seen in a broader perspective, the Korean case adds several clear points to the cross-linguistic picture of indefinites and polarity. First, it shows that interrogative-based indefinites need not drift toward NPI or strong FC status; they can remain in the interrogative/epistemic corner even when they are fully lexicalized. Second, it illustrates how dedicated indefinite roots that start out as plain existentials can shift into strong NPI and FC uses when they are repeatedly embedded under negation and concessive morphology. Third, it shows that when interrogative roots and dedicated roots coexist, they may end up specializing for different tasks—one anchoring Q-based and epistemic indefinites, the other anchoring NPI-based and FC-based forms—rather than merging into a single series.

The broader lesson is therefore simple but, I think, important: what looks at first glance like a simple indefinite opposition may in fact reflect the interaction of

clause structure, particle semantics, discourse status, and paradigm-level competition. The Korean evidence suggests that indefinite systems are not merely inventories of forms. They are structured paradigms in which morphosyntax is built and rebuilt over semantic space. That is the main claim this dissertation has tried to make visible.

References

Aguilar-Guevara, A., Aloni, M., Port, A., Šimík, R., de Vos, M., & Zeijlstra, H. (2010). Semantics and pragmatics of indefinites: Methodology for a synchronic and diachronic corpus study. In *Beyond semantics: Corpus-based investigations of pragmatic and discourse phenomena* (pp. 1–16).

An, D. (2007). On the distribution of NPIs in korean. *Natural Language Semantics*, *15*, 317–350.

Baayen, R. H. (2008). *Analyzing linguistic data: A practical introduction to statistics using r*. Cambridge University Press. https://doi.org/10.1017/CBO9780511801686

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Biber, D. (2012). Register as a predictor of linguistic variation. *Corpus Linguistics and Linguistic Theory*, *8*(1), 9–37.

Biber, D., & Conrad, S. (2019). *Register, genre, and style*. Cambridge University Press.

Bybee, J. (2010). *Language, usage and cognition*. Cambridge University Press. https://doi.org/10.1017/CBO9780511750526

Chang, S.-J. (1996). *Korean*. John Benjamins.

Chierchia, G. (2001). A puzzle about indefinites. In C. Cecchetto, G. Chierchia, & M. T. Guasti (Eds.), *Semantic interfaces: Reference, anaphora, and aspect* (pp. 51–89). CSLI Publications.

Chierchia, G. (2013). *Logic in grammar: Polarity, free choice, and intervention*. Oxford University Press.

Choi, J. (2007). *Free choice and negative polarity: A compositional analysis of korean polarity sensitive items* [Doctoral dissertation, University of Pennsylvania].

Choi, Y.-J. (2011). Correlation between disjunction and modality: Focused on *inka* [In Korean]. *Journal of Korean Linguistics*, *60*, 146–181.

Chung, D. (2012). Is *amwu-N-to* a negative quantifier? *Linguistic Research*, *29*(3), 541–562.

Dayal, V. (1998). Any as inherently modal. *Linguistics and Philosophy*, *21*, 433–476.

Dayal, V. (2004). The universal force of free choice any. *Linguistic Variation Yearbook*, *4*(1), 5–40. https://doi.org/10.1075/livy.4.02day

Degano, M., & Aloni, M. (2022). Indefinites and free choice: When the past matters. *Natural Language & Linguistic Theory*, *40*(2), 447–484.

Dryer, M. S., & Haspelmath, M. (Eds.). (2013). *The world atlas of language structures online* [Data set] (v2020.4). https://doi.org/10.5281/zenodo.13950591

Eckardt, R. (2006). *Meaning change in grammaticalization: An enquiry into semantic reanalysis*. Oxford University Press.

Enç, M. (1991). The semantics of specificity. *Linguistics Inquiry*, *22*(1), 1–25.

Espinal, M. T. (1991). The representation of disjunct constituents. *Language*, *67*(4), 726–762. https://doi.org/10.2307/415075

Farkas, D. (2002). Specificity distinctions. *Journal of Semantics*, *19*, 1–31.

Fodor, J. D., & Sag, I. (1982). Referential and quantificational indefinites. *Linguistics and Philosophy*, *5*, 355–398.

Giannakidou, A. (1998). *Polarity sensitivity as (non)veridical dependency*. John Benjamins.

Giannakidou, A. (2000). Negative ⋯concord? *Natural Language & Linguistic Theory*, *18*, 457–523.

Giannakidou, A. (2001). The meaning of free choice. *Linguistics and Philosophy*, *24*, 659–735.

Giannakidou, A. (2011). Negative polarity and positive polarity: Licensing, variation, and compositionality. In K. von Heusinger, C. Maienborn, & P. Portner (Eds.), *Semantics: An international handbook of natural language meaning, vol. 2* (pp. 1660–1712). De Gruyter.

Giannakidou, A., & Quer, J. (2013). Exhaustive and non-exhaustive variation with free choice and referential vagueness: Evidence from greek, catalan, and spanish. *Lingua*, *126*, 120–149.

Giannakidou, A., & Yoon, S. (2016). Scalar marking without scalar meaning: Non-scalar, nonexhaustive 'even'-marked NPIs in greek and korean. *Language*, *92*, 522–556.

Groenendijk, J., & Stokhof, M. (1991). Dynamic predicate logic. *Linguistics and Philosophy*, *14*(1), 39–100. https://doi.org/10.1007/BF00628304

Haspelmath, M. (1997). *Indefinite pronouns*. Oxford University Press.

Haspelmath, M. (2003). The geometry of grammatical meaning: Semantic maps and cross-linguistic comparison. In M. Tomasello (Ed.), *The new psychology of language, vol. 2* (pp. 211–242). Lawrence Erlbaum.

Haspelmath, M. (2005). Indefinite pronouns. In M. Haspelmath, M. S. Dryer, D. Gil, & B. Comrie (Eds.), *The world atlas of language structures* (pp. 190–193). Oxford University Press.

Haspelmath, M. (2013). Indefinite pronouns. In M. S. Dryer & M. Haspelmath (Eds.), *The world atlas of language structures online*. Max Planck Institute for Evolutionary Anthropology. Retrieved December 11, 2025, from https://wals.info/chapter/46

Heim, I. (1982). *The semantics of definite and indefinite noun phrases* [Doctoral dissertation, University of Massachusetts Amherst]. https://semanticsarchive.net/Archive/jA2YTJmN/Heim%20Dissertation%20with%20Hyperlinks.pdf

Hilpert, M. (2013). *Constructional change in english: Developments in allomorphy, word formation, and syntax*. Cambridge University Press. https://doi.org/10.1017/CBO9781139004206

Hiraiwa, K., & Nakanishi, K. (2021). Japanese free choice and existential indeterminates as hidden clauses. In *Proceedings of the 15th workshop on altaic formal linguistics* (*WAFL 15*) (pp. 45–58, Vol. 93).

Hoeksema, J. (1998). On the (non)loss of polarity sensitivity: Dutch *ooit*. In R. Hogg & L. van Bergen (Eds.), *Historical linguistics 1995* (pp. 101–114). John Benjamins.

Hoeksema, J. (2010). Dutch *enig*: From nonveridicality to downward entailment. *Natural Language & Linguistic Theory*, *28*, 743–784.

Hopper, P. J., & Traugott, E. C. (1993). *Grammaticalization*. Cambridge University Press.

Horn, L. (1989). *A natural history of negation*. University of Chicago Press.

Israel, M. (1996). Polarity sensitivity as lexical semantics. *Linguistics and Philosophy*, *19*, 619–666.

Israel, M. (2004). The pragmatics of polarity. In L. Horn & G. Ward (Eds.), *The handbook of pragmatics* (pp. 701–723). Blackwell.

Jäger, A. (2010). Anything is nothing is something: On the diachrony of polarity types of indefinites. *Natural Language & Linguistic Theory*, *28*, 787–822.

Jang, Y. (1999). Two types of question and existential quantification. *Linguistics*, *37*, 847–869.

Kadmon, N. (1987). *On unique and non-unique reference and asymmetric quantification* [Doctoral dissertation, University of Massachusetts Amherst].

Kadmon, N., & Landman, F. (1993). Any. *Linguistics and Philosophy*, *16*, 353–422.

Kamp, H. (1981). A theory of truth and semantic representation. In J. A. G. Groenendijk, T. M. V. Janssen, & M. J. B. Stokhof (Eds.), *Formal methods in the study of language* (pp. 277–322, Vol. 135). Mathematisch Centrum.

Kang, A. (2014). A novel *wh*-indeterminate in korean: *wh-inka* as a marker of referential vagueness. In M. Kenstowicz, T. Levin, & R. Masuda (Eds.), *Japanese/korean linguistics 23*. CSLI Publications.

Kang, A. (2021). Path to grammaticalization of referentially vague indefinite *wh-inka* in Korean [Chungnam National University].

Karttunen, L. (1976). Discourse referents. In J. D. McCawley (Ed.), *Syntax and semantics 7: Notes from the linguistic underground* (pp. 363–386). Academic Press.

Karttunen, L. (1977). Syntax and semantics of questions. *Linguistics and Philosophy*, *1*, 3–44.

Kim, A.-R. (2000). *A derivational quantification of* "*wh-phrase*" [Doctoral dissertation, Indiana University].

Kim, J. (2020a). On the semantics of an even-based polarity sensitive item, *wh*-(*N*)-*lato*. *Linguistic Research*, *37*(2), 147–186. https://doi.org/10.17250/khisli.37.2.202006.001

Kim, J. (2020b). Semantics of *-lato*. *Korean Journal of Linguistics*, *45*(1), 69–96.

Kim, J.-B. (2024). Fragment answers with negative dependencies in korean: A direct interpretation approach. *Linguistics*, *62*(2), 385–419. https://doi.org/10.1515/ling-2021-0134

Kim, M., & Kaufmann, S. (2006). Domain restriction in freedom of choice: A view from korean *Indet-na* items. In E. Puig-Waldmüller (Ed.), *Proceedings of sinn und bedeutung 11* (pp. 375–389). Universitat Pompeu Fabra.

Kim, M. (2015). From choice to counter-expectation: Semantic–pragmatic connections of the korean disjunctive, concessive, and scalar focus particle *-na*. *Journal of Pragmatics*, *80*, 1–21.

Klima, E. S. (1964). Negation in English. In J. A. Fodor & J. Katz (Eds.), *The structure of language* (pp. 246–323). Prentice Hall.

Kratzer, A. (1998). Scope or pseudoscope? are there wide-scope indefinites? In S. Rothstein (Ed.), *Events and grammar* (pp. 163–196, Vol. 70). Springer. https://doi.org/10.1007/978-94-011-3969-4_8

Kratzer, A., & Shimoyama, J. (2002). Indeterminate pronouns: The view from japanese. In Y. Otsu (Ed.), *Proceedings of the third tokyo conference on psycholinguistics* (pp. 1–25). Hituzi Syobo.

Krifka, M. (1995). The semantics and pragmatics of polarity items. *Linguistic Analysis*, *25*, 209–257. https://amor.cms.hu-berlin.de/~h2816i3x/POLARITY.pdf

Kroch, A. S. (1989). Reflexes of grammar in patterns of language change. *Language Variation and Change*, *1*(3), 199–244. https://doi.org/10.1017/S0954394500000168

Kuroda, S.-Y. (1965). *Generative grammatical studies in the japanese language* [Doctoral dissertation, Massachusetts Institute of Technology].

Labov, W. (1994). *Principles of linguistic change, volume 1: Internal factors*. Blackwell.

Ladusaw, W. A. (1979). *Polarity sensitivity as inherent scope relations* [Doctoral dissertation, University of Texas at Austin].

Ladusaw, W. A. (1980). *Polarity sensitivity as inherent scope relations*. Garland Publishing.

Lahiri, U. (1998). Focus and negative polarity in Hindi. *Natural Language Semantics*, *6*, 57–123.

Lakoff, G. (1974). Syntactic amalgams. In M. Galy, R. Fox, & A. Bruck (Eds.), *Papers from the tenth regional meeting of the chicago linguistic society* (pp. 321–

344). Chicago Linguistic Society. https://george-lakoff.com/wp-content/uploads/2011/01/syntactic-amalgams-lakoff-1974.pdf

Lee, C. (1996). Negative polarity items in English and korean. *Language Sciences*, *18*, 505–523.

Lee, C., Chung, D., & Nam, S. (2000). The semantics of *amwu N-to/-irato/-ina* in korean—arbitrary choice and concession. *Language and Information*, *4*(2), 107–123.

Lee, J.-H. (2010). Nonveridical dependency of korean particle *-ilato* and japanese *-demo*. In S. Iwasaki et al. (Eds.), *Japanese/korean linguistics 17* (pp. 231–245). CSLI Publications.

Lim, D. (2017). The semantics of *-lato* in korean. In C. Lee, F. Kiefer, & M. Krifka (Eds.), *Contrastiveness in information structure, alternatives, and scalar implicatures* (pp. 203–226). Springer.

Linebarger, M. C. (1980). *The grammar of negative polarity* [Doctoral dissertation, Massachusetts Institute of Technology].

Linebarger, M. C. (1987). Negative polarity and grammatical representation. *Linguistics and Philosophy*, *10*, 325–387.

Matthewson, L. (1999). On the interpretation of wide-scope indefinites. *Natural Language Semantics*, *7*, 79–134.

Menéndez-Benito, P. (2010). On universal free choice items. *Natural Language Semantics*, *18*, 33–64.

Montague, R. (1973). The proper treatment of quantification in ordinary english. In J. Hintikka, J. Moravcsik, & P. Suppes (Eds.), *Approaches to natural language* (pp. 221–242). D. Reidel. https://doi.org/10.1007/978-94-010-2506-5_10

Muggeo, V. M. R. (2003). Estimating regression models with unknown breakpoints. *Statistics in Medicine*, *22*, 3055–3071. https://doi.org/10.1002/sim.1545

Muggeo, V. M. R. (2008). Segmented: An r package to fit regression models with broken-line relationships. *R News*, *8*(1), 20–25. https://journal.r-project.org/articles/RN-2008-004/

Nakanishi, K. (2021). A clausal analysis of free choice *demo* in japanese. *Proceedings of the Linguistic Society of America*, *6*(1), 1024–1038.

Park, E.-H. (2009). *Wh-indeterminates, free choice, and expressive content in korean* [Doctoral dissertation, University of Chicago].

Park, J.-H. (2007). The pronominal system of korean viewed from a typological perspective. *Journal of Korean Linguistics*, *50*, 115–147.

Partee, B. H. (1987). Noun phrase interpretation and type-shifting principles. In J. Groenendijk, D. de Jongh, & M. Stokhof (Eds.), *Studies in discourse representation theory and the theory of generalized quantifiers* (pp. 115–143). Foris.

Progovac, L. (1993). *Positive and negative polarity: A binding approach*. Cambridge University Press.

Reinhart, T. (1997). Quantifier scope: How labor is divided between QR and choice functions. *Linguistics and Philosophy*, *20*, 335–397.

Ross, J. R. (1969). Guess who? In R. I. Binnick, A. Davison, G. M. Green, & J. L. Morgan (Eds.), *Papers from the fifth regional meeting of the chicago linguistic society* (pp. 252–286). Chicago Linguistic Society.

Shimoyama, J. (2001). *Wh-constructions in japanese* [Doctoral dissertation, University of Massachusetts Amherst].

Shimoyama, J. (2006). Indeterminate phrase quantification in japanese. *Natural Language Semantics*, *14*(2), 139–173.

Sohn, H.-M. (2001). *The korean language*. Cambridge University Press.

Suh, C.-S. (1989). Interrogatives and indefinite words in korean: With reference to japanese. In *Harvard studies in korean linguistics 3* (pp. 329–340).

Suh, C.-M. (1987). *Study of questions in korean*. Tap Publishing Co.

Suh, C.-M. (1989). Wh-constructions in korean. In *Harvard studies in korean linguistics 3* (pp. 517–526).

Traugott, E. C., & Dasher, R. B. (2002). *Regularity in semantic change*. Cambridge University Press.

Uegaki, W. (2018). A unified semantics for the japanese Q-particle *ka* in indefinites, questions and disjunctions [Article 14]. *Glossa: A Journal of General Linguistics*, *3*(1), 1–45. https://doi.org/10.5334/gjgl.238

von Heusinger, K. (2011). Specificity, referentiality and discourse prominence: German indefinite demonstratives. In I. Reich, E. Horch, & D. Pauly (Eds.), *Proceedings of sinn & bedeutung 15* (pp. 9–30). Saarland University Press.

Winter, Y. (1997). Choice functions and the scopal semantics of indefinites. *Linguistics and Philosophy*, *20*, 399–467.

Yang, C. D. (2002). *Knowledge and learning in natural language*. Oxford University Press.

Yoon, J.-M. (2005). Two historical changes in wh-constructions in korean and their implications. *Studies in Generative Grammar*, *15*, 457–487.

Yoon, S. (2008). From non-specificity to polarity: A compositional account of *even* and n-words [University of Washington]. *Proceedings of the Northwest Linguistics Conference 24*.

Yun, J. (2011). On the meaning of *wh-(N)-ina* and *wh-(N)-itun* in korean. *Language Research*, *47*, 191–218.

Yun, J. (2013). *Wh-indefinites: Meaning and prosody* [Doctoral dissertation, Cornell University].

Yun, J. (2022). Wh-indefinites. In S. Cho & J. Whitman (Eds.), *The cambridge handbook of korean language and linguistics* (pp. 581–606). Cambridge University Press.

Zwarts, F. (1995). Nonveridical contexts. *Linguistic Analysis*, *25*(3–4), 286–312.

Appendix A

Annotation procedure: decision tree, diagnostics, and tie-breaking

This appendix explains, step by step, how functional labels were assigned to the indefinite expressions analyzed in this dissertation. The core protocol was developed for the person-referring series that are central to the study (*nwukwu-*, *amo-/amwu-*, and their particle combinations), but the same logic was applied, mutatis mutandis, to non-human forms where they were included in the analysis.

A *token* is one concrete occurrence of an indefinite expression in the corpus. Each token receives:

- one main functional label from the semantic-map inventory, and
- where relevant, one broader macro-environment label used in the quantitative summaries.

The appendix is organized as follows. Section A.1 gives the overall logic of the annotation scheme and defines the label inventory. Section A.2 states the full workflow in the order in which the annotator applies the tests. Figure A.1 then presents a simplified decision tree. Sections A.3–A.10 spell out the tests in detail, with Korean diagnostics and examples. Section A.11 gives tie-breaking rules for borderline cases. Section A.12 defines the macro-environments used in the main chapters, and Section A.13 summarizes the procedure.

A.1 Overview of the annotation scheme

A.1.1 Core label inventory

The main functional labels used in the dissertation are listed in Table A.1. These are the labels that appear in the semantic-map analyses in Chapters 4–6.

Table A.1 Core functional labels used in the annotation procedure.

Label	Interpretation
SK	specific known: a particular referent, identifiable to the speaker
SU	specific unknown: a particular referent, but not identified to the speaker/addressee
IR	irrealis existential: non-assertive existential in a non-question environment
Q	polar-question existential: existential use in a yes/no question
CA	conditional antecedent universal
CO	comparative universal
DN	direct negation: clausemate sentential negation
AA	anti-additive lexical negative environment
AM	negative implicative / non-clausemate strong negative environment
FC	free choice
GEN	generic universal
UFC	restricted-domain or universal free choice
IND	indiscriminative 'just any X' reading

These labels are assigned one per token. The goal is not to encode every possible nuance of interpretation, but to place each token in a reproducible region of the semantic map.

A.1.2 Auxiliary labels

A small number of auxiliary labels are used where the main inventory does not quite capture the observed function:

- Q_WH: wh-question use of a bare wh-root ('who?', 'what?'). This is treated as a macro-level interrogative category in the main chapters.

- Q_POL: polite or yes/no question use of an already indefinite form. In practice, this is a surface refinement of the broader question region.
- NO_MATTER: clausal 'no matter who/what' constructions. These are kept in the raw annotations but excluded from some of the main quantitative summaries.
- EXPRESSION: fixed idioms or lexicalized expressions in which the form does not transparently realize an indefinite function.

These auxiliary labels are important for data management, but the main semantic-map arguments in the dissertation rely on the core inventory in Table A.1.

A.2 Overall workflow

The annotation of a token proceeds in the following order:

1. Identify the token and its local context: main predicate, clause type, presence and type of negation, conditionals, modals, comparatives, and any explicit domain restrictor.
2. Run Test (a): **specific vs. non-specific**.
3. If specific, run Test (b): **known vs. unknown specific**.
4. If non-specific, run Test (c): **existential vs. universal force**.
5. If non-specific existential, run Test (d): **polar question vs. irrealis existential**.
6. If non-specific universal, run Test (e): **strong negative vs. non-negative universal domain**.
7. If strong negative, run Test (f): **DN vs. AA vs. AM**.

8. If non-negative universal, run Test (g): **GEN vs. UFC**.

9. If the token is not best analyzed as GEN/UFC, run Test (h): **CA vs. CO vs. FC vs. IND**.

10. If more than one label still seems plausible, apply the tie-breaking rules in Section A.11.

The procedure is designed to minimize ad hoc judgement. Each step asks a local, operational question about the token and its context, and the final label is the result of those local decisions rather than a global impression.

A.3 Test (a): specific vs. non-specific reference

Purpose. Test (a) separates *specific* uses from *non-specific* uses. A specific indefinite is used when the discourse can treat the token as introducing one particular individual that later reference can pick up. A non-specific indefinite does not introduce such an anchored individual; it ranges over possibilities ('anyone', 'no one', 'whoever', etc.).

Only clear S^+ cases can receive SK or SU. All other labels are reserved for S^- cases.

Diagnostic. Add a follow-up sentence with a demonstrative NP such as *ku salam* 'that person' or *ku kes* 'that thing'. If this continuation is natural and clearly refers back to the same individual or entity, annotate the token as S^+. If the continuation is infelicitous or strongly marked, annotate it as S^-.

S^+ (specific). Anaphoric continuation with *ku salam* / *ku kes* is natural and understood as referring back to the same individual or entity introduced by the indefinite.

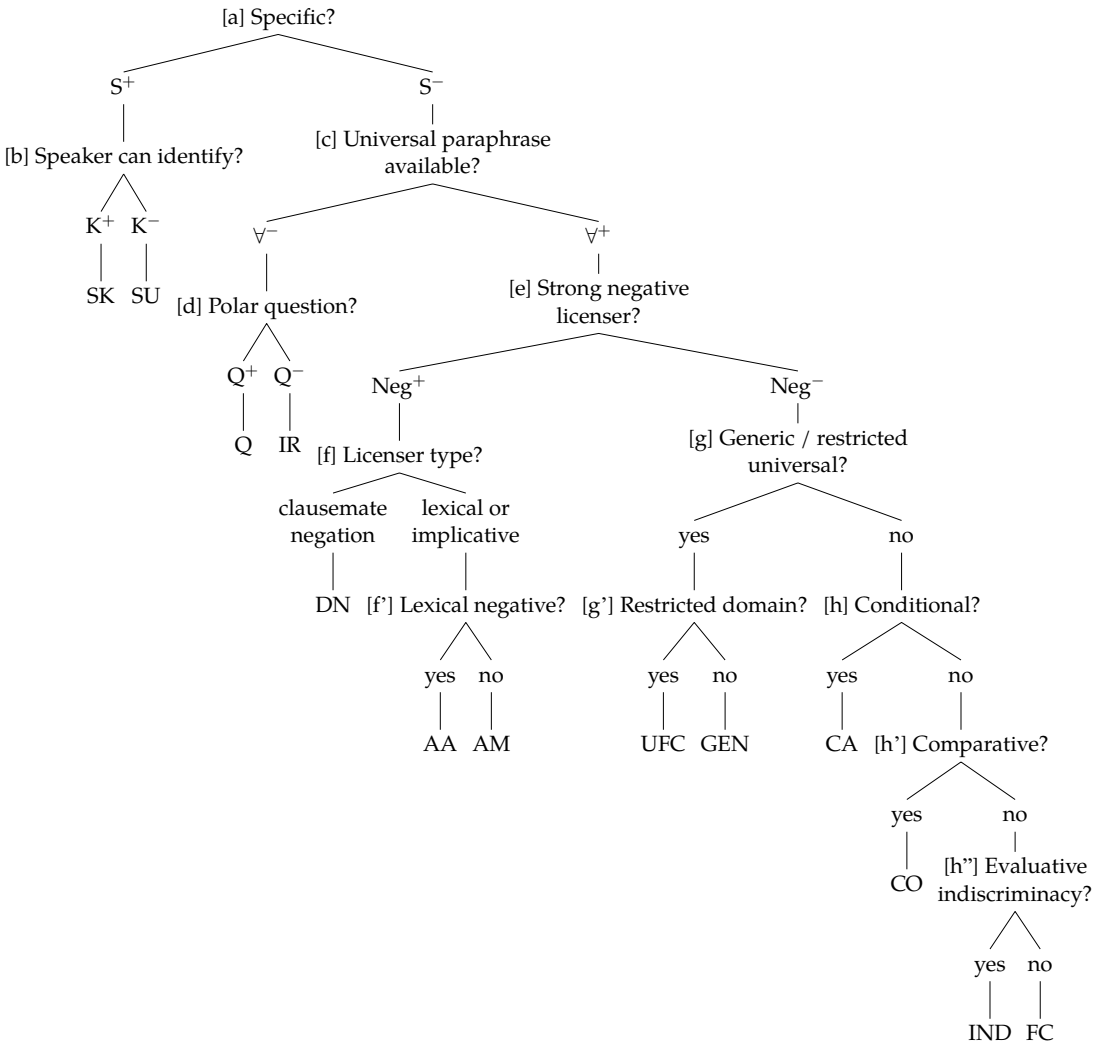


Figure A.1 Simplified decision tree for assigning the main functional label. Auxiliary labels such as Q_WH, Q_POL, NO_MATTER, and EXPRESSION are added by supplementary diagnostics discussed below.

(1) Nwukwu-nka$_i$-ka cenhwa-ha-ess-ta. ku salam$_i$-i
who-NKA$_i$-NOM phone-do-PST-DECL that person$_i$-NOM
sae yaksok-ul wonha-ess-ta.
new appointment-ACC want-PST-DECL
'Somebody$_i$ called. That person$_i$ wanted a new appointment.'
⇒ Specific *nwukwu-nka* allows anaphoric *ku salam*.
[S$^+$]

S$^-$ (non-specific). The demonstrative continuation is infelicitous. This indicates there is no single salient person to pick up.

(2) Nwukwu-na$_i$ i semina-ey tul-e o-l swu iss-ta.
who-NA$_i$ this seminar-LOC enter-CONN come-REL can exist-DECL
ku salam$_i$-un maywu chinchelh-ta.
that person$_i$-TOP very kind-PRS-DECL
'Anyone$_i$ can join this seminar. #That person$_i$ is very kind.' ⇒
Free-choice *nwukwu-na* does not support anaphoric *ku salam*.
[S$^-$]

Borderline cases are treated conservatively. If anaphoric reference is not clearly supported, the token is annotated as S$^-$.

A.4 Test (b): known vs. unknown specific

Purpose. Within S$^+$ tokens, Test (b) distinguishes *known specific* (SK) from *unknown specific* (SU). Both refer to a particular individual, but SK presupposes that the speaker can identify who it is, while SU does not.

Diagnostic. For S$^+$ tokens, add a continuation like 'guess who it was'. If this is natural, annotate SK. If a continuation like 'I don' t know who it was' is natural instead, annotate SU.

K^{+} (**SK**).

(3) Nwukwu-nka-ka akka cenhwa-ha-ess-ta. nwukwu-nci
who-NKA-NOM earlier phone-do-PST-DECL who-COMP
machye pwa.
guess see.IMP
'Someone called earlier. Guess who it was.' ⇒ The speaker knows who the caller is.

[SK]

K^{-} (**SU**).

(4) Nwukwu-nka-ka akka cenhwa-ha-ess-ta. kule-ntey
who-NKA-NOM earlier phone-do-PST-DECL but
na-to nwukwu-nci cenhye molu-keyss-ta.
I-also who-COMP at.all not.know-FUT-DECL
'Someone called earlier, but I don' t know who it was.' ⇒ A particular caller is assumed, but the speaker is ignorant of their identity.

[SU]

SK is reserved for cases with explicit evidence of speaker knowledge (or later identification in context). All other S^{+} tokens are annotated as SU.

A.5 Test (c): existential vs. universal force

Purpose. For S^{-} tokens, Test (c) distinguishes: (i) *existential* readings, where the sentence claims that there exists at least one relevant individual, from (ii) *universal* readings, where the sentence ranges over all relevant individuals in a domain.

Diagnostic. Check whether a faithful paraphrase with a universal expression is available, such as 'for every person x'. If a universal paraphrase is appropriate, annotate $\forall^+$. Otherwise annotate $\forall^-$.

$\forall^-$ (**existential**). IR and Q fall here.

(5) a. Ne-nun tto talun nwukwu-nka-lul manna-pwa-ya
you-TOP again other who-NKA-ACC meet-try-MUST
ha-n-ta.
do-PRS-DECL

'You must meet someone else.' ⇒ No reading 'for every person x: you must meet x'.

[IR ($\forall^-$)]

b. Amwu salam-i-lato ku il-ey tayhay
amwu person-COP-EVEN that matter-LOC about
malha-yess-ni?
tell-PST-Q

'Did anybody tell you anything about it?' ⇒ Asks about the existence of some person, not every person.

[Q ($\forall^-$)]

$\forall^+$ (**universal-like force**). CA, CO, DN, AA, AM, FC, GEN, UFC, and IND fall here.

(6) a. Nwukwu-tunci po-myen kwukcenghi malha-e
who-TUNCI see-if immediately tell-CONN
cwu-e.
give-IMP
'If you see anybody, let me know.' ⇒ Roughly: for every person x, if you see x, tell me.
[CA ($\forall^+$)]

b. Con-un amwu-to po-ci anh-ass-ta.
John-TOP anyone-TO see-CONN not.do-PST-DECL
'John didn' t see anybody.' ⇒ For every person x: John did not see x.
[DN ($\forall^+$)]

c. Nwukwu-lato ku-lul tow-a cwu-e-to
who-LATO him-ACC help-CONN give-CONN-EVEN
toy-n-ta.
may-PRS-DECL
'Anybody may help him.' ⇒ For every relevant person x: helping him is permitted for x.
[FC ($\forall^+$)]

A.6 Test (d): polar question vs. irrealis

Purpose. Within the non-specific existential domain, Test (d) separates yes/no questions (Q) from other non-assertive existential uses (IR).

Diagnostic. If the clause is a yes/no question, annotate Q. Otherwise annotate IR.

A.7 Test (e): strong negative vs. non-negative universal domain

Purpose. Within the universal-like branch, Test (e) separates tokens that are licensed by strong negative environments from tokens that are universal-like in non-negative environments such as conditionals, comparatives, modals, or generic statements.

Operational diagnostic. Annotate **Neg**$^{+}$ if the token is licensed by:

- overt clausemate sentential negation (*an*, *mot*, *-ci anh-*, *-ci mot-*),
- a lexical negative predicate such as *eps-ta* 'not exist' or *molu-ta* 'not know', or
- a negative implicative predicate that licenses the indefinite in an embedded or nominalized complement.

Annotate **Neg**$^{-}$ for universal-like uses in conditionals, comparatives, modals, generic statements, and related non-negative environments.

(7) a. Amwu-to ku-lul ihaey-ha-ci mos-ha-n-ta.
amwu-TO him-ACC understand-do-CONN cannot-do-PRS-DECL
'Nobody understands him.' ⇒ Universal negative under overt sentential negation.
[Neg^{+}]

b. Con-un nwukwu-pota-to khi-ka khu-ta.
John-TOP who-than-even height-NOM big-PRS-DECL
'John is taller than anybody.' ⇒ Universal comparative without overt negation.
[Neg^{-}]

A.8 Test (f): DN vs. AA vs. AM

Purpose. If a token is in a strong negative environment, Test (f) distinguishes DN, AA, and AM. These are not different meanings of the indefinite itself; they are different types of negative licensing environment.

DN (**direct negation**). Use DN when the indefinite is licensed by overt *clause-mate* sentential negation. This includes short-form negation (*an*, *mot*) and long-form negation (*-ci anh-*, *-ci mot-*).

AA (**anti-additive lexical negative environment**). Use AA when the token is licensed by a lexical negative predicate rather than clausemate sentential negation. In this dissertation, the most important AA licensers are *eps-ta* 'not exist, be without' and *molu-ta* 'not know', together with similar lexical negatives where the indefinite is directly in their scope.

AM (**negative implicative environment**). Use AM when the token is licensed by a negative implicative or evaluative predicate outside the clause or nominalization that contains the indefinite. Typical licensors include predicates such as 'dislike', 'avoid', 'refuse', 'give up', 'forbid', or 'be difficult' when the indefinite occurs inside their embedded or nominalized complement.

(8) a. Cinaka-nun amwu-eykey-to mwul-ul phil-yo-ka
pass.by-REL amwu-DAT-TO ask-ACC need-NOM
eps-ta.
not.exist-DECL

'There is no need to ask anyone passing by.' ⇒ Lexical negative predicate *eps-ta* licenses the indefinite.

[AA]

b. Amwu-to manna-ki-ka silt-ta.
amwu-TO meet-NMLZ-NOM dislike-PRS-DECL
'I don' t want to meet anyone.' ⇒ Negative implicative licenser outside the nominalized complement.
[AM]

Note on conditionals and comparatives. Some NPI literature groups conditionals and comparatives with other NPI-friendly environments. In the present annotation, these are not coded as AA/AM/DN. To keep the functional labels aligned with the semantic map, conditionals are coded as CA and comparatives as CO.

A.9 TEST (G): GEN vs. UFC

Purpose. Within the non-negative universal domain, Test (g) separates:

- **GEN**: generic statements about people or entities in general;
- **UFC**: universal quantification over a clearly restricted set.

Diagnostic. If the sentence is best understood as a generalization about people across times and situations ('people in general ⋯'), annotate GEN. If it clearly ranges over a restricted domain ('for every x in this set ⋯'), annotate UFC.

(9) a. Nwukwu-na silswu-nun ha-n-ta.
who-NA mistake-TOP do-PRS-DECL
'Anyone makes mistakes.' ⇒ Generic statement about people in general.
[GEN]

b. Haksayng-i-la-myen nwukwu-na i semina-ey chamye-ha-l
student-COP-IF who-NA this seminar-LOC participate-do-REL
swu iss-ta.
can exist-DECL
'Any student can participate in this seminar.' ⇒ Universal over the restricted domain of students.

[UFC]

A.10 Test (h): CA vs. CO vs. FC vs. IND

Purpose. If a token is universal-like but not best treated as GEN or UFC, Test (h) distinguishes the remaining main types:

- **CA**: conditional antecedent universal,
- **CO**: comparative universal,
- **FC**: modal or permission-like free choice,
- **IND**: evaluative or indiscriminative 'just any X' reading.

Diagnostics. Use the most local structural cue first:

- If the indefinite is in the antecedent of an *if*-clause, annotate CA.
- If it is the standard of a comparative, annotate CO.
- If the reading is modal or permission-like ('anyone may/can ⋯'), annotate FC.
- If the sentence conveys an evaluative or dismissive 'just any X' flavor, annotate IND.

(10) a. Con-un nwukwu-pota-to ppalli ttwi-n-ta.
John-TOP who-than-even fast run-PRS-DECL
'John runs faster than anybody.' ⇒ Comparative universal.
[CO]

b. Amwu-lato tul-e wa-to toy-n-ta.
amwu-LATO enter-CONN come-CONN-EVEN may-PRS-DECL
'Anybody may enter.' ⇒ Permission-like free choice.
[FC]

c. Na-nun amwu-na-hako manna-ko
I-TOP amwu-NA-with meet-date-COMP
siph-ci anh-nun-ta.
want-CONN not.do-PRS-DECL
'I don' t want to date just anybody.' ⇒ Evaluative or indiscriminative 'just any X' reading.
[IND]

A.11 Tie-breaking and unclear cases

A small number of contexts are systematically ambiguous. The following rules were used to keep annotation consistent.

Specificity. If a token could in principle be read either specifically or non-specifically, it is annotated as S^- unless there is *clear* discourse evidence for a recoverable individual.

GEN vs. FC with *swu iss-ta*. Constructions with *swu iss-ta* 'can' are sometimes compatible with both generic ability and permission-like free choice. By default,

such cases are annotated as GEN unless the larger discourse clearly presents a concrete permission or option structure, in which case they are annotated FC.

FC vs. IND. Use IND only when there is a clear evaluative, dismissive, or 'just any X' interpretation. If the sentence simply conveys unrestricted permission or choice, annotate FC.

AA vs. DN. DN is reserved for overt clausemate sentential negation. Lexical negatives such as *eps-ta* and *molu-ta* are coded AA, even if their meaning is strongly negative.

AM vs. AA. Use AM only when the indefinite is licensed inside an embedded or nominalized complement of a negative implicative predicate. If the lexical negative directly licenses the token in the same clause, use AA.

Off-map labels. Tokens that belong to fixed idioms, highly lexicalized expressions, or clausal NO_MATTER constructions are retained in the raw corpus but excluded from the main quantitative analyses unless the chapter explicitly includes them.

Unclear cases. If, after applying all diagnostics, a token still cannot be assigned with confidence, it is labeled UNCLEAR and excluded from the relevant quantitative summary. This was used sparingly.

A.11.1 NPI SUBTYPES

Within the NPI macro-category, I distinguish three types of negative licensing environment. These subtype labels do not name different meanings of the indefinite itself; rather, they record what kind of negative environment licenses the token.

DN (**direct negation**). DN is used when the indefinite is licensed by overt clausemate sentential negation, including short-form negation (*an*, *mot*) and long-form negation (*-ci anh-*, *-ci mot-*).

AA (**anti-additive lexical negative environment**). AA is used when the token is licensed by a lexical negative predicate rather than clausemate sentential negation. In this dissertation, the most important AA licensers are *eps-ta* 'not exist, be without' and *molu-ta* 'not know'.

AM (**negative implicative environment**). AM is used when the token is licensed by a negative implicative or evaluative predicate outside the clause or nominalization that contains the indefinite. Typical licensors include predicates such as 'dislike', 'avoid', 'refuse', 'give up', 'forbid', or 'be difficult' when the indefinite occurs inside their embedded or nominalized complement.

A.12 Macro-environments used in the main chapters

The five macro-environments used in the main quantitative analyses are as follows.

Q_WH. Wh-question uses: the item functions as an interrogative pronoun ('who?', 'what?') rather than as an indefinite.

PPI. Positive existential or speaker-oriented specific uses, corresponding mainly to SK and SU.

Neutral. Non-negative but non-committal existential uses, corresponding mainly to IR, Q, CA, and in some chapter-specific summaries also Q_POL.

NPI. Negative-polarity environments, corresponding to DN, AA, and AM.

Free-choice. Free-choice, generic, comparative, and indiscriminative uses, corresponding to CO, FC, GEN, UFC, and IND. In other words, this macro-environment covers the full FC-side neighbourhood used in the chapter-level summaries, not just FC in the narrow sense.

These macro-environments are analytic groupings used for visualization and regression, not replacements for the finer labels in Table A.1.

A.13 Summary

In summary, each token in the dataset is assigned:

- one core functional label from the semantic-map inventory (SK, SU, IR, Q, CA, CO, DN, AA, AM, FC, GEN, UFC, IND),
- and, where relevant, one broader macro-environment label (Q_WH, PPI, Neutral, NPI, Free-choice).

The procedure is deliberately local and explicit. It begins with specificity, then distinguishes existential from universal force, then separates negative from non-negative universal domains, and finally resolves the finer contrasts inside those domains. The aim is not to force every token into an artificially sharp category, but to provide a consistent annotation protocol that can be applied across the corpora and then used in the diachronic analyses of Chapters 4–6.

In that sense, the appendix provides the bridge between the theoretical framework and the corpus data. It is what makes it possible to track, in a reproducible way, how Korean indefinite forms expand, narrow, compete, and specialize over time.