

Z.AI LEXICOGRAPHY ADVISORY

STRATEGIC ADVISORY · WHITE PAPER

A Chinese–Anishinaabe Government Dictionary

*Lexicographic architecture for the Government
Anishinaabe Programme — integrating the 51,000-entry
corpus, the B1 Coursebook, and the Administrative State
Manual into a treaty-grade trilingual reference work.*

PROGRAMME	Government Anishinaabe
DOCUMENT TYPE	Strategic Advisory, Lexicography
SOURCE CORPUS	corpus_final_100k_enhanced.json v4.0
COMPANION VOLUMES	Book B1 Coursebook; Administrative State Manual
CLASSIFICATION	Internal Advisory — Project Leadership

JULY 2026 · TORONTO



Table of Contents

Executive Summary	3
Part I • Global Best Practices in Lexicography	4
1.1 Corpus-Driven Lexicography	4
1.2 Sense Disambiguation and Polysemy	4
1.3 Entry Structure and Microarchitecture	5
1.4 Bidirectional and Reversible Dictionaries	5
1.5 Polysynthetic and Algonquianist Lexicography	5
Part II • The Government Anishinaabe Problem	7
2.1 Constructed-Diplomatic-Language Lexicography	7
2.2 The Triadic Script Reality	7
2.3 The Administrative State Manual as Semantic Backbone	8
Part III • Audit of the Existing Corpus	10
3.1 Strengths to Preserve	10
3.2 Gaps to Address	10
3.3 Strong Entry vs. Weak Entry: A Side-by-Side	11
Part IV • Recommended Dictionary Architecture	13
4.1 Data Format: Migrate to LIFT	13
4.2 Tri-Directional Indexing	14
4.3 Sense Disambiguation: Frame-Indexed Senses	14
4.4 Morphological Decomposition	15

4.5 Domain Ontology Alignment	15
Part V • Editorial Workflow and Governance	16
5.1 Editorial Roles	16
5.2 Review Pipeline: Five Stages	16
5.3 Governance of Entry Types	17
5.4 Etymological Discipline	18
Part VI • Sovereignty, Community, and Decolonial Lexicography	19
6.1 OCAP Principles and Indigenous Data Sovereignty	19
6.2 The Anti-Colonial Loanword Strategy	19
6.3 Community Consent for Constructions and Neologisms	20
6.4 Intellectual Property and Licensing	20
Part VII • Tooling, Integration, and NLP Readiness	21
7.1 Authoring: FLEx + LIFT	21
7.2 Publication Pipeline	21
7.3 Integration with the Coursebook	21
7.4 NLP and Machine Translation Readiness	22
Part VIII • Risk Register	23
Part IX • Roadmap (90-Day, 1-Year, 3-Year)	25
9.1 90-Day Phase: Foundation	25
9.2 One-Year Phase: Learner Dictionary v1	25
9.3 Three-Year Phase: Full Dictionary	26
Closing • Toward a Treaty-Grade Lexicography	27

Executive Summary

The Government Anishinaabe Programme has reached the point where a dictionary is no longer an optional scholarly artifact but a strategic necessity. The Programme already possesses three substantial assets: a 51,247-entry trilingual corpus with full etymological provenance, an 85-page Level 1 Coursebook teaching the Government Anishinaabe script and twelve units of professional vocabulary, and a 500-page Administrative State Manual that defines the conceptual universe the Programme must express in Anishinaabemowin. The dictionary’s task is to weave these three assets into a single treaty-grade reference work — and to do so under conditions that standard bilingual lexicography was never designed to handle.

This advisory argues that the project should be framed not as a conventional bilingual dictionary but as a **constructed-diplomatic-language lexicography** problem, analogous in important ways to the lexical work done for revived languages (Modern Hebrew, Cornish), planned languages (Bahasa Indonesia, Modern Standard Arabic), and intergovernmental working languages (UN French, EU legal English). The Anishinaabe side is polysynthetic, dialectally diverse, and politically charged; the Chinese side requires pinyin/hanzi coordination and a four-script Anishinaabe representation (Government script, Fiero, IPA, and Hangul bridge); the English side is the existing gloss layer that must be retained for backward compatibility with the corpus and the Coursebook but repositioned as a tertiary index rather than the primary source.

Four headline recommendations follow. First, **migrate the corpus to a proper lexicographic schema** (TEI Lex-0 or LIFT) so that the existing JSON becomes a queryable lexical database rather than a flat list. Second, **introduce Chinese as a third indexed language** with full bidirectional lookup, leveraging the Hangul phonetic bridge that is already present in the corpus. Third, **establish editorial governance** that distinguishes ATTESTED, CONSTRUCTED, LOANWORD, and NEOLOGISM entries with different review thresholds and different community-consent requirements. Fourth, **align the dictionary to the Administrative State Manual’s concept taxonomy** so that every chapter of the Manual has a corresponding lexical subdomain, making the dictionary a true companion to the Programme’s pedagogical and administrative work.

The remainder of this advisory develops these recommendations in nine parts: a survey of global best practice, a diagnosis of the Government Anishinaabe problem, an audit of the existing corpus, the recommended technical architecture, the editorial workflow, the sovereignty and community-consent dimensions, the tooling stack, a risk register, and a phased 90-day / 1-year / 3-year roadmap.

Part I • Global Best Practices in Lexicography

Before turning to the specifics of Government Anishinaabe, it is worth setting out the four main traditions of modern lexicographic practice and the specific contributions each makes to a project of this kind. No single tradition is sufficient on its own; a treaty-grade dictionary for a constructed diplomatic register must draw on all four.

1.1 Corpus-Driven Lexicography

The corpus-driven revolution, associated above all with John Sinclair and the COBUILD project at Birmingham in the 1980s and 1990s, established the principle that dictionary entries should be written from evidence rather than from intuition. A dictionary entry, on this view, is a generalization over a corpus of actual usage: senses are ordered by frequency, examples are drawn from real text, and collocations are recorded as observed rather than as logical possibilities. Patrick Hanks extended this work into the Theory of Norms and Exploitations, distinguishing the normal patterns in which a word occurs from the creative stretches that speakers perform against those patterns.

For the Government Anishinaabe Programme, the corpus-driven principle has a sharp edge. The existing 51,247-entry corpus was largely *constructed* rather than *observed* — only 5,460 entries (10.7%) are ATTESTED, while 43,976 (85.8%) are CONSTRUCTED and 1,748 (3.4%) are LOANWORD. A corpus-driven purist would say that 86% of the corpus is not evidence at all. The Programme’s response must be to treat ATTESTED entries as the empirical backbone and to clearly mark CONSTRUCTED entries with their derivation method, opacity score, and review status — so that users can weight their trust accordingly. This is the model already half-implemented in the corpus’s etym field; it needs to be completed and made central to the dictionary’s editorial policy.

1.2 Sense Disambiguation and Polysemy

Sue Atkins and Michael Rundell’s work on the Oxford Dictionary of English and the Cobuild dictionaries established the modern practice of sense disambiguation: each headword is broken into discrete senses, each sense is given a clear definition written in controlled vocabulary, and senses are ordered by frequency or by logical relatedness. The FrameNet project at Berkeley extended this by anchoring each sense to a semantic frame — a structured representation of the situation the word evokes — with frame elements (participant roles) and typical collocations.

Sense disambiguation is unusually hard for Government Anishinaabe because the source language is polysynthetic. A single Anishinaabe verb form may encode what English would express as a full clause, with subject, object, tense, negation, and modality all fused into one word. The English gloss “discuss/review” (corresponding to *ganawendam*) is not one sense but a family: monitor, supervise, review (a document), deliberate (in council), care for (an obligation). The dictionary must record each of these as a distinct sense with its own frame, examples, and Anishinaabe morphological context — not collapse them into a single gloss. FrameNet-style framing, adapted to the Algonquianist tradition, is the right tool.

1.3 Entry Structure and Microarchitecture

Ladislav Zgust’s classic *Manual of Lexicography* (1971) established the modern architecture of the dictionary entry: headword, pronunciation, part of speech, inflectional paradigm, etymology, sense(s) with definitions and examples, collocations, and cross-references. Robert Lew’s work on dictionary use refined this by showing that users navigate entries in characteristic patterns — skimming the headword and first sense, looking for examples before definitions, and relying on typographic signaling to find what they need quickly. Modern dictionaries are designed to be *scanned* as much as *read*.

For a trilingual dictionary, the entry microarchitecture must serve three audiences at once: an Anishinaabe speaker looking up a Chinese or English equivalent, a Chinese-speaking diplomat looking up an Anishinaabe term, and a scholar checking the etymology. The Government Anishinaabe entry therefore needs an unusually rich header: the headword in Government script, the Fiero double-vowel equivalent, the Hangul phonetic bridge, the IPA pronunciation, the pinyin (for the Chinese equivalent), the part of speech using Algonquianist categories (VAI, VTI, VTA, VII, AI, II), the animacy (animate/inanimate), and the morphological decomposition — all before the first sense is given.

1.4 Bidirectional and Reversible Dictionaries

A bidirectional dictionary serves both Anishinaabe-to-Chinese and Chinese-to-Anishinaabe lookups from the same underlying data. Stefan Heid, Thierry Fontenelle, and others have shown that true bidirectionality is harder than it looks: an Anishinaabe-to-Chinese entry for *nibi* (water) must give the Chinese equivalent 水 with appropriate register notes, while a Chinese-to-Anishinaabe entry for 水 must explain that the Anishinaabe equivalent depends on context — *nibi* for water as substance, *nibiwa* for water as a body, *zaaga’igan* for water as a flow — and that in ceremonial contexts the word may carry obligations that the Chinese term does not. The two directions are not mirror images.

For Government Anishinaabe, the solution is to maintain a single canonical lexical entry for each Anishinaabe headword and to derive multiple *access indexes* from it: by Anishinaabe initial alphabet (Government script order), by Fiero, by Chinese hanzi, by pinyin, by English, and by semantic domain. The underlying entry is one object; the published dictionary presents it through six different lookup lenses. This is technically straightforward with a LIFT or TEI Lex-0 backend.

1.5 Polysynthetic and Algonquianist Lexicography

Anishinaabemowin is a polysynthetic Algonquian language, and its lexicography cannot be modeled on English or French. The foundational work was done by Leonard Bloomfield in the early twentieth century, whose grammar of Eastern Ojibwe established the analytical categories (initial, medial, final; animate/inanimate; the four verb classes VAI, VTI, VTA, VII) that all subsequent work has used. John Nichols’s *Ojibwe-English Dictionary* (1985, with Earl Nyholm) and the online Ojibwe People’s Dictionary extended this into a full lexical resource. Richard Rhodes’s work on Ojibwe syntax and semantics, particularly on the medials and the verb complex, provides the morphological decomposition model that the Programme should adopt.

A key Algonquianist principle is that the *verb* is the primary lexical unit, not the noun. Many concepts that English expresses as nouns are expressed in Anishinaabemowin as verbs: “he is a chief” is *ogimaa* (literally “he is chief”), not “a chief” with a copula. The Government Anishinaabe dictionary must therefore be verb-centric: every verb headword must be cited in its full inflectional context (third-person singular animate intransitive as the citation form), with cross-references to derived nouns and to the initials, medials, and finals that compose it. The Coursebook already teaches this pattern (its vocabulary tables cite VAI, VTI, VTA, VII for every verb); the dictionary must match.

The table below summarizes the four traditions and what each contributes to the Government Anishinaabe dictionary.

Tradition	Key figures	Core principle	Contribution to this project
Corpus-driven	Sinclair, Hanks, COBUILD	Entries from observed usage	Treat ATTESTED entries as empirical backbone; mark CONSTRUCTED with derivation method
Sense disambiguation	Atkins, Rundell, FrameNet	Discrete senses with semantic frames	FrameNet-style frame indexing for administrative concepts
Entry microarchitecture	Zgust, Lew	Scannable entry structure	Trilingual header with Government script + Fiero + Hangul + pinyin + IPA
Algonquianist	Bloomfield, Nichols, Rhodes	Verb-centric, polysynthetic decomposition	Initial-medial-final analysis; VAI/VTI/VTA/VII as primary POS

Table 1. The four lexicographic traditions and their contributions.

Part II · The Government Anishinaabe Problem

The Government Anishinaabe dictionary is not a standard bilingual dictionary. It is a lexical infrastructure project for a deliberately constructed diplomatic register — closer in spirit to the lexical work that supports intergovernmental languages than to a typical learner’s dictionary. Three features define the problem and shape every recommendation that follows.

2.1 Constructed-Diplomatic-Language Lexicography

Government Anishinaabe is a constructed register in two senses. First, it is a *codified subset* of a living language: the Programme has selected from the dialect continuum (largely Southwestern Ojibwe, with Fiero orthography as the reference) a working register suitable for government use, much as Bahasa Indonesia was codified from Malay for administrative use in the 1920s and 1930s. Second, it is *actively expanded* through lexical engineering: the corpus shows 30,283 calques, 11,554 phrase calques with specialized loans, and 1,520 specialized loanwords deliberately sourced from Mandarin, Turkish, Indonesian, and Swahili. This is not documentation of an existing language; it is construction of a working diplomatic language on an Anishinaabe substrate.

The closest modern parallels are Modern Hebrew (revived from liturgical use into a working state language, with deliberate lexical engineering by the Academy of the Hebrew Language), Bahasa Indonesia (codified from Malay and expanded with technical terminology by the Pusat Bahasa), and the lexical work of the EU’s translation services (which maintain IATE, a structured terminology database for EU legal terms across 24 languages). Each of these projects shares three features with Government Anishinaabe: a deliberate codification process, an active terminology development program, and a network of institutional users who both consume and contribute to the lexicon. The dictionary must support all three functions.

2.2 The Triadic Script Reality

Government Anishinaabe operates across three writing systems simultaneously, and the dictionary must handle all three. On the Anishinaabe side there is the **Government Anishinaabe script** — a Hangul-derived featural alphabet with a combining macron for long vowels and a dedicated glottal-stop letter (附) — taught in Book A1 and used throughout Book B1. Alongside it sits the **Fiero double-vowel orthography**, the standard romanized form used in academic Algonquianist literature and in every existing Ojibwe dictionary. The Coursebook correctly presents both: the script is primary, Fiero is given in parentheses for cross-reference.

On the Chinese side, the Programme must handle **hanzi** (simplified characters), **pinyin** (romanization with tone marks), and — crucially — the **Hangul phonetic bridge** that the corpus already provides for every entry. This Hangul bridge is a remarkable design choice: it gives Chinese-speaking learners (whose first writing system is hanzi and whose phonetic training is in pinyin) a transparent phonetic representation of Anishinaabe words using a script (Hangul) that is itself featural, syllable-structured, and easy to learn. The Hangul bridge is, in effect, a third orthography that mediates between the Government Anishinaabe script and pinyin.

The dictionary’s entry header therefore requires a **quadruple representation** for every Anishinaabe headword: Government script, Fiero, IPA, and Hangul. For every Chinese equivalent, it requires: hanzi, pinyin (with tones), and (where useful) the Hangul phonetic bridge in the reverse direction. English glosses are retained as the tertiary index. This is more than most dictionaries handle, but it is not excessive — it is the minimum needed to serve the Programme’s actual user base.

2.3 The Administrative State Manual as Semantic Backbone

The Administrative State Manual is the Programme’s 500-page reference for administrative concepts, organized into 42 chapters across seven parts: Foundations (rule of law, ethics, civil service), Core Functions (organizational design, reform, implementation, regulation, finance, procurement, HR, records, IT), Coordination and Service (intergovernmental relations, service delivery, crisis administration), Process (administrative procedure, standardization, workflow, performance, quality, projects, risk, knowledge), Oversight (audit, inspection, investigation, appeals, transparency), Capacity (assessment, development, reform, technical assistance, comparison), and Craft (toolkit, briefing, meetings, writing, relationships, judgment, career).

This is the semantic universe that the dictionary must cover. Each chapter defines a cluster of concepts that must have Anishinaabe equivalents, and those equivalents must be consistent across the dictionary, the Coursebook, and any future administrative materials the Programme produces. The current corpus has only partial coverage: 734 ADMIN entries, 171 TREATY entries, 167 INFRA entries, 34 STATE STRUCTURE entries, 21 POLITY entries, 21 SOVEREIGNTY entries — roughly 1,400 domain-tagged entries out of 51,247 total (2.8%). The Administrative State Manual has 42 chapters; the corpus has 28 domain tags, and most entries (94.6%) have no domain tag at all. The gap between what the Manual requires and what the corpus provides is the dictionary’s central editorial problem.

The table below maps the Administrative State Manual’s seven parts to the corpus’s existing domain tags and identifies where the dictionary must develop new vocabulary.

Manual Part	Chapters	Existing corpus domain tags	Coverage status	Dictionary priority
I. Foundations	1–7	STATE STRUCTURE, CIVIL SERVICE, CRIMINAL	Thin (75 entries)	High — every manual chapter needs ≥50 headwords
II. Core Functions	8–16	ADMIN, INFRA, INFRASTRUCTURE, BUDGET (1)	Partial (900 entries)	High — procurement, HR, records need expansion
III. Coordination & Service	14–16	COMMUNITY, COMMUNICATION, COMMUNICATIONS	Thin (250 entries)	High — service delivery vocabulary largely absent

Manual Part	Chapters	Existing corpus domain tags	Coverage status	Dictionary priority
IV. Process	17–24	PROCEDURE (5), no workflow/performance tags	Very thin (5 entries)	Critical — process vocabulary needs construction
V. Oversight	25–30	No audit/inspection/investigation tags	Absent	Critical — entire subdomain to be built
VI. Capacity	31–36	KNOWLEDGE (14)	Minimal	Medium — reform and capacity vocabulary
VII. Craft	37–42	No tags	Absent	Medium — craft terms (briefing, meeting, writing)

Table 2. Coverage map: Administrative State Manual vs. corpus domain tags.

The core editorial problem in one sentence: 94.6% of the corpus has no domain tag, and the 5.4% that does covers only a fraction of the Administrative State Manual’s 42 chapters — so the dictionary must simultaneously triage existing entries and construct new vocabulary for entire missing subdomains (oversight, process, craft).

Part III · Audit of the Existing Corpus

The corpus (*corpus_final_100k_enhanced.json*, version 4.0-final-specialized) is a substantial achievement. It contains 51,247 entries with 100% etymological coverage, a five-track generation system, deliberate anti-colonial loanword sourcing, and a Hangul phonetic bridge for every entry. It is, however, a *lexical raw material* rather than a *dictionary*. The audit below sets out strengths and gaps in honest detail, because the dictionary project’s success depends on inheriting the strengths and explicitly addressing the gaps.

3.1 Strengths to Preserve

Scale and coverage ambition. 51,247 entries is a serious corpus by any standard. The Freiburg Ojibwe corpus has roughly 30,000; the online Ojibwe People’s Dictionary has about 12,000; the Comprehensive Ojibwe Dictionary (Nichols & Nyholm) has about 9,000. The Government Anishinaabe corpus is already the largest Ojibwe-lexicon resource in existence, by a factor of two. This scale is the project’s strategic advantage and should be defended.

Etymological transparency. Every entry carries an *etym* field with type (attested, calque, loanword, neologism), source language, method, and (for calques) the components used. This is exactly the editorial discipline that the dictionary will need, and it is already in the data. The 97-entry *calque_override* track — which records cases where a transparent loanword was deliberately replaced by a native calque for political-linguistic reasons — is exemplary: each override records the original loanword, the reason (“opacity < 0.4”), and the replacement. This is the standard the whole corpus should meet.

Anti-colonial loanword strategy. The corpus sources specialized loanwords from Mandarin (442), Indonesian (410), Swahili (355), and Turkish (313) — deliberately avoiding English and French, the two colonial languages most associated with administrative imposition on Indigenous communities. This is a political-linguistic choice of considerable sophistication, and it is consistently applied: the metadata records zero English/French loanwords. The dictionary should preserve this strategy, document it in the front matter, and extend it (for example, by considering Quechua, Navajo, or Māori as additional source languages for specific semantic domains).

Hangul phonetic bridge. Every entry includes *a_hangul*, a Hangul transcription of the Anishinaabe form. This is a clever piece of design: it gives Chinese-speaking learners (whose L1 script is hanzi, whose L2 is often English, and whose phonetic training is pinyin) a transparent featural-phonetic representation that requires minimal additional learning. Hangul is itself a constructed featural alphabet, designed in the 15th century by King Sejong’s institute; using it to bridge to Anishinaabe (whose Government script is also featural) creates a beautifully coherent phonetic infrastructure. The dictionary should treat the Hangul bridge as a first-class lookup index, not as a pedagogical footnote.

3.2 Gaps to Address

Review backlog. 49,555 entries (96.7%) are PENDING review; only 2,743 (5.4%) are CONFIRMED. The CONFIRMED entries are concentrated in the *calque_override* and *track_a_corpus* subsets. The bulk of the

corpus — particularly the 30,283 `track_b_calque` entries and the 11,554 `track_g_phrase_calque_specialized` entries — has never been reviewed by a human Anishinaabe speaker. This is the single largest editorial risk in the project: the dictionary cannot ship with 97% of its entries unreviewed. Triageage is the first priority of the 90-day roadmap.

Part-of-speech shallowness. 41,947 entries (81.9%) are tagged simply as “phrase”, with no further grammatical analysis. Only 1,607 entries (3.1%) are tagged as nouns, 1,077 (2.1%) as verbs, and a scattering across other POS categories. For a polysynthetic language where the verb is the primary lexical unit, this is a critical gap: the dictionary cannot meaningfully present VAI/VTI/UTA/VII distinctions if 82% of entries have no POS at all. POS tagging — ideally automated with a morphological analyzer and then reviewed — is a prerequisite for any usable dictionary.

Domain coverage. As documented in Part II, only 2,788 entries (5.4%) carry domain tags, and those tags cover only 28 of the Administrative State Manual’s 42 chapters. Entire subdomains — audit, inspection, investigation, appeals, transparency, workflow, performance, quality, projects, risk, knowledge management, briefing, meetings, writing — have no vocabulary at all. The dictionary must construct this vocabulary, and the corpus must be retroactively domain-tagged.

Chinese absence. The corpus is currently English-glossed. There are no Chinese headwords, no Chinese equivalents, no pinyin, no hanzi. The Hangul bridge is present but it bridges Anishinaabe to phonetic representation, not to Chinese meaning. To become a Chinese–Anishinaabe dictionary, the corpus must acquire a Chinese semantic layer: for each entry, at least one Chinese equivalent (with hanzi, pinyin, register notes, and context). This is a substantial translation task — on the order of 50,000 entries — and it requires both Sinological expertise (to choose the right Chinese term among near-synonyms) and Anishinaabe expertise (to verify that the Chinese term captures the Anishinaabe sense).

Morphological decomposition absent. Despite the polysynthetic structure, the corpus has no morpheme-level analysis. A word like *gagikino’amaagewin* (which the Coursebook glosses as “application/treaty/agreement”) is stored as a single string, not decomposed into *ga-* (prefix) + *gikino’amaage* (verb stem “to teach/apply”) + *-win* (nominalizer). The dictionary needs this decomposition for three reasons: it lets users recognize related words, it supports morphological search (find all words containing the medial *-aazh-*), and it is essential for any future NLP/MT work. The Rhodes-style initial-medial-final analysis, applied to the 5,460 ATTESTED entries first and then propagated to CONSTRUCTED entries where the derivation is recoverable, should be a core editorial task.

Track_j_final_push passthroughs. The 243 entries in `track_j_final_push` include literal passthroughs: “i” → “I”, “xxi” → “XXI”, “xxii” → “XXII”. These are not Anishinaabe words at all — they are placeholders that should either be properly translated (Roman numerals have conventional Anishinaabe equivalents using ordinal prefixes) or flagged as out-of-scope and removed from the dictionary entirely. The same applies to `track_h_oov_content_words` (90 entries) and `track_i_final_oov` (219 entries), which contain direct English-to-Anishinaabe mappings without full morphological or semantic analysis. These tracks were useful for hitting the 100k coverage target but should be triaged before dictionary publication.

3.3 Strong Entry vs. Weak Entry: A Side-by-Side

The contrast below illustrates what the editorial process must do. Both entries are real entries from the corpus; the difference is editorial care, not data availability.

Field	Strong entry (calque_override)	Weak entry (track_b_calque)
English gloss	administrator	ordinary judiciary
Anishinaabe	izhi-naadamo-ogimaa	siladan-gikendaasowin-gagwe-bizhish
Hangul	(phonetic bridge present)	(phonetic bridge present)
POS	(empty — needs tagging)	phrase
Domain	ADMIN	(empty — needs tagging)
Status	CONSTRUCTED	CONSTRUCTED
Review	CONFIRMED	PENDING
Etymology	native_calque, opacity_override, replaces Turkish <i>ajans</i> , reason recorded	calque, word_by_word_component_composition, “morphological review needed: hyphen-joining is derivation-engine artifact”
Editorial assessment	Ready for dictionary publication	Requires re-analysis before publication

Table 3. Strong vs. weak entries: same corpus, different editorial states.

The strong entry is a model: it has a domain tag, a confirmed review status, a clear etymology that records the political-linguistic reasoning (replacing a transparent Turkish loanword with a native calque), and a morphologically analyzable form. The weak entry has none of these: the etymology explicitly flags that the hyphen-joining is an artifact of the derivation engine (meaning the word may not be morphologically well-formed Anishinaabe), the POS is the generic “phrase”, the domain is empty, and the review status is PENDING. The dictionary’s editorial task is to convert weak entries into strong entries — or to remove them.

Part IV · Recommended Dictionary Architecture

This part sets out the technical architecture for the Government Anishinaabe dictionary. Five layers are addressed: the data format, the indexing model, the sense-disambiguation approach, the morphological decomposition, and the domain-ontology alignment. Together they define what the dictionary *is*, as distinct from what it *contains*.

4.1 Data Format: Migrate to LIFT

The corpus is currently stored as a flat JSON list. This was appropriate for the construction phase, when entries were being generated programmatically and the priority was throughput. It is not appropriate for a dictionary, which must support structured queries (find all VTA verbs in the ADMIN domain with a Mandarin loanword component), incremental editing (a lexicographer working on one entry without affecting others), version control (tracking changes to individual entries over time), and round-trip conversion to multiple output formats (print, web, app, API).

The recommended format is **LIFT** (Lexicon Interchange Format), an XML schema developed by SIL International for fieldwork lexicons. LIFT is supported natively by FLEx (Fieldworks Language Explorer), the standard authoring tool for endangered-language lexicography; it round-trips cleanly to TEI Lex-0 (the Text Encoding Initiative’s dictionary schema, used by academic publishers); and it has well-defined extension points for custom fields (which the Programme will need for the Hangul bridge, the Government script, and the domain taxonomy). LIFT is also the format used by Webonary, the free online dictionary platform, which means a LIFT dictionary can be published to the web with minimal additional engineering.

A sample LIFT entry, showing the full structure needed for a Government Anishinaabe dictionary entry, is below. This is the target format; the migration path from the current JSON is straightforward (a Python script can do the structural conversion in a day, though populating the missing fields is the editorial task of years).

```
izhi-naadamo-ogimaa
izhi-naadamo-ogimaa
이지-나다모-오기마
[itʃi na:tamo ogima:]
```

```
ajans
```

```
Turkish loanword, replaced
```

```
administrator (one who serves a chief)
```

行政官

xíngzhèngguān

An official who carries out administrative duties under the authority of a chief or governing body.

izhi-naadamo-ogimaa gaa-izhichiged

The administrator who decided.

决定的行政官。

行政官

xíngzhèngguān

4.2 Tri-Directional Indexing

A bilingual dictionary has two lookup directions; a trilingual dictionary has six ($A \rightarrow B$, $A \rightarrow C$, $B \rightarrow A$, $B \rightarrow C$, $C \rightarrow A$, $C \rightarrow B$). For Government Anishinaabe, the three languages are Anishinaabe, Chinese, and English, and the dictionary must support all six directions from the same underlying entry. The mechanism is a *reversal index*: each LIFT entry contains one or more *reversal-form* elements for each non-primary language, and the publication pipeline generates a lookup index for each direction.

On the Anishinaabe side, the user can look up a word by Government script, by Fiero, by Hangul, or by IPA — four sub-indexes within the Anishinaabe direction. On the Chinese side, the user can look up by hanzi (with radical/stroke indexes), by pinyin (alphabetical with tone diacritics), or by Hangul phonetic bridge. On the English side, the user looks up alphabetically. The Hangul bridge is particularly important here: a Chinese speaker who has learned the Government Anishinaabe script (which is Hangul-derived) can move seamlessly between the two scripts without going through Latin orthography.

4.3 Sense Disambiguation: Frame-Indexed Senses

Each headword's senses should be anchored to a semantic frame, adapted from FrameNet but extended for administrative concepts. A frame is a structured representation of the situation a word evokes: it has a name (e.g. *GOVERNANCE*, *TREATY_OBLIGATION*), a set of frame elements (participant roles like *governed*, *governor*, *instrument*), and typical collocations. A single Anishinaabe headword may have senses in multiple frames; a single Chinese headword may have senses in multiple frames; the frame index lets the user see exactly which sense is relevant for which administrative context.

For Government Anishinaabe, the Programme should develop a custom frame inventory aligned to the Administrative State Manual's chapter structure. Chapter 10 (Procurement) corresponds to a *PROCUREMENT* frame with elements *procuring_entity*, *supplier*, *contract*, *goods_services*, *evaluation_criteria*. Chapter 12 (Records) corresponds to a *RECORDS_LIFECYCLE* frame with elements *record*, *creator*, *custodian*, *retention_schedule*, *disposition*. The frame inventory is itself a deliverable — roughly 80 to 120 frames should cover the Manual — and it should be published alongside the dictionary as a controlled vocabulary.

4.4 Morphological Decomposition

Every Anishinaabe headword should be decomposed into its initial, medial, and final components, following the Algonquianist tradition. The decomposition serves three purposes: it shows users the morphological family of a word (so they can recognize *gagikinoo'amaagewin* as related to *gikinoo'amaage*), it supports morphological search (find all words with the medial *-aazh-* “to a certain extent”), and it provides the training data for any future morphological analyzer or finite-state transducer.

The decomposition should be stored in the LIFT entry as a structured field, not as a string. For *gagikinoo'amaagewin* the decomposition is: initial *ga-* (prefix marking completed action), root *gikinoo'amaage* (verb stem, VAI, “to teach/apply”), final *-win* (nominalizer, forming abstract nouns). Each component is itself a lexicographic object with its own entry (cross-referenced), so the user can click on *-win* and see all the other nouns formed with the same nominalizer. This is the kind of structural depth that distinguishes a scholarly dictionary from a phrasebook.

4.5 Domain Ontology Alignment

Every entry should be linked to the Administrative State Manual via a stable URI. The Manual’s 42 chapters provide the top-level structure; each chapter’s sections (e.g. Chapter 10.1 “Public Procurement as a Core Function”) provide the second level. The dictionary’s domain field should use a controlled vocabulary that mirrors this structure, so a user looking up procurement vocabulary can ask for “all entries linked to Manual chapter 10” and receive a focused subdictionary. The same linkage supports the reverse: a reader of the Manual who encounters a concept (say, “regulatory capture” in Chapter 8.5) can click through to the Anishinaabe equivalents and see how the concept is rendered, with etymology and example sentences.

The alignment has a third use: it exposes coverage gaps. If the dictionary has no entries linked to Chapter 25 (Audit), the gap is visible and can be addressed by targeted lexical construction. The current corpus has no audit vocabulary at all; the alignment would make this obvious and would direct the editorial team’s construction effort. This is the dictionary’s most important strategic function: it makes the state of the language visible to the Programme’s leadership, who can then direct resources to the gaps that matter most.

Part V · Editorial Workflow and Governance

A dictionary is not just a data structure; it is an editorial process. This part sets out the roles, the review pipeline, the governance of different entry types, and the etymological discipline that the Programme should adopt. The process described here is deliberately modeled on the editorial practices of the Academy of the Hebrew Language and the Pusat Bahasa, both of which face analogous challenges (constructed diplomatic register, active lexical engineering, multiple user communities).

5.1 Editorial Roles

Five distinct roles should be established. None can be combined without risk to editorial quality, though a single person may hold more than one role at different times.

Role	Responsibility	Required expertise	FTE estimate
Senior Lexicographer	Editorial policy, entry structure, final sign-off	Professional lexicographer, ideally with bilingual dictionary experience	1.0
Anishinaabe Language Keeper	Validate forms, senses, examples; community liaison	Native or near-native Anishinaabemowin; community standing	1.0
Sinologist	Chinese equivalents, register notes, pinyin/hanzi	Advanced Chinese; familiarity with administrative Chinese	0.5
Domain Specialist	Administrative concept accuracy; Manual alignment	Public administration, ideally with Indigenous governance experience	0.5
Computational Lexicographer	LIFT infrastructure, morphological analyzer, publication pipeline	NLP/lexicographic engineering; Python, XML, FLEx	1.0

Table 4. Editorial roles and approximate FTE estimates for a 3-year project.

The Language Keeper role is the most consequential and the hardest to fill. The Programme should establish an *Elder’s Council* of three to five Anishinaabe language keepers, compensated at professional rates, who review every CONSTRUCTED and NEOLOGISM entry before it is marked CONFIRMED. This is not only an editorial safeguard; it is a sovereignty requirement (see Part VI). The Council should meet monthly during the active editorial phase, with async review between meetings.

5.2 Review Pipeline: Five Stages

The current corpus has only two review states (PENDING and CONFIRMED). The dictionary needs five, with clear criteria for advancement. The pipeline below applies to every entry; an entry can only advance one stage at a time, and the criteria for each stage are explicit.

Stage	Criteria	Owner	Estimated entries/year
1. PENDING	Entry exists in corpus with etymology; no human review yet	Computational Lexicographer	—
2. TRIAGED	Entry has been examined; POS, domain, and morphological decomposition assigned or flagged for construction	Computational Lexicographer	15,000
3. DRAFTED	Full entry drafted in LIFT: senses, Chinese equivalent, examples, frame, Manual alignment	Senior Lexicographer + Sinologist	8,000
4. REVIEWED	Anishinaabe forms, senses, and examples validated by Language Keeper; Chinese equivalents validated by Sinologist	Language Keeper + Sinologist	5,000
5. CONFIRMED	Elder’s Council has signed off; entry is publishable	Elder’s Council	3,000

Table 5. Five-stage review pipeline with annual throughput estimates.

The throughput estimates assume the FTE staffing in Table 4 and are conservative. At these rates, the full 51,247-entry corpus would reach CONFIRMED status in approximately 17 years — far too slow. The 90-day roadmap (Part IX) therefore prioritizes a 5,000-entry Learner Dictionary drawn from the highest-frequency entries, which can be confirmed in the first year, with the remaining entries confirmed in waves over the subsequent four to six years.

5.3 Governance of Entry Types

The four entry types — ATTESTED, CONSTRUCTED, LOANWORD, NEOLOGISM — have different epistemic statuses and require different governance. The corpus already distinguishes them; the dictionary must make the distinction visible to users and enforce it editorially.

ATTESTED entries (5,460, 10.7% of corpus) are drawn from existing Anishinaabe usage. They require Language Keeper validation of the form and sense, but not Elder’s Council sign-off, unless the sense has been extended for administrative use. The presumption is in favor of inclusion.

CONSTRUCTED entries (43,976, 85.8%) are calques or phrase calques created by the Programme’s derivation engine. They require full review: morphological well-formedness (the corpus itself flags “hyphen-joining is derivation-engine artifact” in many entries, which must be fixed), semantic accuracy (does the calque actually mean what the gloss says it means?), and — for entries in sensitive domains (sovereignty, treaty, ceremony) — Elder’s Council sign-off. The presumption is neutral: each entry must earn its place.

LOANWORD entries (1,748, 3.4%) are specialized loanwords from Mandarin, Turkish, Indonesian, or Swahili. They require phonological adaptation review (does the Anishinaabe form follow the adaptation rules recorded in the etym field?), opacity assessment (the Programme’s threshold is opacity < 0.4 triggers

replacement by a native calque, per the `calque_override` track), and — for high-frequency loanwords — Elder’s Council review. The corpus’s `calque_override` mechanism is the model: every loanword should be subject to potential override, and the override decision should be recorded in the entry’s etymology.

NEOLOGISM entries (63, 0.1%) are completely new coinages — words that have no Anishinaabe source, no calque source, and no loanword source. These are the most politically sensitive: creating a new Anishinaabe word is a cultural act, not just a linguistic one. Every NEOLOGISM requires Elder’s Council approval, with documented reasoning (why was a neologism preferred over a calque or loanword?), and the neologism should be provisional for at least one year before being marked **CONFIRMED**. The current 63 neologisms should be reviewed individually; the bar for new neologisms should be high.

5.4 Etymological Discipline

Every entry must carry a complete etymological record. The `calque_override` track is the standard: type, source language, method, opacity score, replacement history (if applicable), and reason. The dictionary should extend this with two additional fields: *attestation_history* (when was this word first attested in Government Anishinaabe use, and in what source?) and *related_forms* (cross-references to morphologically related entries). Etymology is not optional decoration; it is the user’s primary basis for trusting an entry, and in a constructed register it is the only evidence that the entry is well-formed.

Part VI · Sovereignty, Community, and Decolonial Lexicography

A Government Anishinaabe dictionary is not a neutral scholarly artifact. It is an act of cultural and political construction, and the editorial process must reflect that fact. This part sets out the sovereignty dimensions, the community-consent requirements, and the decolonial lexicographic principles that should govern the project.

6.1 OCAP Principles and Indigenous Data Sovereignty

The First Nations principles of OCAP — Ownership, Control, Access, and Possession — establish that First Nations communities own their cultural and linguistic data, control how it is collected and used, have access to it at all times, and possess it in physical and digital form. The Government Anishinaabe corpus is, in OCAP terms, Anishinaabe cultural data: the words, forms, and meanings it contains are the intellectual property of the Anishinaabe communities from which the language comes. The dictionary project must respect OCAP at every level.

In practical terms, this means: (1) the corpus and dictionary should be hosted on infrastructure that the Programme controls, not on commercial platforms that claim derivative rights; (2) the editorial process should be transparent to the participating communities, with regular reports on what entries have been added, modified, or removed; (3) the dictionary should be freely accessible to Anishinaabe community members, with no paywall or login requirement; (4) the underlying data should be available to communities in usable formats (LIFT, CSV, JSON), not locked in a proprietary tool. The Programme’s existing choice to publish the corpus as open JSON is the right precedent; the dictionary should follow it.

6.2 The Anti-Colonial Loanword Strategy

The corpus’s deliberate sourcing of loanwords from Mandarin, Turkish, Indonesian, and Swahili — and its corresponding avoidance of English and French — is a political-linguistic choice of considerable sophistication. It refuses the default pattern in which Indigenous languages absorb colonial vocabulary for administrative concepts (as happened with Cree, Māori, and Hawaiian, all of which now use English loanwords for many administrative terms). Instead, it draws on languages that have their own histories of decolonial state-building: Turkish (reformed from Ottoman under Atatürk), Indonesian (constructed from Malay as a unifying national language), Swahili (expanded from a regional trade language into a national language of Tanzania and Kenya), and Mandarin (the world’s largest state-language by speakers, with its own modernization history).

The dictionary should preserve this strategy, document it in the front matter, and extend it. Three extensions are worth considering. First, the Programme could add **Quechua** as a source language for vocabulary relating to community and kinship — Quechua has a strong tradition of decolonial linguistic activism in Peru, Bolivia, and Ecuador. Second, **Navajo** (Diné bizaad) could be a source for vocabulary relating to sovereignty and self-determination — Navajo has been used as an administrative language within the Navajo Nation since 1991. Third, **Māori** could be a source for vocabulary relating to treaty relations — Māori has been the working

language of the Waitangi Tribunal since 1985 and has developed a sophisticated treaty lexicon. Each of these extensions would deepen the decolonial character of the project without diluting the Anishinaabe substrate.

6.3 Community Consent for Constructions and Neologisms

Creating a new Anishinaabe word for “procurement” or “judicial review” is a cultural act, not just a linguistic one. The Anishinaabe language is not a passive substrate that can be engineered at will; it is a living cultural inheritance, and the authority to extend it rests with the communities that have preserved it. The Elder’s Council mechanism described in Part V is the operational expression of this principle: no CONSTRUCTED or NEOLOGISM entry should be marked CONFIRMED without explicit community consent.

The consent process should be documented in the entry’s etymology: who approved the entry, when, in what forum (Elder’s Council meeting, community language committee, individual Language Keeper sign-off), and what alternatives were considered. This documentation serves three purposes: it gives users a basis for trust (they can see that the entry is not just a linguistic engineer’s invention), it creates an audit trail for future editorial decisions (a later lexicographer can see why a particular construction was chosen), and it creates a record of cultural contribution that can be acknowledged in the dictionary’s front matter.

The Coursebook’s treatment of ceremonial language is instructive here. The Coursebook notes that certain words — particularly *manidoo* (sacred, spirit) and *semaa* (tobacco) — are sometimes written without the macron even when their vowels are long, “as a mark of ceremonial respect”. This is the only orthographic exception in the Government Anishinaabe script, and it is taught as a convention, not a rule. The same principle should apply to the dictionary: some constructions, particularly those touching ceremonial or spiritual domains, should carry explicit cultural-use notes that the Elder’s Council has approved.

6.4 Intellectual Property and Licensing

The dictionary should be published under a license that respects both open-access principles and Indigenous data sovereignty. The recommended approach is a hybrid: a **Creative Commons Attribution-ShareAlike 4.0** license for the structural and editorial content (entry structure, definitions, examples, etymological notes), with a **community-use addendum** for the Anishinaabe-language content itself. The addendum should specify that Anishinaabe communities may use the dictionary freely for any non-commercial purpose, that commercial use requires community consent, and that derivative works must carry the same community-use addendum.

This hybrid model has been developed by the Local Contexts project (Mukurtu, Traditional Knowledge Labels) and is increasingly used for Indigenous-language digital resources. It is not a perfect solution — intellectual property law was not designed for collective cultural inheritance — but it is the best available compromise, and it signals clearly that the dictionary is not a conventional commercial publication.

Part VII · Tooling, Integration, and NLP Readiness

The dictionary must integrate with the rest of the Programme’s ecosystem: the Coursebook, the Administrative State Manual, the corpus, and any future NLP applications. This part sets out the tooling stack and the integration points.

7.1 Authoring: FLEx + LIFT

The recommended authoring environment is **SIL Fieldworks (FLEx)**, the standard tool for endangered-language lexicography. FLEx is free, open-source, actively maintained, and supports LIFT natively. It handles the full lexicographic workflow: entry creation, morphological decomposition, sense management, example sentences, cross-references, and reversal indexes. It has a built-in interlinear text tool that can be used to extract new entries from texts (useful for extending ATTESTED coverage). It supports multi-user editing through a synchronization mechanism, which allows distributed editorial teams to work on the same lexicon.

FLEx’s weakness is its UI, which is dated and can be slow for very large lexicons. For the Government Anishinaabe project, FLEx should be the canonical authoring tool, with a custom web-based review interface (built on the LIFT data) for the Elder’s Council and other reviewers who do not need FLEx’s full feature set. The web interface should support: browsing entries by all six indexes (Anishinaabe script, Fiero, Hangul, hanzi, pinyin, English); filtering by domain, status, and entry type; commenting on individual senses; and approving or rejecting entries through the five-stage pipeline.

7.2 Publication Pipeline

The publication pipeline converts the LIFT lexicon into multiple output formats: a print-ready PDF (via this ReportLab pipeline or a LaTeX pipeline), a static website with full-text search (via Webonary or a custom static-site generator), a JSON API for programmatic access, and an e-book format. The pipeline should be deterministic (the same LIFT input always produces the same outputs) and reproducible (any published version can be regenerated from the version-controlled LIFT source).

The website should support all six lookup directions, with autocomplete for each index, faceted filtering (by domain, by entry type, by Manual chapter, by review status), and full-text search across definitions, examples, and etymology. It should be mobile-friendly (Anishinaabe community members may access it primarily from phones) and should work offline (a service worker caching the lexicon for use in areas with poor connectivity). The website should also expose a public API, so that other Programme tools (the Coursebook’s companion app, any future translation memory system) can query the dictionary programmatically.

7.3 Integration with the Coursebook

The Coursebook (Book B1) currently maintains its own vocabulary tables, with each unit introducing 10 to 20 new words. The dictionary should be the single source of truth for these tables: the Coursebook’s vocabulary should be drawn directly from the dictionary, with one-to-one correspondence between Coursebook entries

and dictionary entries. This eliminates the risk of divergence (where the Coursebook teaches one form and the dictionary records another) and lets the Coursebook be regenerated automatically from the dictionary when entries are updated.

The integration should work in both directions. When a new unit is being written, the Coursebook author queries the dictionary for entries in the relevant domain (e.g. “all ADMIN entries at ILR Level 1 difficulty”) and selects the appropriate subset. When the dictionary is updated, the affected Coursebook units are flagged for review. This bidirectional integration requires the dictionary to carry a *pedagogical metadata* field on each entry: ILR level, prerequisite entries (entries that should be learned first), and unit references (which Coursebook units introduce this entry).

7.4 NLP and Machine Translation Readiness

The dictionary should be designed from the start as a resource for computational linguistics, not just for human lookup. Three NLP applications are worth preparing for: a morphological analyzer (which can parse any Anishinaabe word into its initial-medial-final components, even if the word is not in the dictionary), a finite-state transducer (which can generate and analyze inflected forms), and a machine-translation system (which can translate between Anishinaabe, Chinese, and English using the dictionary as the lexical backbone).

The morphological analyzer requires the morphological decomposition field on every entry (Part IV.4), plus a list of all known initials, medials, and finals as separate lexical objects. The finite-state transducer requires a formal grammar of Anishinaabe morphology, which is substantial but well-documented in the Algonquianist literature. The machine-translation system requires aligned sentence-level data, which the Coursebook’s bilingual dialogues already provide (each Core Dialogue is a trilingual aligned text). The Programme should publish all three resources — morphological analyzer, FST grammar, aligned corpus — as open-source, with the dictionary as the lexical hub.

Part VIII • Risk Register

The following ten risks are the most consequential threats to the Government Anishinaabe dictionary project. Each is assessed for likelihood and impact, with mitigation strategies. The register should be reviewed quarterly and updated as the project evolves.

#	Risk	Likelihood	Impact	Mitigation
1	Review bottleneck: 96.7% of corpus is PENDING; at conservative throughput, full review takes 17+ years	High	High	Prioritize 5,000-entry Learner Dictionary in Year 1; recruit additional Language Keepers; consider AI-assisted triage with human sign-off
2	Morphological hallucination: track_b_calque entries flag “hyphen-joining is derivation-engine artifact” meaning many CONSTRUCTED entries may not be well-formed Anishinaabe	High	High	Run morphological analyzer on all CONSTRUCTED entries; flag and re-analyze any entry that fails; require Language Keeper sign-off on morphology before CONFIRMED status
3	Semantic drift: English gloss may not capture the Anishinaabe semantic field; Chinese equivalent may drift further	Medium	High	Frame-indexed senses with explicit definitions in all three languages; example sentences drawn from Coursebook dialogues; Sinologist and Language Keeper both required for REVIEWED status
4	Dialect fragmentation: Government Anishinaabe is largely Southwestern Ojibwe; communities speaking Severn Ojibwe, Oji-Cree, or other dialects may reject the dictionary	Medium	Medium	Document dialect choices in front matter; include dialect variants where they exist; invite dialect communities to contribute; mark dialect-specific entries explicitly
5	Chinese pinyin/hanzi inconsistency: simplified vs. traditional characters, tone sandhi, regional Putonghua variants	Medium	Medium	Adopt simplified characters and standard Putonghua as default; mark traditional variants where they exist; record tone sandhi in IPA, not in pinyin
6	Community consent delays: Elder’s Council may take longer than editorial schedule allows	High	Medium	Build consent lead time into editorial calendar; hold Council meetings monthly; allow async review for non-controversial entries; have backup entries queued for each Council meeting
7	Copyright/IP disputes: Anishinaabe communities may challenge Programme’s right to publish the dictionary at all	Low	High	Hybrid CC-BY-SA + community-use addendum license (Part VI.4); transparent governance; community representation on editorial board; revenue-sharing if any commercial use emerges

#	Risk	Likelihood	Impact	Mitigation
8	Loanword opacity thresholds disputed: the 0.4 opacity threshold for calque_override may be too aggressive or too lax	Medium	Low	Publish the threshold in front matter; document each override decision; allow threshold to be revised by Elder’s Council; record alternative forms in entry
9	Hangul script adoption friction: Anishinaabe community members may resist a Hangul-derived script for their language	Medium	Medium	Always present Government script alongside Fiero; ensure Fiero is fully functional as a fallback; document the script’s design rationale; treat script choice as a community decision, not a Programme imposition
10	Long-term sustainability: a constructed diplomatic register requires sustained institutional support; if the Programme ends, the dictionary may become an orphan	Medium	High	Deposit LIFT source with a trusted archive (e.g. ELAR, PARADISEC); establish an editorial succession plan; build a community of user-contributors; document the editorial process so it can be continued by others

Table 6. Top 10 risks with mitigation strategies.

Part IX · Roadmap (90-Day, 1-Year, 3-Year)

The roadmap below is phased to deliver usable products at each milestone while building toward the full dictionary. The phases are cumulative: each phase builds on the deliverables of the previous one. The staffing estimates assume the FTE allocations in Table 4; the budget ranges assume North American professional rates and may need adjustment for the Programme’s actual cost structure.

9.1 90-Day Phase: Foundation

The first 90 days establish the editorial infrastructure and triage the highest-frequency entries. The deliverable is a working LIFT repository, a five-stage review pipeline in operation, and the first 1,000 entries moved to TRIAGED status. No publication in this phase; the focus is on building the editorial system.

Activity	Deliverable	Weeks	Owner
Migrate corpus JSON to LIFT format	LIFT file with all 51,247 entries; conversion script under version control	1–2	Computational Lexicographer
Stand up FLEx authoring environment	FLEx project configured with custom fields for Government script, Hangul, domain	2–3	Computational Lexicographer
Recruit Elder’s Council (3–5 members)	Council charter, compensation agreements, monthly meeting schedule	1–6	Senior Lexicographer
Draft editorial charter	Five-stage pipeline, governance rules, license, OCAP compliance document	3–6	Senior Lexicographer
Develop domain ontology aligned to Administrative State Manual	42-chapter domain tag set; mapping from corpus tags to Manual chapters	4–8	Domain Specialist
Triage top 1,000 highest-frequency entries	1,000 entries at TRIAGED status with POS, domain, morphology assigned	6–12	Computational Lexicographer + Senior Lexicographer
Build web-based review interface	Browse, filter, comment, approve; deployed to internal server	8–12	Computational Lexicographer

Table 7. 90-day phase activities.

Estimated cost for the 90-day phase: \$80,000 to \$120,000 (CAD), covering 1.0 FTE Computational Lexicographer, 0.5 FTE Senior Lexicographer, Elder honoraria, and infrastructure setup. The result is a working editorial system ready to scale.

9.2 One-Year Phase: Learner Dictionary v1

Year 1 delivers the first published product: a 5,000-entry Learner Dictionary covering the vocabulary introduced in Book B1 and the most frequent administrative terms from the Administrative State Manual. The Learner Dictionary is Anishinaabe-to-Chinese-to-English, with full morphological decomposition and example sentences from the Coursebook dialogues. It is published as a PDF and a static website.

The 5,000 entries are selected by frequency (highest-frequency English glosses first), domain coverage (at least 100 entries per Administrative State Manual chapter), and pedagogical priority (all vocabulary introduced in Book B1 must be included). All 5,000 entries must reach CONFIRMED status, which requires Elder’s Council sign-off on every CONSTRUCTED and NEOLOGISM entry. The Council therefore meets weekly during the second half of Year 1, with a target throughput of 100 entries per meeting.

Year 1 also delivers the first integration with the Coursebook: Book B2 (planned for Year 2) can be written directly against the dictionary, with vocabulary tables auto-generated from LIFT. The publication pipeline (LIFT to PDF, LIFT to website, LIFT to API) is operational by end of Year 1. Estimated cost for Year 1: \$400,000 to \$600,000 (CAD), covering full FTE staffing, Elder honoraria, web infrastructure, and publication costs.

9.3 Three-Year Phase: Full Dictionary

Year 3 delivers the full 51,247-entry dictionary, with all entries at REVIEWED status and at least 20,000 at CONFIRMED. The full dictionary is published in three editions: (1) a **scholarly edition** with full etymology, dialect notes, and bibliographic references, aimed at linguists and Algonquianists; (2) a **learner edition** with simpler entries focused on the most frequent 15,000 words, aimed at Programme students and government users; (3) a **computational edition** as LIFT, JSON, and API, aimed at NLP researchers and MT developers. All three editions are generated from the same LIFT source.

Year 3 also delivers the derivative products: a terminology database (extracted from the domain-tagged entries, aligned to the Administrative State Manual), a thesaurus (organized by semantic frame), a reverse index (all six lookup directions in one volume), and the morphological analyzer (open-source, with documentation). The Programme should also publish an editorial retrospective documenting what worked, what did not, and what the next dictionary project should do differently. Estimated cost for the full 3-year project: \$1.5 to \$2.2 million (CAD), depending on staffing levels and publication scale.

***A note on scope.** The roadmap above is ambitious but not unprecedented. The Dictionary of American Regional English took 50 years and approximately \$10 million (USD); the Oxford English Dictionary took 70 years for its first edition. The Government Anishinaabe dictionary benefits from a much smaller scope (one language pair, one register, one user community) and from modern editorial tooling. A 3-year timeline for the first full edition is realistic; a 5-year timeline for the scholarly edition with full community consent is more realistic still.*

Closing • Toward a Treaty-Grade Lexicography

A treaty-grade lexicography is one in which every entry is a small treaty — an agreement between linguistic universes, between political communities, between the past and the future of a language. The Government Anishinaabe dictionary, done well, will be such a work. Each Anishinaabe headword is an act of cultural continuity; each Chinese equivalent is an act of diplomatic translation; each English gloss is an acknowledgment of the colonial medium that the Programme is working to move beyond. The Hangul bridge that runs through every entry is itself a small treaty: a Korean script, designed by a Korean king in the fifteenth century for a Korean language, repurposed in the twenty-first century to carry an Algonquian language into a Chinese-speaking diplomatic space. No one designed this convergence; it emerged from the Programme's own design choices, and it is the dictionary's job to honor it.

The recommendations in this advisory are technical, but the project is not. It is a project of language reclamation, of diplomatic imagination, of decolonial statecraft. The technical architecture matters because it is the substrate on which the cultural and political work can proceed; without proper editorial governance, the dictionary becomes an inventory of unverified forms; without proper sovereignty framework, it becomes another act of extraction; without proper integration with the Coursebook and the Manual, it becomes a reference work that no one consults. The architecture proposed here is designed to make the cultural and political work possible — nothing more, and nothing less.

The Programme has already done the hardest part: it has built a 51,000-entry corpus with full etymological provenance, it has written a Coursebook that teaches the language, it has defined an Administrative State Manual that defines what the language must say. The dictionary is the work of weaving these three into a single reference that can be consulted, trusted, and — eventually — cited as authority. That work is the work of years, but it is the work that makes the Programme durable. A language without a dictionary is a language that depends on the memory of its speakers; a language with a dictionary is a language that can be inherited by speakers not yet born.

Niigaan-bimoseyaang maamawi. — We go forward together. (Coursebook, Unit 12 closing)