

Tetum Spellchecking Resources for Microsoft Office

1. Language Identification

- **Language name:** Tetum
 - **ISO 639-3 code:** tet
 - **Locale:** tet-TL
 - **Country:** Democratic Republic of Timor-Leste
 - **Status:** Official National Language
 - **Orthographic authority:** Instituto Nacional de Linguística (INL), Timor-Leste
-

2. Scope of the Resource

This resource provides:

- Spellchecking (word-level) only
- No grammar checking
- No style or contextual suggestions

The objective is to ensure correct orthography in digital documents written in Tetum, particularly for educational, governmental, and journalistic use.

3. Dictionary Content

The dictionary consists of:

- A curated word list based on official Tetum (INL) orthography
- One lexical item per line
- All entries normalized to lowercase
- UTF-8 and UTF-16 Unicode compliant versions provided

The dictionary includes:

- Core Tetum vocabulary
 - Standardized loanwords as accepted in official usage
 - Function words, particles, and affixed forms where applicable
-

4. Orthographic Conventions

4.1 Glottal Stop Representation

In Tetum, the glottal stop is a phonemic element and part of the alphabet.

For technical compatibility with Microsoft Office, the glottal stop is represented using the ASCII apostrophe:

- Character: '
- Unicode codepoint: U+0027

This choice is technical, not linguistic, and does not alter the phonemic or orthographic status of the glottal stop in Tetum. It ensures reliable handling by existing spellchecking systems.

5. Encoding and Format

- Microsoft Word custom dictionary format (.dic)
- Encoding: Unicode UTF-16 LittleEndian
- Plain text, one entry per line

An additional UTF-8 word list is provided for interoperability with Hunspell-based systems (e.g. LibreOffice, browsers).

6. Intended Use

The resource is intended for:

- Microsoft Word and other Office applications
- Educational institutions
- Government ministries and agencies
- Journalistic and publishing workflows

It is designed to function as a foundational linguistic resource that can be extended to full language support in the future.

7. Maintenance and Governance

The word list is maintained according to INL orthographic standards.

Future updates may include:

- Vocabulary expansion
 - Standardization refinements
 - Compatibility updates for additional platforms
-

8. Contact and Stewardship

This resource is developed and maintained in coordination with Tetum language stakeholders in Timor-Leste.

Institutional stewardship is intended to be assumed by Instituto Nacional de Linguística at Universidade Nacional Timor Lorosa'e.

(End of document)