v7: Toward a Faster Vietnamese Typing Toolkit

Duc Phan Tri Nguyen[1,3,5⋆], **Khang Hoang Nhat Vo**[1,3,5⋆], **Nghia Hieu Nguyen**[2,4,5]
[1]Faculty of Computer Science and Engineering
[2]Faculty of Information Science and Engineering
[3]Ho Chi Minh City University of Technology
[4]University of Information Technology
[5]Vietnam National University, Ho Chi Minh City, Vietnam
{duc.nguyen, khang.vo872003}@hcmut.edu.vn, nghiangh@uit.edu.vn

Abstract

This paper introduces v7, an AI-powered Vietnamese typing method that replaces traditional modifier-based input with an intelligent predictive system. By optimizing both initial consonants, rhymes, and tones — the core components of Vietnamese phonetics and orthography, v7 drastically reduces keystrokes while maintaining high prediction accuracy. Through real-time responsiveness, auto-spacing, and linguistic pattern recognition, v7 enhances both typing speed and efficiency. As the first predictive-based approach tailored to the Vietnamese language, v7 offers a more natural, intuitive, and effortless typing experience, redefining digital text entry for Vietnamese users.

1 Introduction

Efficient text input methods are crucial for digital communication, particularly in writing system relies heavily on diacritics like Vietnamese where it modifying base characters to represent distinct sounds. This complexity makes Vietnamese typing time-consuming.

Over the past few decades, Telex [Nguyen, 2025a] and VNI [Nguyen, 2025b] have been developed as two dominant conventions. Both follow a *modifier-based* approach, where users first type a base letter (e.g., `o`) and then apply a modifier key (e.g., `7` in VNI to produce *"ơ"*).

Vietnamese is a **monosyllabic** language [Xuan, 1998; Giáp, 2008; Giáp, 2011], where word and morpheme are identical and each morpheme has at most one syllable. It is also a high phonetic-orthography language in which the structure of its syllables always consists of *initial consonants*, *rhymes*, and *tonal markers* [Đoàn Thiện Thuật, 2016]. This suggests that traditional *modifier-based* methods, which treat each diacritics as independent modifications, may not be the most efficient approach.

In this paper, we introduce **v7**, an advanced input method designed to optimize Vietnamese text entry by integrating linguistic patterns and modern computational techniques. By rethinking Vietnamese character input and leveraging NLP-driven optimizations, v7 enhances both speed and usability, offering a more intuitive typing experience. The name v7 is inspired by the fact that typing **as little as** `v7` in this system will predict the word *"Việt"* as in *"Việt Nam"* - Vietnam.

2 Related Works

2.1 Vietnamese Typing Methods

Both VNI and Telex require the same number of keystrokes for typing any Vietnamese word. VNI uses numeric keys for diacritics, making it English-friendly. In contrast, Telex is often considered faster due to its exclusive use of alphabetic keys, eliminating the need for number row to allow more fluid typing experience, particularly beneficial for mobile users. However, Telex can be annoying when typing in English.

2.2 Pinyin Keyboard

Vietnamese and Chinese share structural similarities as monosyllabic languages, where each morpheme corresponds to a syllable that consist of *initial consonant*, *rhyme*, and *tonal marker*, making their phonetic structures highly systematic.

Unlike Vietnamese, which relies on *modifier-based* input methods, Chinese typing has evolved toward *predictive-based*. The most widely used system, **Pinyin input** [Wang and Andrews, 2021], allows users to type as little as the initial consonants of syllables, the system then predict the intended characters through statistical modeling and linguistic patterns. This significantly enhances typing speed despite the vast number of Chinese characters.

Inspired by Pinyin's efficiency, v7 adopts a *predictive-based* approach for Vietnamese typing to minimize diacritical input.

2.3 GPT-2

GPT-2 [Radford *et al.*, 2019] is a transformer-based language model capable of unsupervised multitask learning. Its small variants offer a balance between efficiency and predictive power, making them well-suited for lightweight NLP applications.

For an input method designed to work on various devices, including low resource, a compact yet robust model is essential. GPT-2's ability to efficiently generate contextual predictions with minimal computational overhead makes it a viable foundation for predictive task for v7.

⋆The two authors contributed equally to this paper.

3 Approach

3.1 Influence

Preliminary on Vietnamese phonological structure

In Vietnamese, morphemes and words are identical [Giáp, 2008; Giáp, 2011], each word has at most one syllable. Syllables in Vietnamese follow a well-defined phonetic structure [Đoàn Thiện Thuật, 2016]. Each syllable consists of three components: an *initial consonant* (**C**), a *rhyme* (**R**), and a *tonal marker* (**T**).

- **Initial Consonants**: Vietnamese *initial consonants* include both single consonants and consonant clusters.
- **Rhymes**: A *rhyme* consists of one or more vowels, optionally followed by a final consonant. Vietnamese also includes the following diacritical vowels: *"ă, â, ê, ô, ơ, ư"*.
- **Tones**: While Vietnamese is commonly described as having a six-tone system, early linguistic studies [Maspero, 1912] have suggested that it actually consists of eight tones, with two *checked* tones previously grouped within the six-tone framework. Figure 1 illustrates their distribution across a large dataset.

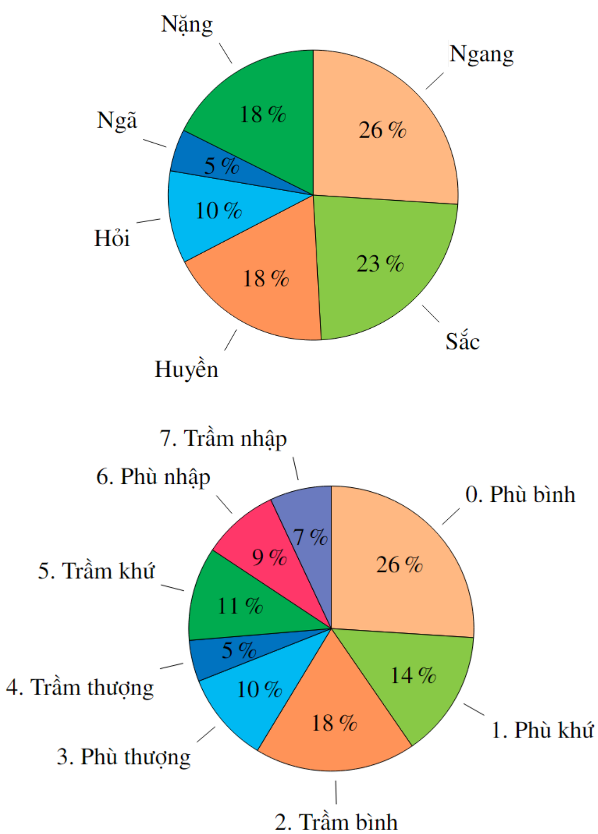

Figure 1: Comparison of tone distributions through a large dataset of 10B tokens (refer to Table 6 for tone categorization).

.

Motivation for Potential Improvements

Given the phonetic and orthographic nature of Vietnamese, there is room for a more optimized typing method beyond traditional *modifier-based* approaches. By leveraging structural patterns in Vietnamese syllables, we can develop a more efficient system that minimizes keystrokes and improve typing efficiency.

- **Initial Consonants**:
  Since the Vietnamese alphabet was developed by Western missionaries, it inherited certain traits of Romance languages, such as using *"gh"* and *"ngh"* for the /ɣ/ and /ŋ/ sounds when preceding *"i"* or *"e"*, and *"gi"* for the /z/ sound. While these conventions improve readability, they introduce redundancy in typing. v7 will simplify this by reducing unnecessary keystrokes while preserving linguistic clarity.

- **Rhymes**:
  Typing Vietnamese traditionally with *modifier-based* methods requires cumbersome inputs for diacritical vowels, especially when combined with diphthongs (e.g., *"ưa"*), triphthongs (e.g., *"iêu"*), and complex rhymes (e.g., *"uyên"*).

  v7 revolutionizes Vietnamese typing by eliminating the need for direct rhyme input **almost entirely**. Instead of typing and then modifying rhymes, v7 employs a *predictive-based* approach, where only the initial consonant (**C**) and tone (**T**) keys provide enough context for accurate rhyme prediction. This drastically reduces keystrokes, making v7 a significant leap forward in efficiency and user experience.

- **Tones**:
  Figure 1 suggest that adopting an 8-tone system improves prediction accuracy compared to the conventional 6-tone system. The additional distinction helps resolve ranking collisions between the two high-frequency tones, *"Sắc"* and *"Nặng"*, which are densely distributed in the 6-tone framework. By utilizing the 8-tone system, v7 enhances predictive accuracy and ensures a more balanced tone distribution.

Typing Rule

v7 inherits features from both VNI and Telex while introducing optimizations for *initial consonants*, *rhymes*, and *tones* according to the above motivations. As shown in Table 1, v7 simplifies the input for initial consonants, reducing the number of keys. Similarly, Table 2 illustrates how tone input in v7 is streamlined, with key mappings for different tone categories, ensuring better typing efficiency and accuracy.

Key(s)	Equivalence	+*"a"*	+*"i"*
`g`	Both *"g"* and *"gh"*	*"ga"*	*"ghi"*
`ng`	Both *"ng"* and *"ngh"*	*"nga"*	*"nghi"*
`z`	*"gi"*	*"gia"*	*"gi"*
`dd`	*"đ"*	*"đa"*	*"đi"*

Table 1: Initial consonant input correspondence in v7 with examples for the rhymes *"a"* and *"i"*.

Key	Equivalence	Example	Tone
`0`	No tone mark	*"tan"*	Phù bình
`1`	Acute	*"tán"*	Phù khứ
`2`	Grave	*"tàn"*	Trầm bình
`3`	Hook	*"tản"*	Phù thượng
`4`	Tilde	*"mãn"*	Trầm thượng
`5`	Underdot	*"mạn"*	Trầm khứ
`6`	*Checked* acute	*"mát"*	Phù nhập
`7`	*Checked* underdot	*"mạt"*	Trầm nhập

Table 2: Tone input correspondence in v7 with examples for the rhyme *"an"*.

3.2 Language Model

Tokenizer. To enable the completeness of word prediction, the vocabulary includes all unique Vietnamese tokens, treating orthographic variants as distinct. It contains $17,788$ tokens plus one padding token. Since training is on a continuous dataset, SOS and EOS tokens are unnecessary, as the model learns to rank the next word based on context.

Architecture. v7 employs v7-GPT, a lightweight, CPU-friendly variant of GPT-2. The model features a context length of 32, with 8 layers, 8 attention heads, and a hidden size of 256.

Dataset. We use the Vietnamese Curated Text Dataset, processed with NVIDIA NeMo Curator for high-quality language data. It includes *C4_Vietnamese*, *Binhvq_News*, *OSCAR_Vietnamese*, and *Wiki_Vietnamese*, covering domains like business, news, health, and technology, ensuring broad generalization[*].

4 Evaluation

4.1 Response Time

Fast response time is essential for real-time typing. v7 achieves rapid predictions, ensuring smooth user experience across devices.

Hardware	Prediction Time (s)
GPU	0.0172
CPU	0.0348

Table 3: Average prediction time per call

With low latency even on CPUs, v7 enables fluid and uninterrupted text input.

4.2 Keystroke Reduction

One of v7's most impactful innovations is its significant reduction in keystrokes required for Vietnamese text input. Since Telex and VNI require the same number of keystrokes per word, we compare v7 directly with VNI.

[*]For more information, please refer to the following link: https://developer.nvidia.com/blog/processing-high-quality-vietnamese-language-data-with-nvidia-nemo-curator

Target	VNI		v7	
	Input	Keys	Input	Keys[†]
"một"	`mot65`	5	`m7`	**3**
"thất vọng"	`that61␣vong5`	12	`th6v5`	**6**
"tưởng tượng"	`tuong73␣tuong75`	15	`t3t5`	**5**
"ý"	`y1`	**2**	`y1`	3
"ba"	`ba`	**2**	`b0`	3

Table 4: Keystroke comparison between VNI and v7

v7 achieves this efficiency through predictive modeling and **automatic spacing**, allowing users to type diacritic-heavy words, phrases, and even full sentences with fewer keystrokes. Though it may be less efficient for simple words like *"ý"* or *"ba"*. Empirical testing on our training dataset shows that v7 reduces keystrokes by 58.36%[‡], allowing users to type with **less than half** the keystrokes required by VNI.

4.3 Top-k Accuracy

v7 eliminates the need for explicit rhyme input by leveraging predictive modeling. While predictions may not always be perfect, the model is optimized to ensure that the intended word is almost always among the top suggestions.

Model	Params	A@1	A@9
v7-GPT	10M	66.54%	95.59%

Table 5: Top-*k* accuracy of v7-GPT in v7

With a **66.54%** top-1 accuracy, most words are correctly predicted at the top. Moreover, **95.59%** of intended words appear within the top-9, ensuring users can quickly find their words in the first UI page[§].

5 Conclusion

v7 revolutionizes Vietnamese text input by replacing the traditional *modifier-based* approach with an intelligent predictive system that optimizes initial consonants, rhymes, and tones. This reduces keystrokes by more than half while maintaining high prediction accuracy. With seamless auto-spacing, real-time responsiveness, and an intuitive design, v7 enables faster and more efficient text entry.

This shift to *predictive-based* input is a necessary evolution, as *modifier-based* methods add unnecessary complexity. By aligning with the phonetic and orthographic nature of Vietnamese, v7 ensures a more natural and effortless typing experience, making it a transformative advancement over traditional input methods.

[†]An additional key is required for prediction selection.

[‡]Assuming the intended word is always the top prediction.

[§]See Appendix A for details on UI design.

A Appendix

Typing Feature Example. v7 eliminates the need for explicit rhyme input, leveraging predictive modeling to infer the intended word efficiently. Below are examples of predictions with the default context *"tôi"*:

- **No rhyme needed** (core functionality): `dd0` → *"đang, đi, đưa, đâu, đây"*
- **Filtering rhyme**: `dda0` → *"đang, đâu, đây, đau, đăng"*
- **Full rhyme input**: `ddang0` → *"đang, đăng, đâng"*
- **Rare rhyme**: `boo0` → *"boong"*
- **Usage of `z` for *"gi"***: `z1` → *"giới, giống, gió, gióng, giớ"*

Although v7 does not require explicit rhyme input, additional keystrokes can refine predictions.

User Interface. When v7 is enabled, an application UI tracks the typed keys. For example, given the input `x0ch2m5ng2`, v7 optimally calls the prediction function at each tone-marking keystroke, ensuring context-aware suggestions.

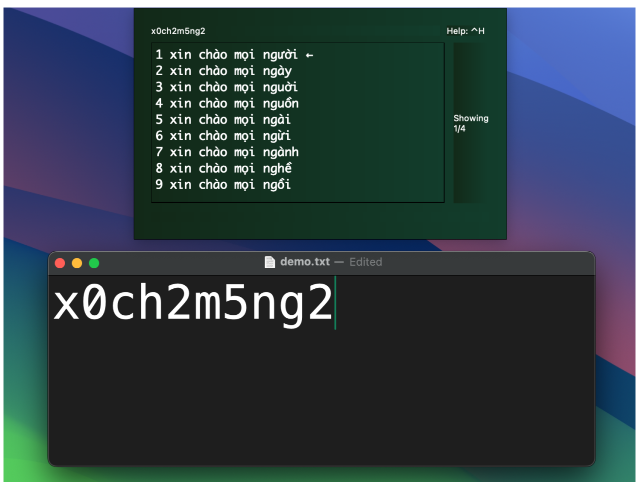

Figure 2: User Interface in MacOS when user is typing using v7

In this case, with `ng2` as the latest input and the prior context *"xin chào mọi"*, the model generates the top predictions, displaying 9 most relevant completions in the first UI page.

Training Configuration. The optimizer used is AdamW with a learning rate of 0.003. The model is trained for 2 epochs with a batch size of 256, using 2 Tesla T4 GPUs.

Vietnamese Tone Category. Vietnamese tones can be classified into two systems: the traditional eight-tone system and the modern six-tone system. These classifications originate from the four-tone classes (四声) of Middle Chinese: *"Bình"* (平 - Level), *"Khứ"* (去 - Departing), *"Thượng"* (上 - Rising), and *"Nhập"* (入 - Entering). Within this framework, *"Phù"* represents high pitch, while *"Trầm"* represents low pitch (阴 and 阳 respectively).

Table 6 illustrates the correspondence between the eight-tone and six-tone systems, along with their English tone mark names and examples.

8-Tone	6-Tone	Tone Mark	Example
"Phù bình"	Ngang	Level (None)	*"an"*
"Trầm bình"	Huyền	Grave	*"àn"*
"Phù khứ"	Sắc	Acute	*"án"*
"Trầm khứ"	Nặng	Underdot	*"ạn"*
"Phù thượng"	Hỏi	Hook	*"ản"*
"Trầm thượng"	Ngã	Tilde	*"ãn"*
"Phù nhập"	Sắc	Acute	*"át"*
"Trầm nhập"	Nặng	Underdot	*"ạt"*

Table 6: Vietnamese Tone Categories in Six-Tone and Eight-Tone Systems

Acknowledgments

This research was supported by The VNUHCM-University of Information Technology's Scientific Research Support Fund.

References

[Giáp, 2008] Nguyễn Thiện Giáp. *Từ vựng học tiếng Việt*. Nhà xuất bản Giáo dục Việt Nam, 2008.

[Giáp, 2011] Nguyễn Thiện Giáp. *Vấn đề "từ" trong tiếng Việt*. Nhà xuất bản Giáo dục Việt Nam, 2011.

[Maspero, 1912] Henri Maspero. Étude sur la phonétique historique de la langue annamite: les initiales [studies in annamese historical phonetics: initial consonants]. *Bulletin de l'École Française d'Extrême-Orient*, 12:1–126, 1912.

[Nguyen, 2025a] Thi Nguyen. Typing vietnamese, part 1: Language, identity and technology at a crossroad. *Saigoneer*, 2025. Illustration by Hannah Hoang.

[Nguyen, 2025b] Thi Nguyen. Typing vietnamese, part 2: The vietnamese diaspora, unicode and the ubiquity of unikey. *Saigoneer*, 2025. Illustration by Hannah Hoang.

[Radford *et al.*, 2019] Alec Radford, Jeff Wu, Rewon Child, David Luan, Dario Amodei, and Ilya Sutskever. Language models are unsupervised multitask learners. 2019.

[Wang and Andrews, 2021] Qiuying Wang and Jean F. Andrews. Chinese pinyin: Overview, history and use in language learning for young deaf and hard of hearing students in china. *American Annals of the Deaf*, 166(4):446–461, 2021.

[Xuan, 1998] Hao Cao Xuan. The problem of phoneme in vietnamese. *Vietnamese studies*, 1998.

[Đoàn Thiện Thuật, 2016] Đoàn Thiện Thuật. *Ngữ âm tiếng Việt*. Nhà xuất bản Đại học Quốc gia Hà Nội, 2016.