

# TỐI ƯU HÓA ỨNG DỤNG CHUYÊN VĂN BẢN TIẾNG VIỆT CÓ CHỨA TỪ VỰNG TIẾNG ANH THÀNH TIẾNG NÓI

Trần Văn Nhuộm<sup>7</sup>, Trần Thị Thùy<sup>8</sup>

**Tóm tắt:** Ngày nay, một số văn bản tiếng Việt thường chứa nhiều từ không chuẩn như chữ viết tắt, chữ số và từ ngữ nước ngoài (facebook, showbiz, internet, smartphone,...). Thông qua các nghiên cứu sự tương đồng về cách phát âm và ngữ âm giữa tiếng Anh và tiếng Việt. Chúng tôi đã nghiên cứu đưa ra một số thuật toán để chuyển từ vựng tiếng Anh thành âm tiết phát âm được bằng tiếng Việt. Từ đó, phát triển thành công cụ tự động phiên âm một từ vựng tiếng Anh bất kỳ thành một từ tiếng Việt. Công cụ này là giải pháp tối ưu hóa các ứng dụng chuyển văn bản tiếng Việt có chứa từ vựng tiếng Anh thành tiếng nói tiếng Việt.

**Từ khóa:** Tối ưu hóa; ứng dụng; văn bản tiếng Việt; từ điển CMU; IPA; t2p.

**Abstract:** Currently, some Vietnamese texts consist of many non-standard words, such as abbreviations, numbers and foreign lexical items (e.g., facebook, showbiz, internet, smartphone, etc). By reviewing the research literature on similarities between Vietnamese and English pronunciation and phonetics, we have come up with several algorithm to convert English words into syllables can be pronounced in Vietnamese. By which, a tool is developed to automatically transcribe any English word into Vietnamese equivalent. Such a tool helps to optimizing an application for converting Vietnamese texts containing English vocabulary into speech.

**Keyword:** Optimizing; application; Vietnamese texts; CMU Dictionary; IPA; t2p.

## 1. Đặt vấn đề

Chuẩn hóa văn bản là một quá trình quyết định xem làm thế nào có thể đọc được những từ không chuẩn chẳng hạn như Nato, Facebook,... Quá trình này sẽ quyết định chất lượng của một hệ thống tổng hợp tiếng nói. Tuy nhiên, nhiều hệ thống tổng hợp tiếng nói chỉ chú trọng vào việc làm thế nào để tạo ra được âm thanh nhân tạo. Như một kết quả tất yếu, họ giả sử rằng những văn bản đầu vào luôn luôn ở dạng có thể phát âm được. Tuy nhiên, những văn bản thực tế thì lại không phải lúc nào cũng bao gồm toàn những từ ở dạng chuẩn có thể phát âm chính xác. Ví dụ, chúng có thể là những số liệu, những chữ viết tắt (như GD viết tắt cho “Giáo dục”), những cấu trúc biểu diễn thời gian (như 12h30), tên nước ngoài và tên địa danh (như New York), những chữ số La Mã,... Ở Việt Nam, đã có nhiều hệ thống tổng hợp tiếng nói đã đạt được nhiều kết quả như SAOMAI, HOASUNG, VOICE OF SOUTHERN, VieTalk,... nhưng hầu hết các hệ thống này

<sup>7</sup> Thạc sĩ Trường Đại học Nam Cần Thơ

<sup>8</sup> Tiến sĩ Trường Đại học Nam Cần Thơ

vẫn chưa dành đủ sự quan tâm cho chuẩn hóa văn bản. Những hệ thống này không thể đọc được những từ không chuẩn, hoặc nếu đọc được thì chỉ đọc bằng một cách rất đơn sơ mà hoàn toàn sai. Vì vậy, trong nghiên cứu này nhằm đưa ra giải pháp góp phần vào việc chuẩn hóa văn bản bằng xây dựng công cụ phiên âm một từ vựng tiếng Anh bất kỳ thành một từ tiếng Việt. Công cụ được tích hợp vào ứng dụng chuyển văn bản thành tiếng nói tiếng Việt có chứa từ vựng tiếng Anh.

## 2. Một số định nghĩa và thuật ngữ

### 2.1. Arpabet

Arpabet là hệ thống mã sao chép âm vị phát triển bởi cơ quan Advanced Research Projects Agency (ARPA) như là một phần của Dự án Thông hiểu tiếng nói (Speech Understanding Project) [3].

### 2.2. International Phonetic Alphabet

International Phonetic Alphabet (IPA) là một sản phẩm của International Phonetic Association (Hiệp hội Ngữ âm quốc tế). Mục đích của IPA là ghi lại và sắp xếp âm trong các ngôn ngữ trên thế giới dựa vào những quy tắc ngữ âm khớp nối (articulatory phonetics principles) [2].

### 2.3. Temp

Temp là hệ thống mã sử dụng để ghi các âm vị tiếng Việt trên máy tính.

### 2.4. t2p

2.4.1. *Định nghĩa*: t2p là một chương trình cài đặt bằng ngôn ngữ Perl dùng để xây dựng những quy tắc biến tự vị thành âm vị dựa trên từ điển phát âm. Nói cách khác, nó xây dựng quy tắc biến chữ cái thành âm để phát âm một từ cho trước dựa trên ví dụ là những từ đã được phát âm trước đó. Ví dụ khi áp dụng t2p cho từ “FACEBOOK” kết quả sau khi chạy chương trình là: “F EY S B UH KD”.

### 2.4.2. Nguyên lý hoạt động

t2p sử dụng từ điển CMU và xây dựng Cây quyết định (Decision Tree) để tạo mô hình cho các từ. Cây quyết định là một cây phân cấp có cấu trúc được dùng để phân lớp các đối tượng dựa vào dãy các luật (series of rules). Về bản chất thì giống như câu lệnh “if then else”. Điều này được thể hiện rõ qua Hình 1.

```

if ($att{'L'} eq 'A') {
  if ($att{'R1'} eq 'A') {
    if ($att{'L1'} eq 'H') {
      if ($att{'R2'} eq 'F') {
        return 'AA'; # unique at depth 4
      }
      if ($att{'R2'} eq 'G') {
        if ($att{'R3'} eq '-') {
          return '_'; # unique at depth 5
        }
        return 'AA';
      }
    }
    if ($att{'R2'} eq 'N') {
      if ($att{'L2'} eq '-') {
        return '_'; # unique at depth 5
      }
      return 'AA';
    }
  }
}

```

Hình 1: Kết quả cây quyết định được xây dựng bởi t2p [7]

### 3. Tổng quan về ngữ âm

#### 3.1. Ngữ âm là gì?

Ngữ âm được hiểu là toàn bộ âm thanh ngôn ngữ và tất cả các quy luật, quy tắc kết hợp âm thanh, giọng điệu ở trong từ, trong câu của ngôn ngữ [1].

+ *Âm thanh ngôn ngữ* là âm thanh do bộ máy cấu âm của con người tạo ra. Nó có nghĩa và đảm nhận chức năng giao tiếp trong cộng đồng.

+ *Kiến trúc ngữ âm* là tổng hợp tất cả các quy luật, quy tắc kết hợp âm thanh, giọng điệu ở trong từ, trong câu của ngôn ngữ.

#### 3.2. Đơn vị kết cấu ngữ âm

+ *Âm tiết* là chuỗi lời nói con người dùng để giao tiếp có thể chia tách thành những khúc đoạn từ lớn đến nhỏ, khúc đoạn nhỏ nhất cuối cùng không còn có thể phân chia, ta gọi là âm tiết. Tuy nhiên, về mặt thính giác, dựa vào kinh nghiệm sử dụng ngôn ngữ trong hoạt động giao tiếp, khi nghe một âm tiết ta có cảm giác là có thể tách thành những yếu tố nhỏ hơn.

+ *Âm tố và âm vị*

Âm tố là đơn vị phát âm tự nhiên nhỏ nhất của lời nói. Âm tố được phân làm hai loại chính là âm tố nguyên âm và âm tố phụ âm. Gọi tắt là nguyên âm và phụ âm.

Âm vị là đơn vị nhỏ nhất của ngôn ngữ dùng để cấu tạo hình vị và phân biệt nghĩa các hình vị.

Âm tố là sự thể hiện của các âm vị trong lời nói.

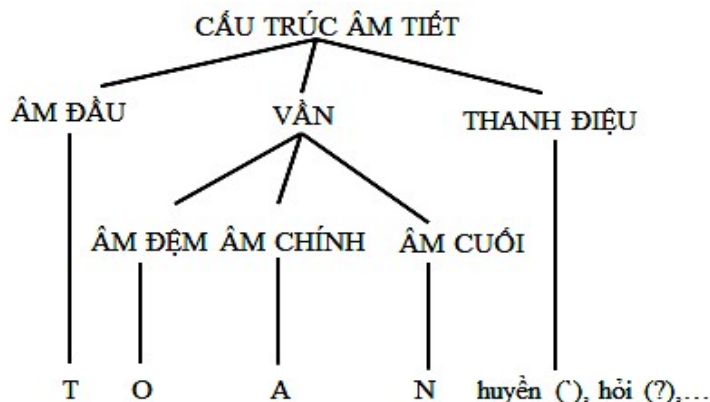
+ *Thanh điệu, trọng âm và ngữ điệu*

*Thanh điệu* là khái niệm dùng để chỉ cao độ của âm tiết. Cao độ này có được là do sự rung bật của dây thanh. Tùy thuộc vào sự rung động đó nhanh hay chậm, mạnh hay yếu, biến chuyển ra sao,... mà ta có các thanh điệu khác nhau. Thanh điệu là một âm vị siêu đoạn tính. Nó được biểu hiện trong toàn âm tiết, hay đúng hơn là toàn bộ phần thanh tính của âm tiết (bao gồm cả âm đầu, âm đệm, âm chính và âm cuối).

*Trọng âm* là hiện tượng phát âm nhấn mạnh vào một yếu tố ngữ âm nào đó trong chuỗi lời nói làm cho nó nổi bật lên. Sự nhấn mạnh đó thường được thể hiện bằng các cách như tăng độ mạnh phát âm, tăng độ dài phát âm, lên xuống giọng.

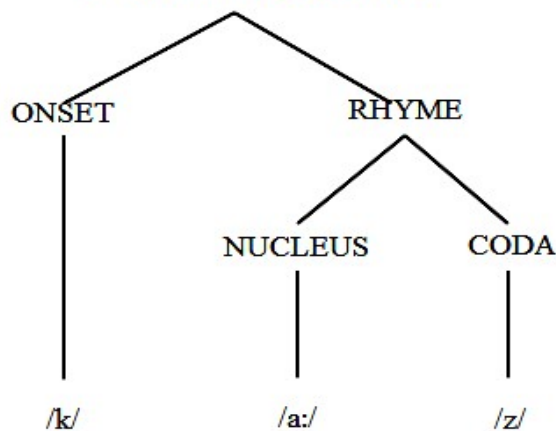
*Ngữ điệu* là sự thay đổi âm điệu của toàn bộ câu nói hay trong hoạt động giao tiếp, dòng ngữ lưu của người phát ngôn truyền đi không phải đều đều mà ngược lại âm điệu của câu nói do người nói phát ra lúc mạnh, lúc yếu, lúc nhanh, lúc chậm, lúc lên, lúc xuống, có lúc liên tục, có lúc ngắt quãng,... sự thay đổi này gọi là ngữ điệu.

### 3.3. Cấu trúc âm tiết tiếng Việt, tiếng Anh



**Hình 2:** Cấu trúc âm tiết tiếng Việt

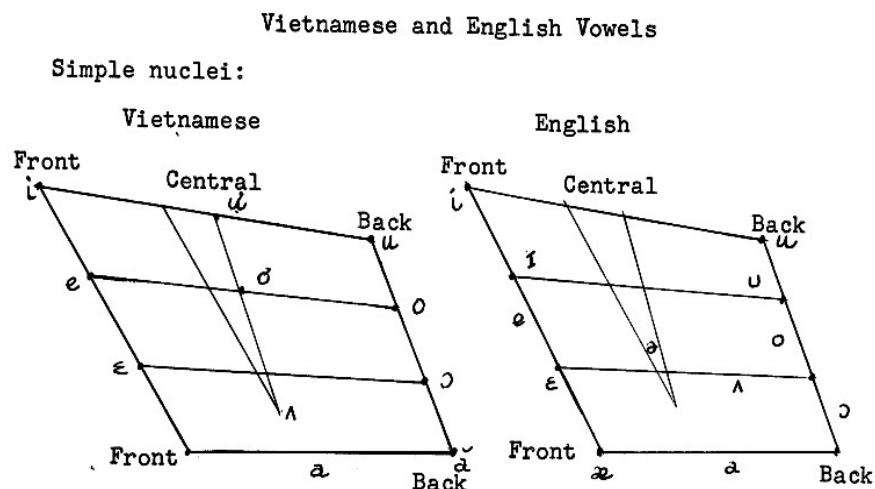
### SYLLABLE STRUCTURE



**Hình 3:** Cấu trúc âm tiết tiếng Anh

### 4. Bảng ánh xạ âm vị tiếng Anh, âm vị tiếng Việt

Dựa trên cơ sở sự tương đồng về phát âm, ngữ âm trong tiếng Anh và tiếng Việt xây dựng các bảng âm vị tương ứng như sau:



Hình 4: So sánh giữa âm vị trong tiếng Việt và tiếng Anh [6]

<b>Monophthongs</b>				
<b>Vowels</b>	<b>Arpabet</b>	<b>IPA</b>	<b>Temp</b>	<b>Word examples</b>
	AO	ɔ	O	off (AO F); fall (F AO L); frost (F RAO S T)
	AA	ɑ	a	father (F AA DH ER), cot (K AA T)
	IY	i	i	bee (BIY); she (SHIY)
	...	...	...	.....

Hình 5: Âm vị tiếng Anh ánh xạ sang Temp

Temp	IPA	Orthography	Examples
<b>Vowels</b>			
u	u	u	tu
W	ʊ	ư	tư
o	o	ô	tô
oU	ɤ	ơ	tơ
O	ɔ	o, oo	to, xoong
Os	ɔ̃	o	vòng, tóc
Es	ɛ̃	a	xanh, sách
a	a	a	ba
---	---	---	----

**Hình 6:** Temp ánh xạ sang âm vị tiếng Việt

### 5. Thuật toán tổng quát

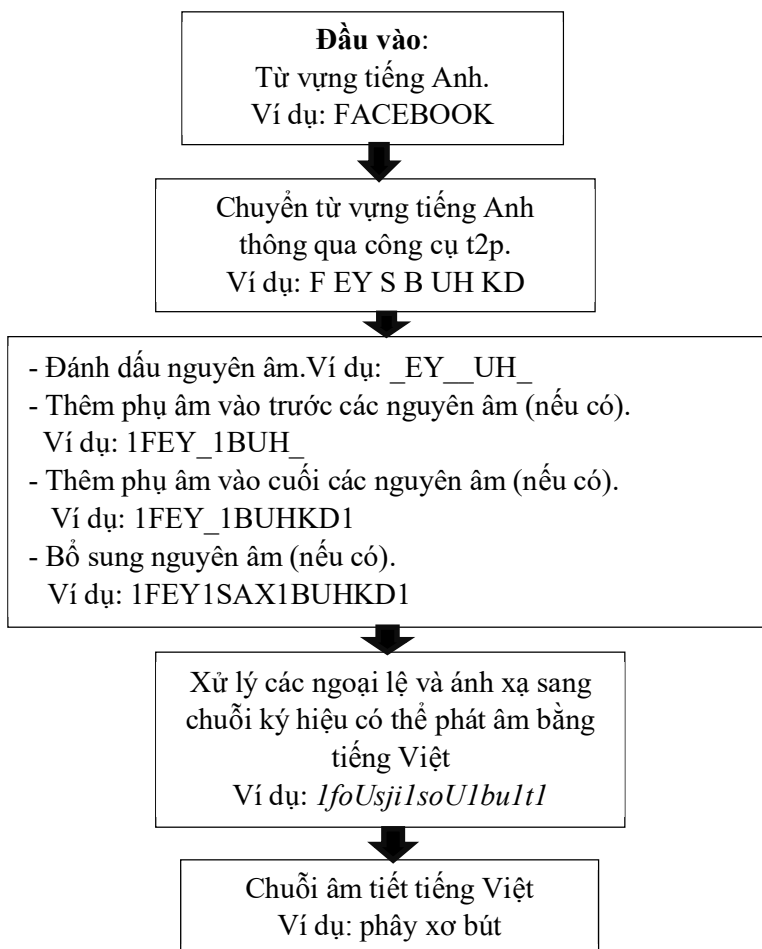
**Bước 1:** Từ vựng chưa chuẩn hóa. Ví dụ là “FACEBOOK”.

**Bước 2:** Sử dụng công cụ t2p chuyển một từ vựng tiếng Anh thành chuỗi âm vị, ví dụ chuyển từ vựng “FACEBOOK” thành chuỗi âm vị “F EY S B UH KD”.

**Bước 3:** Trên cơ sở lý thuyết đã phân tích, vị trí nguyên âm đóng vai trò chính trong một âm tiết nên cần xác định vị trí nguyên âm đầu tiên, tiếp theo thêm phụ âm đầu, cuối và bổ sung nguyên âm nếu cần để hoàn thiện các âm tiết tương đối hoàn chỉnh. Vì vậy, từ chuỗi âm vị như ví dụ trên sẽ cho kết quả là chuỗi âm vị “1FEY1SAX1BUHKD1”.

**Bước 4:** Từ chuỗi âm vị tiếng Anh ban đầu sau khi đã qua các bước xử lý để được chuỗi âm vị mới có thể ánh xạ thành một chuỗi ký hiệu có thể phát âm tiếng Việt, cụ thể như ví dụ trên là chuỗi “1FEY1SAX1BUHKD1” khi ánh xạ chuỗi ký hiệu và xử lý các ngoại lệ để đạt kết quả là chuỗi “1foUsjilsoU1bul1”.

**Bước 5:** Ánh xạ chuỗi ký hiệu có thể phát âm tiếng Việt thành chuỗi âm tiết tiếng Việt. Ví dụ: “1foUsjilsoU1bul1” → “phây xơ bút”.



### 6. Kết quả chương trình

Từ một từ vựng trong văn bản tiếng Việt chưa được chuẩn hóa, ví dụ từ “facebook” chương trình tự động phiên dịch thành chuỗi âm tiết tiếng Việt “phây xơ bút”. Kết quả được minh họa bằng chương trình sau:



**Hình 7:** Kết quả chương trình phiên âm

**Kết luận:** Khi bắt đầu nghiên cứu để tìm ra thuật toán tổng quát giải quyết bài toán tối ưu hóa ứng dụng chuyển văn bản tiếng Việt có chứa thành tiếng nói đã phát hiện nhiều vấn đề vô cùng khó khăn vì giữa chúng tuy có tương đồng về ngữ âm nhưng bên cạnh đó có không ít sự khác biệt. Vì vậy, kết quả từ các thuật toán tìm được cũng mang tính tương đối nhằm góp chút ít công sức cho lĩnh vực nghiên cứu về ngôn ngữ tự nhiên nói chung và chuẩn hóa văn bản nói riêng.

## TÀI LIỆU THAM KHẢO

### Tiếng Việt

- [1]. Võ Xuân Hào (2009). *Giáo trình Ngữ âm tiếng Việt hiện đại*. Trường Đại học Quy Nhơn, trang 1 - 58.

### Tiếng Anh

- [2]. John Kominek, “*TTS From Zero Building Synthetic Voices for New Languages*”, CMU-LTI-09-006 trang 16-18.
- [3]. The CMU Pronouncing Dictionary. <http://www.speech.cs.cmu.edu/cgi-bin/cmudict>.
- [4]. Luis Pedro Hurtarte Caceres, “*Translation of Poetry’s Structures to Architecture*”, trang 26-28, luận văn thạc sĩ.
- [5]. Hoang Gia Ngo, Nancy F. Chen, Sunil Sivadas, Bin Ma, Haizhou Li, “*A Minimal-Resource Transliteration Framework for Vietnamese*”, Proc. Interspeech, 2014.
- [6]. Hoang Thi Quynh Hoa, “*A Phonological Contrastive Study of Vietnamese and English*”, A thesis in English, submitted to the Graduate Faculty of Texas Technological College.
- [7]. Kevin Lenzo, “*t2p: Text-to-phoneme converter Builder*”, retrieved from Carnegie Mellon University: <http://www.cs.cmu.edu/afs/cs.cmu.edu/user/lenzo/html/areas/t2p>,” 1998, December 28.
- [8]. Thang Tat Vu, Mai Chi Luong, Satoshi Nakamura, “*An HMM-based Vietnamese Speech Synthesis System*”, Proc. Oriental COCODA, 2009.
- [9]. The CMU Pronouncing Dictionary, <http://svn.code.sf.net/p/cmuspinx/code/trunk/cmudict/sphinxdict>.