**HỌC VIỆN CÔNG NGHỆ BƯU CHÍNH VIỄN THÔNG**

**KHOA** **CÔNG NGHỆ THÔNG TIN I**

****

**BÁO CÁO BÀI TẬP LỚN**

**MÔN HỌC: CÁC HỆ THỐNG DỰA TRÊN TRI THỨC**

*Đề tài*:

**Xây dựng hệ thống phiên dịch câu Tiếng Việt sang câu Tiếng Anh**

Giảng viên hướng dẫn: Thầy Nguyễn Đình Hóa

Nhóm lớp: D18-017

Nhóm bài tập: 03

Sinh viên thực hiện:

Trần Đức Quang - B18DCCN485

Ninh Thị Ánh Ngọc - B18DCCN456

*Hà Nội, tháng 01 năm 2022*

Mục lục

[I. Kiến thức chuyên môn 3](#_Toc93100791)

[A. Lời mở đầu 3](#_Toc93100792)

[B. Tiếng Việt 3](#_Toc93100793)

[1. Từ loại 3](#_Toc93100794)

[2. Cấu trúc câu Tiếng Việt 5](#_Toc93100795)

[3. Các thì trong câu Tiếng Việt 7](#_Toc93100796)

[C. Tiếng Anh 8](#_Toc93100797)

[1. Từ loại 8](#_Toc93100798)

[2. Cấu trúc câu tiếng Anh 9](#_Toc93100799)

[3. Các thì trong tiếng Anh 12](#_Toc93100800)

[II. Dữ liệu 18](#_Toc93100801)

[A. Tiếng Việt 18](#_Toc93100802)

[B. Tiếng Anh 20](#_Toc93100803)

[1. Tập dữ liệu số 1 20](#_Toc93100804)

[2. Tập dữ liệu số 2 21](#_Toc93100805)

[3. Tập dữ liệu số 3 22](#_Toc93100806)

[4. Tập dữ liệu số 4 23](#_Toc93100807)

[III. Luật 25](#_Toc93100808)

[IV. Kịch bản 29](#_Toc93100809)

[A. Kịch bản số 1 29](#_Toc93100810)

[B. Kịch bản số 2 31](#_Toc93100811)

[C. Kịch bản số 3 33](#_Toc93100812)

[D. Kịch bản số 4 33](#_Toc93100813)

[V. Công cụ lập trình 34](#_Toc93100814)

[VI. Chạy hệ thống 34](#_Toc93100815)

[VII. Tài liệu tham khảo 34](#_Toc93100816)

[A. Tài liệu tiếng Việt 34](#_Toc93100817)

[B. Tài liệu tiếng Anh 34](#_Toc93100818)

[C. Nguồn dữ liệu 34](#_Toc93100819)

[D. Công cụ 35](#_Toc93100820)

# Kiến thức chuyên môn

## Lời mở đầu

Trong thời kỳ hội nhập hiện nay, Tiếng Anh là thứ ngôn ngữ Quốc tế thông dụng của thế giới. Một trong những lý do mà người ta chọn thứ ngôn ngữ này là vì cấu trúc linh hoạt của nó, cũng như sự đơn giản trong việc truyền tải nội dung câu so với các ngôn ngữ khác.

Do đó nhu cầu sử dụng tiếng Anh ngày càng cao trong đó có nhu cầu dịch thuật Tiếng Việt sang Tiếng Anh. Từ nhu cầu đó, chúng em xây dựng một hệ thống dịch câu Tiếng Việt sang Tiếng Anh để phục vụ cho trao đổi thông tin giữa mọi người ở khắp mọi nơi trở nên dễ dàng hơn.

Để có thể dịch một câu Tiếng Việt sang câu Tiếng Anh đạt kết quả truyền tải tốt nhất về mặt nội dung và cấu trúc thì trước hết chúng ta cần hiểu rõ về ngôn ngữ Tiếng Việt và Tiếng Anh.

## Tiếng Việt

### Từ loại

Ngữ pháp Tiếng Việt rất đa dạng từ cấu trúc ngữ pháp đến các chức năng trong câu. Trong đó, từ là đơn vị ngôn ngữ nhỏ nhất dùng để đặt câu. Các từ giống nhau về mặt đặc điểm ngữ pháp và ý nghĩa biểu đạt khái quát gọi là từ loại.

|  |  |  |
| --- | --- | --- |
| STT | Từ loại | Định nghĩa |
| 1 | Danh từ | - Danh từ là những từ chỉ người, vật, khái niệm, hiện tượng, …  - Danh từ được chia thành một số loại như:   * Danh từ riêng là tên riêng của từng người, từng vật, ... viết hoa chữ cái đầu tiên Hà Nội, Ngọc, New York … * Danh từ chung: mèo, vở, chiến tranh, hoà bình, sấm sét… * Danh từ chỉ loại: con, cái, cục, hòn, ... * Danh từ chỉ đơn vị: lít, tấn, tạ, giờ, ...   - Trong câu, danh từ thường giữ chức vụ làm chủ ngữ, khi là vị ngữ danh từ cần có từ *là* đứng trước.  - Danh từ chỉ loại kết hợp với danh từ chung để bổ sung loại cho danh từ chung.  - Danh từ riêng, danh từ chung kết hợp với từ chỉ số lượng ở phía trước, các tính từ, động từ, chỉ từ phía sau để tạo thành cụm danh từ: những bông hoa xinh đẹp, 10 bạn học sinh, ... |
| 2 | Động từ | - Động từ là những từ chỉ các hoạt động, trạng thái sự vật  - Động từ được chia thành một số loại như:   * Động từ tình thái (thường đòi hỏi động từ khác đi kèm): nên, cần, phải, có thể, bị, được, định, … * Động từ chỉ hoạt động, trạng thái (không đòi hỏi động từ khác đi kèm): chạy, nhảy, đi, ….   - Trong câu, động từ thường giữ chức vụ vị ngữ trong câu, khi làm chủ ngữ thì mất khả năng kết hợp với các từ *đã, đang, hãy, đừng*, …  - Kết hợp với tính từ, danh từ, phó từ, từ phủ định để tạo ra các cụm động từ: chạy nhanh, đánh mèo, còn thương, … |
| 3 | Tính từ | - Tính từ là những từ miêu tả đặc điểm hoặc tính chất của sự vật, hoạt động và trạng thái.  - Tính từ được chia thành một số loại như:   * Tính từ chỉ đặc điểm tương đối: nhiều, ít, nặng, vơi, … * Tính từ chỉ đặc điểm tuyệt đối: xanh, đỏ, vàng, xanh ngắt, …   - Trong câu, tính từ giữ chức vụ vị ngữ trong câu, có tác dụng bổ sung ý nghĩa cho danh từ.  - Kết hợp với phó từ chỉ thời gian, động từ, danh từ để tạo ra các cụm động từ, cụm danh từ: chạy chậm, mèo xinh, ... |
| 4 | Số từ | - Số từ là những từ chỉ số lượng và thứ tự của sự vật: một, năm trăm, thứ hai, …  - Số từ chỉ lượng kết hợp với danh từ phía sau, số từ chỉ số thứ tự kết hợp với danh từ đứng trước để có tác dụng bổ sung ý nghĩa cho danh từ: năm tên cẩu tặc, quyển sách thứ hai … |
| 5 | Lượng từ | - Lượng từ là từ chỉ lượng ít hay nhiều của sự vật: mấy chục, các, tất cả, …  - Kết hợp với danh từ ở phía sau để có tác dụng bổ sung ý nghĩa cho danh từ: mấy chục tên cướp, những 10 người, … |
| 6 | Phó từ | - Phó từ là những từ chỉ ý nghĩa thì, cách thức, mức độ hay kết quả của hành động   * Phó từ chỉ quan hệ thời gian: sẽ, từng, đang, hôm qua… * Phó từ chỉ sự phủ định: không, chưa, chẳng, chả, … * Phó từ chỉ sự cầu khiến: hãy, đi, chớ, … * Phó từ chỉ kết quả và hướng: mất, được, ra, lại, … * Phó từ chỉ sự tiếp diễn tương tự: thường xuyên, luôn, … * Phó từ chỉ mức độ: rất, lắm, khá, hơi, …   - Kết hợp với động từ, tính từ để bổ sung cho động từ, tính từ: chưa ăn, đừng đi, rất nhiều … |
| 7 | Đại từ | - Đại từ là những từ dùng để trỏ người, sự vật, hoạt động, tính chất, được nói đến trong một ngữ cảnh nhất định của lời nói hoặc dùng để hỏi.  - Đại từ được chia thành một số loại như:   * Đại từ để trỏ:   + Đại từ xưng hô: anh, nó, tôi, chúng nó, …  + Đại từ trỏ số lượng: bấy, bấy nhiêu, …  + Đại từ trỏ hoạt động, tính chất, sự việc: vậy, thế, …   * Đại từ để hỏi:   + Đại từ hỏi về người, sự vật: ai, gì, …  + Đại từ hỏi về số lượng: bao nhiêu, mấy, …  + Đại từ hỏi về hoạt động, tính chất, sự việc: sao, thế nào, …  - Trong câu, đại từ có thể làm chủ ngữ, vị ngữ hoặc phụ ngữ của danh từ, động từ, tính từ: tôi bơi rất giỏi, chuyện chỉ có vậy thôi, ... |
| 8 | Chỉ từ | - Chỉ từ là từ dùng để trỏ vào sự vật nhằm xác định vị trí của sự vật trong không gian, thời gian: đó, kia, này, nọ, …  - Kết hợp với danh từ ở phía trước: tên kia, cô này, chú ấy, ... |
| 9 | Quan hệ từ | - Quan hệ từ là những từ dùng để biểu thị ý nghĩa quan hệ như so sánh, sở hữu, nhân quả... giữa các bộ phận của câu hoặc giữa các câu trong một đoạn văn:   * Quan hệ từ sở hữu: của * Quan hệ từ so sánh: như, bằng, hơn, … * Quan hệ từ liệt kê (để kết hợp các từ cùng loại): và, hoặc, … * Cặp quan hệ từ: tuy … nhưng …, mặc dù … nhưng …, … * Quan hệ từ định vị: ở, trên, dưới, trái, phải, bên cạnh, …   - Quan hệ từ sở hữu kết hợp với danh từ để thể hiện danh từ/đại từ đứng sau sở hữu danh từ/đại từ phía trước: con mèo của tôi, …  - Quan hệ từ so sánh kết hợp với tính từ/danh từ đứng trước và danh từ phía sau nhằm thể hiện sự so sánh: lớn hơn tôi, tiền bằng tôi, …  - Các quan hệ từ còn lại thể hiện sự ngang bằng về cấu tạo giữa từ đứng trước và sau |
| 10 | Thán từ | - Thán từ là từ dùng để bộc lộ tình cảm, cảm xúc hoặc để gọi đáp của con người: a, ô, hay, dạ, này, …  - Trong câu, thán từ thường tạo thành một câu độc lập hoặc đứng ở đầu: Ôi. Con mèo xinh quá, … |

### Cấu trúc câu Tiếng Việt

Cấu tạo thành phần chính của một câu đơn Tiếng Việt đơn giản:

Chủ ngữ + Vị ngữ

Trong đó:

* Chủ ngữ là thành phần chính của câu nêu tên sự vật, hiện tượng có hoạt động, đặc điểm, trạng thái, … được miêu tả ở vị ngữ và trả lời cho câu hỏi *Ai?, Con gì?, Cái gì?*. Chủ ngữ thường là danh từ, đại từ hoặc cụm danh từ. Trong trường hợp nhất định, động từ hoặc cụm động từ, tính từ hoặc cụm tính từ có thể làm chủ ngữ.
* Vị ngữ là thành phần chính của câu, có khả năng kết hợp với các phó từ chỉ quan hệ thời gian và trả lời cho câu hỏi *Làm gì?, Làm sao?, Như nào?, Là gì?.* Vị ngữ thường là động từ hoặc cụm động từ, tính từ hoặc cụm tính từ, danh từ hoặc cụm danh từ.

Ngoài ra, cấu trúc câu còn có thêm thành phần phụ là trạng ngữ:

* Trạng ngữ được thêm vào câu để xác định thời gian, nơi chốn, nguyên nhân, mục đích, phương tiện, cách thức diễn ra sự việc nêu trong câu. Trạng ngữ còn giúp kết nối các câu, các đoạn, góp phần làm cho đoạn văn, bài văn được mạch lạc.
* Trạng ngữ có thể đứng đầu câu, cuối câu hoặc giữa câu, ngăn cách với thành phần chính của câu bằng dấu phẩy.
* Khi đổi vị trí của trạng ngữ, ý nghĩa của câu có thể bị thay đổi:
* Tôi sẽ học *tới 3 giờ*.
* *Tới 3 giờ*, tôi sẽ học

Vị trí các từ loại trong cấu tạo của một câu Tiếng Việt:

* Chủ ngữ thường thường là danh từ, đại từ, hoặc là một cụm chủ vị, hầu hết đứng ở đầu câu:
* Chủ ngữ là danh từ: *Con mèo* đen
* Chủ ngữ là đại từ: *Tôi* đói
* Chủ ngữ là cụm chủ vị: *Cách mạng thành công* đem lại độc lập cho nước nhà
* Chủ ngữ là mang tính phủ định: *Không ai* được ra khỏi đó
* Chủ ngữ chứa từ ‘có’: *Có* 1 con mèo trên cây
* Vị ngữ thường là động từ, tính từ hoặc cụm động từ, cụm tính từ và một số từ loại khác như đại từ, số từ, danh từ hoặc là một cụm chủ vị, thường đứng sau chủ ngữ:
* Vị ngữ là động từ: Cô ấy *làm việc khuya*
* Vị ngữ là động từ đặc biệt “là”: Cô ấy *là kỹ sư phần mềm*
* Vị ngữ là tính từ: Cô ấy *chăm chỉ*
* Vị ngữ là danh từ: Nhà *5 tầng*
* Vị ngữ là cụm chủ vị: Cán bộ phải *chân đi*, *mắt thấy*, *tai nghe*
* Trạng ngữ có thể đứng ở bất kỳ vị trí nào trong câu:
* Trạng ngữ thời gian: Cô ấy ngồi vậy *năm phút rồi*
* Trạng ngữ nơi chốn: Cô ấy ngồi *ở công viên*
* Trạng ngữ chỉ nguyên nhân: Cô ấy vất vả *vì con mèo*
* Trạng từ chỉ phương tiện: Cô ấy đi học *bằng xe bus*

Bảng 1: Các tri thức về sắp xếp cụm từ, câu trong tiếng Việt

|  |  |  |
| --- | --- | --- |
| STT | Tri thức | Ví dụ minh họa |
| 1 | - Cụm tính từ: phụ trước + trung tâm + phụ sau   * Phụ trước biểu thị quan hệ thời gian mức độ, tính chất, phủ định: phó từ * Trung tâm là 1 tính từ * Phụ sau biểu thị mức độ: phó từ   - Khái quát:  phó từ chỉ quan hệ thời gian/mức độ/phủ định + **tính từ** + phó từ chỉ mức độ | còn ngây thơ lắm  *phó từ tính từ phó từ* |
| 2 | - Cụm danh từ: phụ trước + trung tâm + phụ sau   * Phụ trước bổ sung ý nghĩa về mặt số lượng: số từ, lượng từ * Trung tâm là một danh từ chung hoặc 1 danh từ chỉ đơn vị + 1 danh từ chung * Phụ sau bổ sung ý nghĩa về mặt chất lượng: động từ, tính từ, chỉ từ, quan hệ từ sở hữu   - Khái quát:  số từ/lượng từ + **danh từ chỉ đơn vị + danh từ chung** + động từ + tính từ + chỉ từ + quan hệ từ sở hữu | một con mèo của tôi  *số từ danh từ quan hệ từ sở hữu*  những con mèo này  *lượng từ danh từ chỉ từ*  cái bàn học cũ  *danh từ* *động từ* *tính từ* |
| 3 | - Cụm động từ: phụ trước + trung tâm + phụ sau   * Phụ trước bổ sung ý nghĩa về quan hệ thời gian, phủ định: phó từ * Trung tâm là một động từ hoặc một chuỗi động từ * Phụ sau bổ sung các chi tiết về đối tượng, hướng, địa điểm, thời gian, cách thức, …: tính từ, danh từ, chỉ từ , …   - Khái quát:  phó từ chỉ quan hệ thời gian/mức độ/phủ định + **động từ tình thái** + **động từ chỉ hành động, trạng thái**  + cụm danh từ/tính từ | đã cầm chục trứng  *phó từ động từ* *cụm danh từ*  từng học rất chăm chỉ  *phó từ động từ* *cụm tính từ*  tới đây  *động từ* *chỉ từ* |
| 4 | - Trạng ngữ: quan hệ từ định vị + cụm danh từ | Nó ở trên nhà  *quan hệ từ định vị danh từ* |
| 5 | Cấu trúc câu đơn giản:  - Cụm danh từ/cụm động từ + cụm động từ/cụm tính từ/cụm danh từ  - Trạng ngữ có thể được bổ sung và nằm ở bất kì vị trí nào trong cấu trúc trên  - Thán từ thường nằm ở đầu câu, thường ngăn cách với câu bởi dấu phẩy | Ồ, con mèo ở bên đó xinh  *thán từ danh từ trạng từ tính từ* |

### Các thì trong câu Tiếng Việt

Thì là phạm trù ngữ pháp quan trọng trong Tiếng Anh để xác định thời gian xảy ra và mức độ thường xuyên mà nội dung câu Tiếng Anh truyền tải. Trong Tiếng Việt không có phạm trù thì, nhưng ta dùng phó từ chỉ thời gian như *đã, đang, sẽ* để chỉ thì hoặc căn cứ vào nghĩa trong bài mà biết được thì của câu Tiếng Việt. Việc xác định thì trong câu Tiếng Việt sẽ giúp ngữ pháp, cấu trúc câu Tiếng Anh khi dịch trở nên đúng hơn. Thường ngay sau phó từ chỉ thời gian sẽ là động từ, tính từ.

Các thì cơ bản, bao gồm:

* Thì hiện tại đơn mô tả một hành động thường xuyên diễn ra trong hiện tại hoặc một chân lý, biểu thị bằng từ *thường xuyên, luôn, luôn luôn, …*
* Thì quá khứ đơn mô tả một hành động, tính chất của sự vật, sự việc xảy ra trong quá khứ, biểu thị bằng từ *đã, từng, hôm qua, ngày xưa, ...*
* Thì tương lai đơn một hành động, tính chất của sự vật, sự việc xảy ra sắp diễn ra, biểu thị bằng từ *sẽ, ngày mai, ngày kìa, …*
* Thì hiện tại tiếp diễn mô tả một hành động đang diễn ra, biểu thị bằng từ *đang, hôm nay..*

Các từ biểu thị đứng trước động từ chính hoặc tính từ của câu sẽ mang ý nghĩa thì của câu đó so với thời điểm mốc đứng ở trước đó:

* Tới năm cô ấy 20 tuổi, cô ấy chắc *đã* có em bé

*Thời điểm mốc thì quá khứ*

## Tiếng Anh

### Từ loại

|  |  |  |
| --- | --- | --- |
| STT | Từ loại | Định nghĩa |
| 1 | Danh từ | - Danh từ là từ loại trong tiếng Anh chỉ tên người, đồ vật, sự việc hay địa điểm, nơi chốn.  - Danh từ trong tiếng Anh là Nouns, viết tắt (n).  - Danh từ được chia thành 2 loại: Danh từ riêng và Danh từ chung.  - Ví dụ: Mr. Quang, Miss Ngoc, student, lecturer, Vietnam, Mr. Hoa,... |
| 2 | Động từ | - Động từ là từ loại trong tiếng Anh diễn tả hành động, một tình trạng hay một cảm xúc. Động từ trong tiếng Anh giúp xác định chủ từ đang làm hay chịu đựng điều gì.  - Động từ trong tiếng Anh là verb, viết tắt (v). Ngoài ra, còn có trợ động từ to-be.  - Ví dụ: study, learn, read, buy, walk, sleep, ... |
| 3 | Tính từ | - Tính từ là từ loại trong tiếng Anh chỉ tính chất của sự vật, sự việc, hiện tượng.  - Tính từ trong tiếng Anh là Adjective, viết tắt là (adj).  - Tính từ được chia thành 9 nhánh nhỏ bao gồm:   * Số lượng * Phẩm chất, ý kiến * Kích thước * Tuổi đời * Hình dạng * Màu sắc * Địa chỉ * Chất liệu * Mục đích   - Ví dụ: beautiful, smart, tall, short, fat, thin, red, old, ... |
| 4 | Trạng từ | - Trạng từ là từ loại trong tiếng Anh nêu ra trạng thái hay tình trạng, dùng để bổ sung ý nghĩa cho động từ.  - Trạng từ trong tiếng Anh là Adverb, viết tắt là (adv).  - Trạng từ được chia thành một số loại:   * Trạng từ đứng trước động từ: thời gian, tần suất, mức độ, … * Trạng từ đứng sau động từ: cách thức, địa điểm, khoảng, ...   - Ví dụ: very, usually, quickly, so, now... |
| 5 | Giới từ | - Giới từ là từ loại trong tiếng Anh dùng để diễn tả mối tương quan về hoàn cảnh, thời gian hay vị trí của các sự vật, sự việc được nói đến.  - Giới từ trong tiếng Anh là Prepositions, viết tắt là (Pre).  - Ví dụ: in, on, at, above, behind, for, with, ... |
| 6 | Từ hạn định | - Từ hạn định trong tiếng Anh thường đứng trước một tính từ, một danh từ hoặc một cụm danh từ, mang nghĩa xác định cho tính từ hay danh từ (cụm danh từ đó). Từ hạn định, ở một số trường hợp, có thể được sử dụng như một tính từ.  - Từ hạn định được chia làm 6 loại nhỏ:   * Mạo từ là từ đứng trước danh từ và cho biết danh từ ấy nhắc đến một đối tượng xác định hay không xác định: a, an, the, … * Từ chỉ định (bao gồm đại từ hoặc tính từ) là những từ sử dụng để chỉ người hoặc vật, dựa vào khoảng giữa phi vật lý của người nói đến đối tượng họ đang muốn nhắc đến: this, that, these, those,... * Từ sở hữu (bao gồm đại từ hoặc tính từ) là những từ chỉ việc sở hữu của con người, đối tượng với vật thể khác: mine, her, his, our, … * Lượng từ là từ chỉ số lượng, chúng được đặt trước danh từ để bổ nghĩa về số lượng cho danh từ: all, few, little, some, … * Số từ là từ chỉ số đếm, thường đặt trước danh từ để bổ nghĩa về số lượng cho danh từ: one, two, three, five,... |
| 7 | Đại từ | - Đại từ là những từ được sử dụng thay thế cho danh từ trong câu, nhằm tránh tình trạng lặp lại danh từ.  - Đại từ được chia thành 3 loại:   * Đại từ nhân xưng thường chỉ đồ vật, con người, nhóm người cụ thể, gồm có Đại từ số ít và Đại từ số nhiều: I, he, me, we,... * Đại từ sở hữu chỉ vật thuộc về ai, thường được sử dụng thay thế cho tính từ sở hữu hoặc danh từ đã sử dụng trước đó: his, her, mine, ours, yours,... * Đại từ phản thân xuất hiện khi chủ ngữ và tân ngữ cùng là một đối tượng. Đại từ phản thân thường đứng ngay sau giới từ for, to: myself, yourself, himself,... |
| 8 | Liên từ | - Liên từ trong tiếng Anh được sử dụng để liên kết các cụm từ, mệnh đề, câu và đoạn văn.  - Liên từ được chia thành 3 loại:   * Liên từ kết hợp kết nối 2 hoặc nhiều hơn đơn vị từ tương đương nhau (2 câu văn, 2 mệnh đề, 2 cụm từ): and, or, so, but, for… * Liên từ tương quan luôn đi thành cặp, dùng để kết nối 2 đơn vị từ với nhau: Not only… but also, Either… or, ... * Liên từ phụ thuộc đứng trước mệnh đề phụ thuộc, gắn kết mệnh đề phụ thuộc với mệnh đề chính trong câu. |
| 9 | Thán Từ | - Thán từ là những từ mang nghĩa cảm thán, không có giá trị về mặt ngữ pháp nhưng thường được sử dụng rất thường xuyên trong văn nói.  - Ví dụ: oh, wow, yeah, oops, oh my god, ... |

### Cấu trúc câu tiếng Anh

Một cấu trúc câu đơn giản và khái quát của tiếng Anh sẽ như sau:



Trong đó:

* Chủ ngữ: Là thành phần chứa người/vật thực hiện hành động và là **thành phần bắt buộc** tham gia trong câu.
* Động từ: Là thành phần biểu hiện hành động của câu và là **thành phần bắt buộc** tham gia trong câu.
* Tân ngữ: Là thành phần chứa người/vật bị hành động tác động vào và là **thành phần không bắt buộc,** có thể có, có thể không có, tùy theo động từ.
* Các thông tin nền: Là thành phần bổ sung các thông tin khác liên quan đến hành động, ví dụ như nơi chốn, thời điểm, thời gian, cách thức, lý do mà hành động xảy ra. Đây là **thành phần không bắt buộc,** nếu có sẽ làm rõ ý nghĩa của câu hơn, còn nếu không có cũng không ảnh hưởng đến câu về mặt ngữ pháp.

Giải thích rõ hơn về tính bắt buộc của các thành phần:

* Câu văn đưa ra: Ngoc studied Math last night.
* Phân tích thành phần:
  + - Chủ ngữ: Ngoc
    - Động từ: Studied
    - Tân ngữ: Math
    - Các thông tin nền: last night
* Câu dịch: Ngọc đã học môn toán vào tối qua.
* Nếu chúng ta thử bỏ từng thành phần và dịch ra, ta sẽ có 3 biến thể sau:
* Bỏ chủ ngữ: Studied last night. Dễ thấy, câu không còn có nghĩa nữa (Đã học tối qua).
* Bỏ vị ngữ: Ngoc last night. Một lần nữa, câu không còn có ý nghĩa (Ngọc tối qua).
* Bỏ tân ngữ: Ngoc studied last night. Câu dù có thay đổi về nghĩa, ít chi tiết hơn nhưng vẫn là câu đúng (Ngọc đã học vào tối qua)
* Bỏ thông tin nền: Ngoc studied Math. Câu dù có thay đổi về nghĩa, ít chi tiết hơn nhưng vẫn là câu đúng (Ngọc đã học môn toán)

Xác định vị trí và tính chất của các thành phần câu:

* Chủ ngữ, thường sẽ là danh từ, cụm danh từ hoặc thuật ngữ; hầu hết sẽ luôn đứng đầu câu:

I love Math

* Tuy nhiên, nó có thể xuất hiện ở dưới hoặc cuối câu (đều có thể chuyển về dạng đầu câu):
* Đảo vị trí trạng từ lên đầu câu:

Usually I take a nap at noon.

* Tăng tính kịch cho câu:

On the table lay his books

* Câu mệnh lệnh:

Sit down!

* Khác với Tiếng Việt, phần vị ngữ của Tiếng Anh luôn luôn chứa động từ. Kết hợp với các thông tin ngoài để bổ ngữ cho động từ, ta sẽ có vị ngữ. Điều này không có nghĩa là vị ngữ chỉ chứa duy nhất 1 động từ.
* Ngoài ra, động từ có một nguyên tắc quan trọng: Một câu đơn chỉ có một động từ chính, và nếu một câu có nhiều động từ chính thì nó là một câu ghép, ghép từ nhiều câu con.
* Cách xác định động từ chính:
* Một động từ có thể chia được thành 6 dạng:

|  |  |
| --- | --- |
| Dạng | Ví dụ cho từ “eat” |
| Nguyên thủy/Đơn giản | eat |
| Thêm s/es để chia ngôi | eats |
| Quá khứ (V-ed) | ate |
| Thêm to + nguyên mẫu | to eat |
| V-ing (tiếp diễn) | eating |
| V-ed phân từ 2 | eaten |

* Động từ chính là động từ ở 3 dạng đầu do nó có thể xác định chính xác thì của từ; còn 3 dạng sau không thể xác định rõ được thì của câu chứa nó, vì vậy nó không được coi là động từ chính.
* Trong Tiếng Anh, còn có một loại động từ là động từ *to be*. Đây là một loại động từ rất quan trọng, dùng để bổ sung cho sự tồn tại, trạng thái với đặc điểm của một vật hoặc một người. Bảng tri thức của động từ *to be*:

|  |  |  |  |
| --- | --- | --- | --- |
| Dạng | Ngôi | Chủ ngữ số ít/Danh từ riêng | Chủ ngữ số nhiều |
| Khẳng định | Thứ nhất | I am | We are |
| Thứ hai | You are | You are |
| Thứ ba | She/He/It + is | They are |
| Phủ định | Thứ nhất | I am not | We are not |
| Thứ hai | You are not | You are not |
| Thứ ba | She/He/It + is not | They are not |
| Quá khứ | Thứ nhất | I was | We were |
| Thứ hai | You were | You were |
| Thứ ba | She/He/It + was | They were |

* Về các thành phần phụ trong câu, ta có thể sử dụng những loại từ đã được liệt kê ở trên, bao gồm:
* Tính từ
* Trạng ngữ
* Giới từ
* Liên từ
* Thán từ
* Từ hạn định

### Các thì trong tiếng Anh

Trong Tiếng Anh có tổng cộng là 12 thì, tuy nhiên do việc thiết lập các luật cho hệ thống cho cả 12 thì là rất phức tạp nên trong hệ thống này chỉ sử dụng 4 thì cơ bản chính là: Hiện tại đơn, Hiện tại tiếp diễn, Quá khứ đơn và Tương lai đơn. Chi tiết về tri thức của từng thì như sau:

* 1. Thì Hiện tại đơn:
* Định nghĩa: thì Hiện tại đơn dùng để diễn tả một sự thật hiển nhiên hay một hành động diễn ra lặp đi lặp lại theo thói quen, phong tục, khả năng.
* Công thức của thì Hiện tại đơn:

|  |  |  |
| --- | --- | --- |
| **Loại câu** | **Đối với động từ thường** | **Đối với động từ “to be”** |
| Khẳng định | S + V(s/es) + O | S + be (am/is/are) + O |
| Phủ định | S + do not /does not + V\_inf | S + be (am/is/are) + not + O |
| Ví dụ | * I study Math   (Tôi học môn toán)   * He doesn’t like candy   (Anh ấy không thích kẹo) | * She is a student.   (Cô ấy là học sinh)   * She is not a teacher   (Cô ấy không phải là giáo viên) |

* Trường hợp sử dụng:
* Diễn tả một sự thật hiển nhiên, chân lý.
* Diễn tả 1 hành động xảy ra thường xuyên, một thói quen ở hiện tại.
* Diễn tả một năng lực của con người.
* Dấu hiệu nhận biết: Thường là các câu theo định nghĩa, tính chất luôn đúng; các sự việc đang xảy ra (nhưng không được xét là “ngay bây giờ” và không có tính chất dừng lại) hoặc xảy ra thường xuyên. Vì vậy, các câu không có dấu thời gian thì sẽ được đưa về dạng hiện tại đơn.
  1. Thì Hiện tại tiếp diễn:
* Định nghĩa: thì Hiện tại tiếp diễn dùng để diễn tả những sự việc xảy ra ngay lúc chúng ta nói hay xung quanh thời điểm chúng ta nói, và hành động đó vẫn chưa chấm dứt (còn tiếp tục diễn ra).
* Công thức của thì Hiện tại tiếp diễn:

|  |  |
| --- | --- |
| **Loại câu** | **Đối với động từ thường** |
| Khẳng định | S + be (am/is/are) + V\_ing |
| Phủ định | S + be (am/is/are) + not + V\_ing |
| Ví dụ | * She is doing some homework.   (Cô ấy đang làm một số bài tập)   * I’m not watching TV now.   (Tôi đang không xem TV bây giờ) |

* Trường hợp sử dụng:
* Diễn tả một hành động đang diễn ra và kéo dài tại một thời điểm ở hiện tại.
* Thường tiếp theo sau mệnh lệnh, câu đề nghị.
* Dấu hiệu nhận biết: Trong câu Việt có từ “đang” và một số từ như “hôm nay hay còn gọi là các từ thể hiện mốc thời gian hiện tại và tiếp diễn.
  1. Thì Quá khứ đơn:
* Định nghĩa: thì Quá khứ đơn dùng để diễn tả một hành động, sự việc diễn ra và kết thúc trong quá khứ.
* Công thức của thì Hiện tại tiếp diễn:

|  |  |  |
| --- | --- | --- |
| **Loại câu** | **Đối với động từ thường** | **Đối với động từ “to be”** |
| Khẳng định | S + V2/ed + O | S + was/were + O |
| Phủ định | S + didn’t + V\_inf + O | S + was/were + not + O |
| Ví Dụ | - I watched TV yesterday  (Hôm qua tôi đã xem TV)  - I didn’t go to school yesterday.  (Hôm qua tôi đã không đi học) | - He was sad yesterday.  (Hôm qua anh ấy đã buồn)  - The shop was not open yesterday.  (Hôm qua cửa hàng không cửa) |

* Trường hợp sử dụng:
* Diễn tả hành động đã xảy ra và chấm dứt trong quá khứ.
* Diễn tả thói quen trong quá khứ.
* Dấu hiệu nhận biết: Trong câu Việt có các từ như “đã”, “hôm qua”, “hôm qua”, “ngày xưa”, … những từ đánh dấu mốc thời gian trong quá khứ.
  1. Thì Tương lai đơn:
* Định nghĩa: thì Tương lai đơn trong tiếng Anh được dùng khi không có kế hoạch hay quyết định làm gì nào trước khi chúng ta nói. Chúng ta ra quyết định tự phát tại thời điểm nói.
* Công thức của thì Hiện tại tiếp diễn:

|  |  |
| --- | --- |
| **Loại câu** | **Đối với động từ thường** |
| Khẳng định | S + will/shall/ + V\_inf + O |
| Phủ định | S + will/shall + not + V\_inf + O |
| Ví dụ | * I will study tomorrow   (Tôi sẽ học vào ngày mai)   * I’m not watching TV now.   (Tôi đang không xem TV bây giờ) |

* Trường hợp sử dụng:
* Diễn tả một hành động được xác định trước trong tương lai.
* Diễn tả một quyết định đột xuất ngay lúc nói.
* Dấu hiệu nhận biết: Trong câu Việt có các từ như “ngày mai”, “ngày kìa”, “sẽ”, … những từ đánh dấu mốc thời gian trong tương lai.

Bảng 3: Các tri thức về cơ sở chia động từ trong tiếng Anh

|  |  |  |  |
| --- | --- | --- | --- |
| STT | Tri thức | Cơ sở tri thức/suy luận | Ví dụ minh họa |
| 1 | Trong tiếng Việt là động từ trạng thái, tồn tại, đặc điểm hoặc không có động từ thì ta sẽ thêm động từ tobe.  Đối với danh từ/đại từ số ít và danh từ riêng, sử dụng be, is, was phụ thuộc vào thì; danh từ/đại từ số nhiều dùng are, were phụ thuộc vào thì; am dùng trong thì hiện tại đơn cho danh từ I (tôi). Động từ tobe là bất biến. | Danh sách: am, are, is, was, were.  Động từ tobe dùng để bổ sung cho sự tồn tại, trạng thái với đặc điểm của một vật hoặc một người. Vì vậy, nếu trong tiếng Việt có từ tương ứng giống như trên thì cần xử lý theo quy tắc của tiếng Anh. | She **is** beautiful  ( Cô ấy đẹp)  I **am** a student  (Tôi là một học sinh) |
| 2 | Đại từ số ít, danh từ không đếm được và đếm được nhưng ở số ít thì động từ sẽ chia thêm đuôi s hoặc es, còn lại giữ nguyên | Tập dữ liệu về động từ đã được chia trong cơ sở tri thức (Tập số 2).  Động từ cần được chia theo ngôi của chủ ngữ (danh từ và đại từ) | He **eats** noodles  (Anh ăn mì)  Quang **teaches** Ngoc  (Quang dạy học Ngoc) |
| 3 | Thì hiện tại đơn, động từ được giữ nguyên theo thể đã dịch, tuy nhiên cần chia theo loại danh/đại từ tương ứng:  - Động từ thường: Luật số 2, bảng 3  - Động từ tobe: Luật số 1, bảng 3  Với câu phủ định, nếu sử dụng động từ thường thì thêm động từ phụ trợ là “do” (cũng được chia theo Luật số 2, bảng 3) và từ :not” | Được ánh xạ từ tập dữ liệu ánh xạ (Tập số 1)  Áp dụng cơ sở tri thức của bảng Hiện tại đơn đã được trình bày ở trên: | I **do not study** Art  (Tôi không học mỹ thuật)  She does not like him  (Cô không thích hắn)  I **am not** a boy  (Tôi không phải là một con nít) |
| 4 | Thì hiện tại tiếp diễn, động từ được thêm đuôi ing (V-ing) và luôn có động từ tobe đứng trước. Động từ tobe được chia theo Luật số 1, bảng 2  Các động từ được chia theo đuôi ing sẽ có động từ thường và bất quy tắc, tuân theo Tập số 2 | Được ánh xạ từ tập dữ liệu ánh xạ (Tập số 1)  Động từ V-ing được tuân theo quy tắc của Tập số 2  Áp dụng cơ sở tri thức của bảng Hiện tại tiếp diễn đã được trình bày ở trên: | He **is learning**  (Anh đang học)  The computer **is not running**  (Cái máy tính đang không chạy)  We **are doing** some homeworks  (Chúng tôi đang làm vài bài tập) |
| 5 | Thì quá khứ đơn, động từ được thêm đuôi ed (V-ed). Tuy nhiên, cần tuân theo quy định sau để chia động từ chuẩn xác:  - Động từ thường: Luật số 2, bảng 3  - Động từ tobe: Luật số 1, bảng 3 | Được ánh xạ từ tập dữ liệu ánh xạ (Tập số 1)  Động từ V-ed được tuân theo quy tắc của Tập số 2  Áp dụng cơ sở tri thức của bảng Quá khứ đơn đã được trình bày ở trên: | He **was** here yesterday  (Chúng tôi đã ở đây hôm qua)  He **studied** last night  (Anh ấy đã học tối hôm trước) |
| 6 | Thì tương lai đơn, động từ được giữ nguyên thể, không áp dụng Luật số 1-2, bảng 3. Tuy nhiên, ta cần thêm phụ từ “will” vào trước các động từ (phụ từ “will” luôn luôn giữ nguyên) | Được ánh xạ từ tập dữ liệu ánh xạ (Tập số 1)  Áp dụng cơ sở tri thức của bảng Tương lai đơn đã được trình bày ở trên: | I **will** eat vegetables tomorrow  (Tôi sẽ ăn rau ngày mai)  He **will** not go  (Anh sẽ không đi) |
| 7 | Nếu trong Chủ ngữ có 2 Danh từ/Đại từ trở lên và có kết nối với nhau bởi liên từ thì sẽ coi đó là Danh từ số nhiều, và Động từ được chia theo Luật số 2, bảng 3 | Khi có 2 Danh từ/Đại từ được nối với nhau, lúc đó sẽ thành 2 Danh/Đại từ và sẽ thành Danh/Đại từ số nhiều | Quang **and** Ngoc **study**  (Quang và Ngọc học) |

Bảng 5: Các tri thức về sắp xếp câu trong tiếng Anh

|  |  |  |  |
| --- | --- | --- | --- |
| STT | Tri thức | Cơ sở tri thức/suy luận | Ví dụ minh họa |
| 1 | Cấu trúc đơn thuần:  Danh từ/Đại từ + Động từ + Tân ngữ (Danh từ/Đại từ)  Các thành phần phụ trợ sẽ tùy vào chức năng mà thêm vào.  S + V + O | Phần 2: Cấu trúc câu tiếng Anh | I play football  (Tôi chơi bóng đá)  I study English  (Tôi học tiếng anh) |
| 2 | Tính từ có vị trí đứng trước danh từ trong câu  Custom N = Adj + Noun  Tuy nhiên, nếu trong từng phần (Chủ ngữ, vị ngữ) có chứa Tính từ mà không chứa Danh từ, Tính từ sẽ được giữ nguyên vị trí và dịch thuần sang tiếng anh | Tính từ bổ sung ý nghĩa cho danh từ, tạo thành một cụm danh từ | This is a **beautiful** picture  (Đây là một bức tranh đẹp)  I study very **good**  (Tôi học rất giỏi) |
| 3 | Trạng từ có thể đứng ở mọi vị trí trong câu (trước Danh từ, trước Động từ, sau Động từ, sau Tính từ) nhưng ta có thể quy ước nó như sau  - Nếu là trạng từ chỉ thời gian, tần suất, mức độ, trạng từ sẽ đứng luôn ở trước Động từ/Tính từ chính mà nó bổ trợ:  Custom V/Adj = Adverb + Verb/Adj  - Nếu là các loại trạng từ khác thì trạng từ sẽ luôn đứng sau động từ chính mà nó bổ trợ:  Custom V = Verb + Adverb | Trạng từ có các chức năng chính như sau:  Bổ nghĩa cho động từ.  Bổ nghĩa cho trạng từ khác.  Dễ thấy, trạng từ phụ thuộc vào động từ nên ta có thể đưa về trước hoặc sau động từ. | I **usually** swim  (Tôi thường bơi)  He is concentrating **completely** on something  (Anh đang tập trung hoàn toàn vào gì đó) |
| 4 | Giới từ phải đứng trước danh từ hoặc đại từ (phụ) và đứng sau động từ hoặc danh từ  Noun/Pronoun + Preposition + Verb/Objects | Do giới từ là những từ dùng để diễn tả mối quan hệ của cụm từ đứng phía sau nó với các thành phần khác trong câu, cho nên đứng trước nó phải là một vật xác định, còn sau là một hành động hoặc một vật khác | He is **near** the computer  (Anh bên cạnh cái máy tính)  I am **behind** you  (Tôi bên sau bạn) |
| 5 | Liên từ phải ở giữa 2 loại từ cùng loại nhau hoặc 2 mệnh đề  S1 + Conjunction + S2 | Do liên từ có chức năng kết nối 2 loại từ cùng lại nó phải ở giữa 2 từ | You **and** me love dog  (Bạn và tôi thích chó)  He **or** she will go  (Anh hoặc cô sẽ đi) |
| 6 | Nếu là danh từ số ít xuất phát bởi ‘ueoai’, thì từ đứng ở trước là ‘a’ sẽ phải chuyển đổi thành ‘an’ | Cần chia lượng từ đứng trước danh từ | I have an apple  (Tôi có một quả táo) |
| 7 | Tất cả từ hạn định luôn đứng trước danh từ trong mọi trường hợp. | Từ hạn định luôn bổ sung ý nghĩa cho danh từ ,vì vậy nó luôn đứng ngay trước danh từ. | This is **an** apple  (Đây là một quả táo)  I love **this** cat  (Tôi thích con mèo này) |
| 8 | 2 danh từ đi liền nhau trong Tiếng Việt thì khí chuyển sang Tiếng Anh thì cần đảo chỗ 2 danh từ đó với nhau | Do sự khác biệt về cấu tạo cụm từ giữa hai ngôn ngữ | I have a **tree garden**  (Tôi có một vườn cây) |

# Dữ liệu

Tiếng Việt và Tiếng Anh đều có cấu trúc chủ vị; chủ ngữ, vị ngữ đều là được cấu tạo từ sự kết hợp của các từ loại. Tuy nhiên từ loại và cấu tạo cụm từ của Tiếng Việt và Tiếng Anh có sự khác biệt, do đó cần xác định đúng từ loại, động từ chính, thì của câu Tiếng Việt để thực hiện việc ánh xạ sang từ loại Tiếng Anh, ghép thành cụm từ, câu Tiếng Anh đúng về mặt ngữ pháp và nội dung. Dựa vào đó, chúng ta sẽ tách dữ liệu thành 2 nhánh là Tiếng Việt và Tiếng Anh. Về cấu trúc cơ sở tri thức sẽ như sau:

## Tiếng Việt

Từ điển tiếng Việt sẽ được chia theo 14 file từ loại, dạng file .txt. Việc chia nhỏ từ loại Tiếng Việt dựa trên sự tương đồng, có thể ánh xạ sang được từ loại Tiếng Anh, và khả năng thực hiện bài toán dịch câu Tiếng Việt sang câu Tiếng Anh của bọn em, bao gồm:

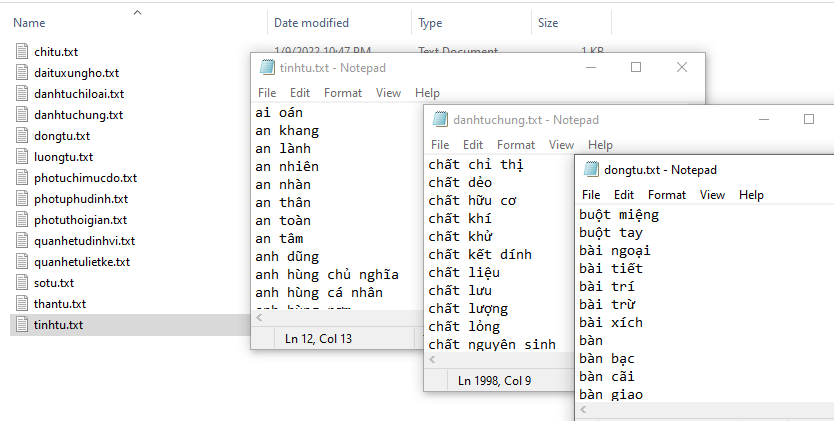
|  |  |  |  |
| --- | --- | --- | --- |
| STT | Tên file | Từ loại | Số lượng từ/cụm từ |
| 1 | chitu.txt | Chỉ từ | 8 |
| 2 | daituxungho.txt | Đại từ xưng hô | 36 |
| 3 | danhtuchiloai.txt | Danh từ chỉ loại | 45 |
| 4 | dangtuchung.txt | Danh từ chung | 14982 |
| 5 | dongtu.txt | Động từ | 9860 |
| 6 | luongtu.txt | Lượng từ | 35 |
| 7 | photuchimucdo.txt | Phó từ chỉ mức độ | 22 |
| 8 | photuphudinh.txt | Phó từ chỉ sự phủ định | 4 |
| 9 | photuthoigian.txt | Phó từ chỉ quan hệ thời gian | 18 |
| 10 | quanhetudinhvi.txt | Quan hệ từ định vị | 20 |
| 11 | quanhetulietke.txt | Quan hệ từ liệt kê | 4 |
| 12 | sotu.txt | Số từ | 16 |
| 13 | thantu.txt | Thán từ | 62 |
| 14 | tinhtu.txt | Tính từ | 7281 |

Mỗi file từ loại sẽ bao gồm các từ hoặc cụm từ thuộc từ loại đó. Mỗi từ, cụm từ được viết trên một dòng và không có trùng lặp trong từ, cụm từ trong một file từ loại.

Một số đặc điểm của bộ dữ liệu Tiếng Việt:

* Một từ có thể có nhiều từ loại.
* Mỗi từ, cụm từ nằm trong file chitu.txt, daituxungho.txt, luongtu.txt, photuchimucdo.txt, photuphudinh.txt, photuthoigian.txt, quanhetulietke.txt, thantu.txt chỉ có duy nhất 1 từ loại
* Các từ, cụm từ nằm trong các file còn lại có thể có nhiều hơn 1 từ loại

Minh hoa tập dự liệu Tiếng Việt:



## Tiếng Anh

### Tập dữ liệu số 1

Do một từ tiếng Việt sau khi được tách sẽ được ánh xạ bao gồm 2 tri thức là “từ” và “loại” sau khi đã được chuyển đổi bởi các luật. Vì vậy, từ điển tiếng anh sẽ có cấu trúc như sau (định dạng JSON):

{

"word": "ăn",

"type": [

"verb"

],

"trans1": [

"eat",

"feed",

"take",

"have"

]

},

* Đánh giá về cấu trúc trên, có 3 thành phần chính gồm:
* “word”: Từ có được so sánh sang tiếng Việt.
* “type”: Loại từ của từ đó trong tiếng Anh, bao gồm các loại là: noun, verb, adv, adj, conj, excl, prep, pronoun.
* “trans1”: Là phiên bản dịch của từ tiếng Việt sang tiếng Anh. Có thể tồn tại nhiều bản dịch cho 1 từ tiếng Việt và số 1 thể hiện cho sự ánh xạ của kiểu từ với từ được dịch. Ví dụ, trường hợp trên là trường hợp 1 từ chỉ ra được 1 loại là động từ, và trong tiếng Anh chỉ tồn tại 1 từ có ý nghĩa như vậy. Ta có thể ví dụ thêm một vài trường hợp, ví dụ như một loại từ có nhiều từ đồng nghĩa:

{

"word": "ăn",

"type": [

"verb"

],

"trans1": [

"eat",

"feed",

"take",

"have"

]

},

* Hoặc trong trường hợp, có nhiều loại từ cho 1 từ tiếng Việt như:

{

"word": "sáng tạo",

"type": [

"verb",

"adj"

],

"trans1": [

"create"

],

"trans2": [

"creative"

]

},

* Quá trình khi tìm ánh xạ, ta sẽ tìm một từ tiếng Việt trong danh sách, dùng các luật để ánh xạ kiểu loại từ trong tiếng Việt sang kiểu loại từ trong tiếng Anh và chọn lựa từ có trong từ điển trên. Trong trường hợp có nhiều khả năng lựa chọn, sẽ hiển thị ra từ xuất hiện đầu tiên làm mặc định, còn lại sẽ hiển thị tất cả từ còn lại, cho phép người nhập có thể tự chọn theo ý muốn hoặc để mặc định.

### Tập dữ liệu số 2

Tuy nhiên, như đã nhắc ở tri thức trên, động từ sẽ cần được chia theo các thì, và có các quy tắc nhất định tuân theo từng loại thì. Tuy nhiên, do tồm tạo một số từ khi được chia ở thì quá khứ (V-ed) sẽ không theo quy tắc truyền thống (động từ bất quy tắc) nên ta sẽ tạo lập ra một từ điển thứ 2, là về danh sách tất cả các động từ đã được chia theo các dạng của nó:

[

"run",

"runs",

"ran",

"run",

"running"

],

* Nhận xét về cấu trúc trên, ta có một số nhận xét:
  + - Động từ sẽ ở tổng cộng là 5 dạng, bao gồm:
* Nguyên thể (V-inf)
* Chia theo ngôi
* Quá khứ (V-ed)
* Phân từ 2 (V-pII)
* Tiếp diễn (V-ing)
* Do toàn bộ các từ có ở từ điển 1 đều là động từ ở thể nguyên thể, ta chỉ cần dựa vào các luật để xác định thì của câu, từ đó lựa chọn dạng chia động từ phù hợp theo từ điển 2 này.

### Tập dữ liệu số 3

Liệt kê các danh từ thuộc kiểu “danh từ số ít” phục vụ cho việc chia động từ theo ngôi. Dữ liệu gồm hơn 48000 danh từ số ít trong tiếng anh, được xếp trong file JSON như sau:

 "auranofin": {},

  "aurantiasis": {},

  "aureomycin": {},

  "auric acid": {},

  "auriculoacupuncture": {},

  "auriculotherapy": {},

  "auricupride": {},

  "aurigraphy": {},

  "aurin": {},

  "aurintricarboxylic acid": {},

  "auriscopy": {},

### Tập dữ liệu số 4

Liệt kê các đại từ thuộc kiểu “đại từ số ít” và “đại từ số nhiều” phục vụ cho việc chia động từ theo ngôi. Dữ liệu đã được chia theo 2 loại trên, được xếp trong file JSON như sau:

[

    {

        "singular pronouns":[

            "he",

            "him",

            "his",

            "himself",

            "she",

            "her",

            "herself",

            "it",

            "itself",

            "something",

            "someone",

            "nobody",

            "nothing",

            "anybody"

        ]

    },

    {

        "plural pronouns":[

            "I",

            "mine",

            "my",

            "myself",

            "your",

            "yourself",

            "you",

            "we",

            "us",

            "our",

            "ours",

            "ourselves",

            "yours",

            "yourselves",

            "its",

            "they",

            "them",

            "their",

            "theirs",

            "themself",

            "themselves",

            "all",

            "everyone",

            "everything"

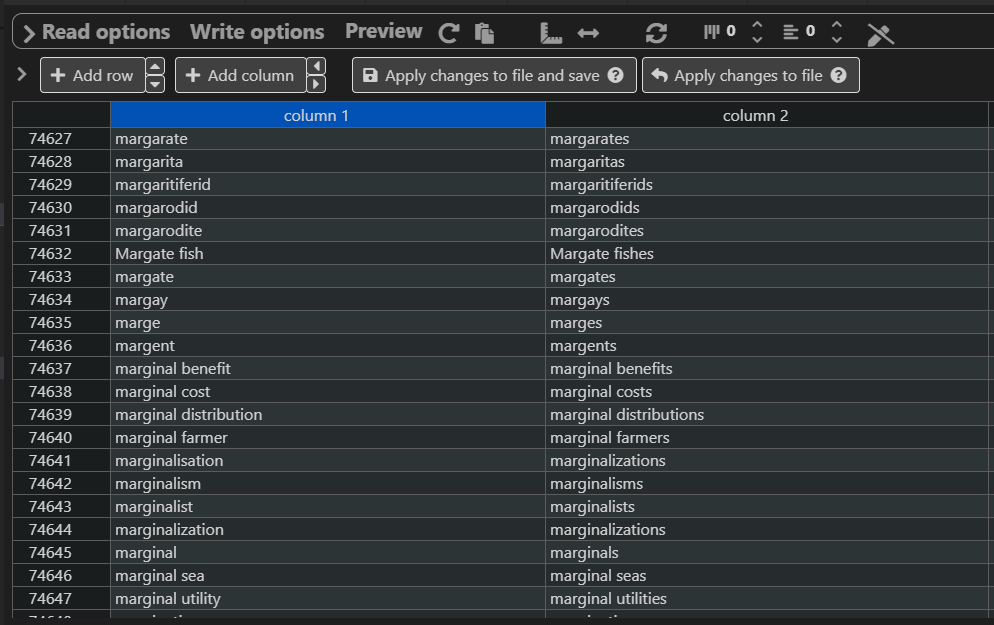
        ]

    }

]

### Tập dữ liệu số 5

Liệt kê các danh từ số ít và danh từ số nhiều tương ứng của nó phục vụ cho việc nhận biết và đưa danh từ về dạng danh từ số ít hay danh từ số nhiều theo mong muốn. Dữ liệu đã được chia 2 cột tương ứng với cột danh từ số ít và cột danh từ số nhiều, mỗi hàng là một cặp danh từ số ít-nhiều, được xếp trong file CSV như sau:



### Tập dữ liệu số 6

Liệt kê danh từ không đếm được, được coi là danh từ số ít khi chia động từ. Mỗi dòng sẽ có một danh từ không đếm, được viết ở đầu dòng, được xếp trong file .txt như sau:

meitnerium

melody

membership

mendelevium

merchandise

mercury

mercy

metal

might

migration

milk

minute

misconception

misdemeanour

misery

# Luật

Bảng 2: Các luật về cơ sở phân loại từ trong tiếng Việt

|  |  |  |  |
| --- | --- | --- | --- |
| STT | Quy tắc | Cơ sở tri thức | Ví dụ minh họa |
| 1 | Từ không tồn tại trong từ điển thì từ loại của nó là “unknown” | Do dữ liệu Tiếng Việt không thể bao quát tất cả các từ trong thực tế |  |
| 2 | Kí tự “,” (dấu phẩy) được gán từ loại là “lietke” | Do dấu phẩy để ngăn cách các vế tương đồng, được sử dụng để nhận biết tự loại |  |
| 3 | Các kí tự đặc biệt khác như -:%(),… được gán từ loại “kitu” | Do cả 2 ngôn ngữ đều có các kí tự này, và chỉ cần chuyển trực tiếp sang bên dịch mà không cần thao tác gì thêm |  |
| 4 | Xác định thì cho câu tiếng Anh bằng các phó từ chỉ quan hệ thời gian.  - Câu có từ *đã*, *từng, hôm qua, hôm xưa, ngày xưa* là thì quá khứ đơn  - Câu có từ *đang, hôm nay* là thì quá khứ đơn  - Câu có từ *sẽ*, *ngày mai, ngày kìa* thì là thì tương lai đơn  - Các trường hợp còn lại coi là thì hiện tại đơn | Dựa trên tri thức các thì trong câu tiếng Việt | con mèo *đã* ăn cơm  -> thì quá khứ đơn  con mèo ăn cơm  -> thì hiện tại đơn |
| 5 | Từ viết hoa chữ cái đầu tiên thì coi là danh từ riêng | Dựa trên tri thức danh từ riêng | Quang làm bài tập  *danh từ riêng* |
| 6 | Từ đứng ở đầu câu không có từ loại thì coi là danh từ chung | Dựa trên tri thức về chủ ngữ trong cấu trúc câu Tiếng Việt |  |
| 7 | Nhận biết danh từ chỉ loại hay động từ:  - Nếu trước nó là số từ, lượng từ, động từ hoặc quan hệ từ định vị thì là danh từ chỉ loại  - Trường hợp còn lại là động từ | Dựa trên tri thức Bảng 1: Các tri thức về sắp xếp cụm từ, câu trong tiếng Việt | tôi có *một* bó gạo  số từ  => bó là danh từ chỉ loại  tôi *đang* bó gạo  phó từ chỉ thời gian  => bó là động từ |
| 8 | Nhận biết quan hệ từ định vị:  - Nếu trước nó là động từ hoặc tính từ thì là quan hệ từ định vị  - Trường hợp còn lại có thể là danh từ, động từ hoặc tính từ | Dựa trên tri thức Bảng 1: Các tri thức về sắp xếp cụm từ, câu trong tiếng Việt | *ở* có thể là động từ hoặc quan hệ từ định vị  nó *chơi*  ở nhà  động từ  => trước *ở* là động từ nên *ở* là quan hệ từ định vị  *nó* ở nhà  đại từ xưng hô  => trước *ở* là đại từ xưng hô nên *ở* là động từ |
| 9 | Từ, cụm từ viết liền nhau, cách nhau bởi dấu phẩy hoặc quan hệ từ liệt kê thì chúng cùng từ loại | Dựa trên tri thức của quan hệ từ liệt kê tiếng Việt | cô ấyxinh đẹp*,* giỏi giang*,* *và* yêu kiều  *xinh đẹp, giỏi giang, yêu kiều* cùng từ loại |
| 10 | Nhận biết danh từ chung:  - Nếu trước nó là số từ, lượng từ, danh từ chỉ loại hoặc quan hệ từ định vị thì là danh từ chung  - Nếu nó là danh từ chỉ loại nhưng nó nằm ở cuối câu hoặc sau nó không phải là danh từ thì là danh từ chung  - Nếu sau nó là chỉ từ, động từ, tính từ hoặc phó từ thì là danh từ chung | Dựa trên tri thức các thì trong câu tiếng Việt, Bảng 1: Các tri thức về sắp xếp cụm từ, câu trong tiếng Việt | *quan niệm* có thể là danh từ hoặc động từ.  *một* quan niệm *đúng*  số từ tính từ  => trước *quan niệm* là số từ, sau *quan niệm* là tính từ nên nó là danh từ chung.  nó có ba con  danh từ chỉ loại  *con* là danh từ chỉ loại nhưng nó ở cuối cau nên là danh từ chung. |
| 11 | Nhận biết động từ:  - Nếu trước nó là phó từ, danh từ chung, đại từ xưng hô thì là động từ  - Nếu sau nó là quan hệ từ chỉ mức độ, quan hệ từ định vị, số từ, lượng từ, danh từ, tính từ thì là động từ | Dựa trên tri thức các thì trong câu tiếng Việt, Bảng 1: Các tri thức về sắp xếp cụm từ, câu trong tiếng Việt | *quan niệm* có thể là danh từ hoặc động từ.  *em* quan niệm  đại từ xưng hô  => trước *quan niệm* là đại từ xưng hô nên *quan niệm* là động từ |
| 12 | Nhận biết tính từ:  - Nếu trước nó là phó từ, danh từ chung , đại từ xưng hô, động từ thì là tính từ  - Nếu sau nó là quan hệ từ chỉ mức độ | Dựa trên tri thức các thì trong câu tiếng Việt, Bảng 1: Các tri thức về sắp xếp cụm từ, câu trong tiếng Việt | *vặn vẹo* có thể là tính từ hoặc động từ.  con mèo *nằm* vặn vẹo  động từ  => trước *vặn vẹo* là động từ nên *vặn vẹo* là tính từ |
| 13 | Xác định động từ chính:  - Nếu trước động từ/tính có phó từ chỉ thời gian, phó từ phủ định thì là động từ chính  - Trường hợp còn lại thì động từ/tính từ đầu tiên trong câu là động từ chính | Dựa trên nhu cầu cần có động từ chính để chia thì và chia động từ khi dịch sang câu Tiếng Anh | tôi *từng* chơi game  => *từng* phó từ chỉ thời gian, *chơi* là động từ chính |
| 14 | Các từ Tiếng Việt sẽ bỏ qua, không gửi sang bên dịch:  - Danh từ chỉ loại  - Phó từ chỉ sự phủ định  - Các từ chỉ thời gian: đã, từng, sẽ ,đang | Các từ định nghĩa thì không tồn tại trong câu Tiếng Anh mà thể hiện qua động từ  Dựa trên tri thức về đặc điểm của bộ dữ liệu Tiếng Anh | *con* mèo *đang* ăn cơm  *con* là danh từ chỉ loại, *đang* là phó từ chỉ thời gian cần loại bỏ nên câu chuyển sang là:  mèo ăn cơm |

Bảng 4: Các luật về cơ sở ánh xạ từ Việt - Anh

|  |  |  |  |
| --- | --- | --- | --- |
| STT | Quy tắc | Cơ sở tri thức/suy luận | Ví dụ minh họa |
| 1 | Các kí tự đặc biệt sẽ được giữ nguyên |  | I**,** you and cat eat fish  (Tôi**,** bạn và mèo ăn cá) |
| 2 | - Danh từ chung trong tiếng Việt sẽ được ánh xạ sang kiểu loại Danh từ trong tiếng Anh.  - Danh từ riêng được giữ nguyên như tiếng Việt, coi như là Danh từ số ít trong Tiếng Anh | Danh từ tiếng Việt hay tiếng Anh đều đóng vai trò như nhau. | I like **orange**  (tôi thích **cam**)  **Ngoc** is a girl  (**Ngọc** là một cô gái) |
| 3 | - Động từ trong Tiếng Việt sẽ được chuyển thành Động từ trong Tiếng Anh, dạng thêm động từ tobe và được chia theo thì như quy định của bảng 3, theo ngôi như quy định của luật 1 bảng 3  Động từ thường ánh xạ không thay đổi, chỉ phải chia theo quy tắc bảng 3 | Như đã giải thích ở luật 1 bảng 3 | She **is** a teacher  (cô là một giáo viên) |
| 4 | Tính từ trong tiếng Việt sẽ được ánh xạ sang kiểu loại Tính từ trong tiếng Anh. | Tính từ tiếng Việt hay tiếng Anh đều đóng vai trò như nhau. | She runs very **fast**  (cô chạy rất **nhanh**) |
| 5 | Số từ trong tiếng Việt sẽ được ánh xạ sang kiểu loại Tính từ - Số lượng trong tiếng Anh. | Về cơ bản, 2 loại từ đều là một, chỉ là việc phân loại giữa 2 ngôn ngữ là khác nhau. | We have **one** last opportunity.  (chúng ta có **một** cơ hội cuối cùng) |
| 6 | Lượng từ trong tiếng Việt sẽ được ánh xạ sang kiểu loại Từ hạn định - Lượng từ trong tiếng Anh. | Về cơ bản, 2 loại từ đều là một, chỉ là việc phân loại giữa 2 ngôn ngữ là khác nhau. | I am doing **a few** test  (tôi đang làm một vài kiểm tra) |
| 7 | Phó từ trong tiếng Việt sẽ được ánh xạ sang kiểu loại Trạng từ trong tiếng Anh (ánh xạ với từng loại)  Trạng từ luôn bổ sung ý nghĩa cho từ gần nhất đứng với nó (trong tiếng Việt, Phó từ thường sẽ đứng trước Động từ/Tính từ mà nó bổ ngữ) | Phó từ tiếng Việt hay Trạng từ tiếng Anh đều đóng vai trò như nhau. | He is texting **skillfully**  (anh đang khéo léo nhắn tin) |
| 8 | Đại từ trong tiếng Việt sẽ được ánh xạ sang kiểu loại Đại từ trong tiếng Anh | Đại từ tiếng Việt hay tiếng Anh đều đóng vai trò như nhau. | **He** is doing something  (anh đang làm gì đó) |
| 9 | Chỉ từ trong tiếng Việt sẽ được ánh xạ sang kiểu loại Từ hạn định - Từ chỉ định trong tiếng Anh | Về cơ bản, 2 loại từ đều là một, chỉ là việc phân loại giữa 2 ngôn ngữ là khác nhau. | I love **this** girl  (tôi yêu cô gái này) |
| 10 | Quan hệ từ trong tiếng Việt sẽ được ánh xạ sang nhiều kiểu loại từ khác nhau trong tiếng Anh, bao gồm:  - Quan hệ từ Sở hữu sẽ được ánh xạ sang Từ hạn định - Từ sở hữu  - Quan hệ từ Liệt kê, Tương phản, Nguyên nhân - Kết quả sẽ được ánh xạ sang Liên từ  - Quan hệ từ Định vị sẽ được ánh xạ sang Giới từ | Quan hệ từ là một tập hợp các loại từ có vai trò kết nối các từ - mệnh đề trong câu với nhau. Nó có đặc điểm khá tương đồng với Liên từ trong tiếng Anh. Tuy nhiên, Quan hệ từ của tiếng Việt bao trùm rất nhiều ý nghĩa khác nhau. Vì vậy, ta sẽ chia thành các từ loại khác nhau trong tiếng Anh. | I have a computer **and** a book  (tôi có một máy tính và một cuốn sách) |
| 11 | Thán từ trong tiếng Việt sẽ được ánh xạ sang kiểu loại Thán từ trong tiếng Anh | Thán từ tiếng Việt hay tiếng Anh đều đóng vai trò như nhau. | **Oh**, I have a plan  (ồ, tôi có một kế hoạch) |
| 12 | Đại từ trong vế vị ngữ (bên phải động từ chính) sẽ chuyển về đại từ chỉ đối tượng | Do vị ngữ là thành phần bổ ngữ cho chủ ngữ nên đại từ thuộc chủ ngữ phải trở thành 1 đối tượng, không phải một nhân xưng | I like **her**  (tôi thích cô ấy) |
| 13 | Nếu gặp từ “của”  - Nếu từ sau nó là Đại từ, chuyển nó về Từ hạn định - Từ sở hữu rồi đảo lên phía trước danh từ đứng trước từ “của”  - Nếu sau nó là 1 Danh từ riêng, thêm đuôi ‘s ở sau Danh từ riêng rồi đảo lên phía trước danh từ đứng trước từ “của” | Do trong Tiếng Anh, từ của đối với Đại từ được thể hiện qua Từ hạn định – Từ sở hữu | **My** cat eats fish  (con mèo của tôi ăn cá)  **Ngoc’s** cat eats fish  (mèo của Ngọc ăn cá) |
| 14 | Xác định ngôi của chủ ngữ theo các luật  - Nếu chứa quan hệ từ liệt kê, đưa về chủ ngữ số nhiều  - Nếu chứa lượng từ hoặc số từ số nhiều, đưa chủ ngữ về số nhiều  - Nếu là đại từ số nhiều, đưa chủ ngữ về số nhiều  - Nếu là danh từ số nhiều, đưa chủ ngữ về số nhiều  - Còn lại là danh từ số ít | Do chủ ngữ được quy định bởi các luật về việc phân chia số nhiều và số ít nên ta cần chia như vậy | I have two **cats**  (tôi có hai con mèo)  She **uses** hammer  (cô ấy sử dụng búa) |
| 15 | Sắp xếp Danh từ và Trạng từ |  |  |
| 16 | Sắp xếp Từ định danh đứng trước Danh từ:  - Nếu gặp Từ định danh, tìm Danh từ đứng ngay trước nó và thực hiện đảo vị trí của chúng | Dựa trên Bảng 5: Các tri thức về sắp xếp câu trong tiếng Anh | con mèo này xinh  cat this is pretty  => chuyển thành  This cat is pretty |
| 17 | Sắp xếp 2 danh từ đứng liền kề nhau:  Thực hiện việc đảo vị trí 2 Danh từ cho nhau | Dựa trên Bảng 5: Các tri thức về sắp xếp câu trong tiếng Anh | tôi có một **vườn** **cây**  I have a **garden tree**  **=>** chuyển thành  I have a **tree garden** |
| 18 | Xác định **a, an** cho Danh từ số ít:  - Nếu danh từ bắt đầu bằng nguyên âm **u, e, o, a, i** thìchọn **an**  - Trường hợp còn lại chọn **a** | Dựa trên Bảng 5: Các tri thức về sắp xếp câu trong tiếng Anh | I have **a** cat  (tôi có một con mèo)  I have **an** umbrella  (tôi có một cái dù) |

# Kịch bản

Dòng thực thi chung: Người dùng nhập câu đầu vào => Tách các từ trong câu đầu vào => Áp dụng tri thức bảng 1 và luật bảng 2 để xác định loại từ của các từ tiếng Việt và các thành phần (động từ chính để tách chủ - vị) trong câu => Áp dụng tri thức của thì trong tiếng Việt, xác định thì của câu đầu vào => Áp dụng luật bảng 4 để ánh xạ các từ sang tiếng Anh (trong đó có sử dụng luật bảng 3 để chia động từ) => In ra cho người dùng câu văn đã được dịch và các lựa chọn về từ loại đối với từng từ hoặc cụm từ Tiếng Việt được tách ra.

* 1. Kịch bản số 1

Mọi từ đều có trong từ điển, các luật được sử dụng chặt chẽ trong mọi từ.

Người dùng nhập vào câu văn:

tôi và Ngọc đang chạy rất nhanh

Tách các từ ra ta được:

“tôi”, “và”, “Ngọc”, “đang”, “chạy”, “rất”, “nhanh”

|  |  |  |  |
| --- | --- | --- | --- |
| Bước | Dữ liệu vào | Dữ liệu ra | Các luật/tập dữ liệu được áp dụng |
| 1 | “tôi”, “và”, “Ngọc”, “đang”, “chạy”, “rất”, “nhanh” | “tôi”:”Đại từ xưng hô, 1”, “và”:”Quan hệ từ liệt kê, 2”, “Ngọc”:”Danh từ riêng, 3”, “đang”:”Phó từ chỉ thời gian, 4”, “chạy”:”Động từ, 5”, “rất”:”Phó từ chỉ mức độ, 6”, “nhanh”:”Tính từ, 7” | Tập dữ liệu Tiếng Việt  Luật bảng 2 và các tri thức tiếng Việt để tìm ra các loại từ có trong câu  Đánh số vị trí của chúng trong câu tiếng Việt ban đầu |
| 2 | “tôi”:”Đại từ xưng hô, 1”, “và”:”Quan hệ từ liệt kê, 2”, “Ngọc”:”Danh từ riêng, 3”, “đang”:”Phó từ chỉ thời gian, 4”, “chạy”:”Động từ, 5”, “rất”:”Phó từ chỉ mức độ, 6”, “nhanh”:”Tính từ, 7”, “Động từ chính”:”5” | “tôi”:”Đại từ xưng hô, 1”, “và”:”Quan hệ từ liệt kê, 2”, “Ngọc”:”Danh từ riêng, 3”, “chạy”:”Động từ, 4”, “rất”:”Phó từ chỉ mức độ-Tính từ-6, 5”, “nhanh”:”Tính từ, 6”, “Động từ chính”:”4”, “Thì”:”Hiện tại tiếp diễn” | Luật bảng 2 để tìm ra thì của câu  Xóa phó từ thời gian phân biệt thì  Sắp xếp lại thứ tự từ  Tìm động từ chính dựa vào luật bảng 2  Tìm chủ từ của các Phó từ còn lại theo Luật số 7, bảng 4 |
| 3 | “tôi”:”Đại từ xưng hô, 1”, “và”:”Quan hệ từ liệt kê, 2”, “Ngọc”:”Danh từ riêng, 3”, “chạy”:”Động từ, 4”, “rất”:”Phó từ chỉ mức độ-Tính từ-6, 5”, “nhanh”:”Tính từ, 6”, “Động từ chính”:”4”, “Thì”:”Hiện tại tiếp diễn” | “tôi”:”Đại từ, 1”, “và”:”Liên từ, 2”, “Ngọc”:”Danh từ riêng, 3”, “chạy”:”Động từ, 4”, “rất”:”Trạng từ mức độ-Tính từ-6, 5”, “nhanh”:”Tính từ, 6”, “Động từ chính”:”4”, “Thì”:”Hiện tại tiếp diễn” | Luật bảng 4 để tìm ra sự ánh xạ của các từ tiếng Việt sang tiếng Anh |
| 4 | “tôi”:”Đại từ xưng hô, 1”, “và”:”Quan hệ từ liệt kê, 2”, “Ngọc”:”Danh từ riêng, 3”, “chạy”:”Động từ, 4”, “rất”:”Phó từ chỉ mức độ-Tính từ-6, 5”, “nhanh”:”Tính từ, 6”, “Động từ chính”:”4”, “Thì”:”Hiện tại tiếp diễn” | “tôi”:”Đại từ, 1”, “và”:”Liên từ, 2”, “Ngọc”:”Danh từ riêng, 3”, “chạy”:”Động từ, 4”, “rất”:”Trạng từ mức độ-Tính từ-6, 5”, “nhanh”:”Tính từ, 6”, “Động từ chính”:”4”, “Thì”:”Hiện tại tiếp diễn” | Luật bảng 4 để tìm ra sự ánh xạ của các từ tiếng Việt sang tiếng Anh |
| 5 | “tôi”:”Đại từ, 1”, “và”:”Liên từ, 2”, “Ngọc”:”Danh từ riêng, 3”, “chạy”:”Động từ, 4”, “rất”:”Trạng từ chỉ mức độ-Tính từ-6, 5”, “nhanh”:”Tính từ, 6”, “Động từ chính”:”4”, “Thì”:”Hiện tại tiếp diễn” | “I”:”Đại từ, 1”, “and”:”Liên từ, 2”, “Ngoc”:”Danh từ riêng, 3”, “run”:”Động từ, 4”, “very”:”Trạng từ chỉ mức độ-Tính từ-6, 5”, “fast”:”Tính từ, 6”, “Động từ chính”:”4”, “Thì”:”Hiện tại tiếp diễn” | Dùng bảng dữ liệu 1 để ánh xạ sang tiếng anh với loại từ tương ứng |
| 7 | “I”:”Đại từ, 1”, “and”:”Liên từ, 2”, “Ngoc”:”Danh từ riêng, 3”, “run”:”Động từ, 4”, “very”:”Trạng từ chỉ mức độ-Tính từ-6, 5”, “fast”:”Tính từ, 6”, “Động từ chính”:”4”, “Thì”:”Hiện tại tiếp diễn” | “I”:”Đại từ, 1”, “and”:”Liên từ, 2”, “Ngoc”:”Danh từ riêng, 3”, “are running”:”Động từ, 4”, “very”:”Trạng từ chỉ mức độ-Tính từ-6, 5”, “fast”:”Tính từ, 6”, “Động từ chính”:”4”, “Thì”:”Hiện tại tiếp diễn” | Dùng bảng dữ liệu 2 để ánh xạ sang đúng kiểu của thì (kiểu V-ing)  Dùng bảng dữ liệu 3 và 4 để tìm ngôi cho danh từ và đại từ  Dùng luật chia động từ theo thì và ngôi của bảng 3  Ngoài ra, xét tri thức số 7 của bảng 3 để chia động từ do có xuất hiện liên từ trong danh từ |
| 8 | “I”:”Đại từ, 1”, “and”:”Liên từ, 2”, “Ngoc”:”Danh từ riêng, 3”, “are running”:”Động từ, 4”, “very”:”Trạng từ chỉ mức độ-Tính từ-6, 5”, “fast”:”Tính từ, 6”, “Động từ chính”:”4”, “Thì”:”Hiện tại tiếp diễn” | “I”:”Đại từ, 1”, “and”:”Liên từ, 2”, “Ngoc”:”Danh từ riêng, 3”, “are running”:”Động từ, 4”, “very”:”Trạng từ chỉ mức độ-Tính từ-6, 5”, “fast”:”Tính từ, 6”, “Động từ chính”:”4”, “Thì”:”Hiện tại tiếp diễn” | Dùng các tri thức có trong bảng 5 để sắp xếp các loại từ:  - Chủ ngữ đứng đầu câu (Luật 1), ở đây là giữ vị trí  - Động từ ở ngay sau chủ ngữ (Luật 1), ở đây là giữ vị trí  - Trạng từ đứng ngay trước tính từ (Do là trạng từ mức độ) (Tri thức số 3), ở đây là giữ vị trí (do đã đứng trước tính từ số 6)  - Tính từ theo trí thức số 2, do không có Danh từ/Đại từ trong phần vị ngữ nên giữ nguyên vị trí |
| 9 | “I”:”Đại từ, 1”, “and”:”Liên từ, 2”, “Ngoc”:”Danh từ riêng, 3”, “are running”:”Động từ, 4”, “very”:”Trạng từ chỉ mức độ-Tính từ-6, 5”, “fast”:”Tính từ, 6”, “Động từ chính”:”4”, “Thì”:”Hiện tại tiếp diễn” | I and Ngoc are running very fast | Ghép các từ theo thứ tự đã chuyển đổi. |

* 1. Kịch bản số 2

Mọi từ đều có trong từ điển, nhưng không tìm được loại từ theo luật kết nối của bên Tiếng Anh.

Người dùng nhập vào câu văn:

“anh ấy có lẽ thích cô ấy”

Tách các từ ra ta được:

“anh ấy”, “có lẽ”, “thích”, “cô ấy”

|  |  |  |  |
| --- | --- | --- | --- |
| Bước | Dữ liệu vào | Dữ liệu ra | Các luật/tập dữ liệu được áp dụng |
| 1 | “anh ấy”, có lẽ”, “thích”, “cô ấy” | “anh ấy”:”Đại từ xưng hô, 1”, “có lẽ”:”Phó từ chỉ mức độ, 2”, “thích”:”Động từ, 3”, “cô ấy”:”Đại từ xưng hô, 4”, “Động từ chính”:”3” | Tập dữ liệu tiếng Việt  Luật bảng 2 và các tri thức tiếng Việt để tìm ra các loại từ có trong câu và vị trí của chúng  Đánh số vị trí của chúng trong câu tiếng Việt ban đầu  Tìm động từ chính dựa vào luật bảng 2 |
| 2 | “anh ấy”:”Đại từ xưng hô, 1”, “có lẽ”:”Phó từ chỉ mức độ, 2”, “thích”:”Động từ, 3”, “cô ấy”:”Đại từ xưng hô, 4”, “Động từ chính”:”3” | “anh ấy”:”Đại từ xưng hô, 1”, “có lẽ”:”Phó từ chỉ mức độ - Động từ - 3, 2”, “thích”:”Động từ, 3”, “cô ấy”:”Đại từ xưng hô, 4”, “Động từ chính”:”3”, “Thì”:”Hiện tại đơn” | Tri thức bảng 1 để tìm ra thì của câu  Tìm chủ từ của các Phó từ còn lại theo Luật số 7, bảng 4 |
| 3 | “anh ấy”:”Đại từ xưng hô, 1”, “có lẽ”:”Phó từ chỉ mức độ - Động từ - 3, 2”, “thích”:”Động từ, 3”, “cô ấy”:”Đại từ xưng hô, 4”, “Động từ chính”:”3”, “Thì”:”Hiện tại đơn” | “anh ấy”:”Đại từ, 1”, “có lẽ”:”Trạng từ chỉ mức độ - Động từ - 3, 2”, “thích”:”Động từ, 3”, “cô ấy”:”Đại từ đối tượng, 4”, “Động từ chính”:”3”, “Thì”:”Hiện tại đơn” | Luật bảng 4 để tìm ra sự ánh xạ của các từ tiếng Việt sang tiếng Anh  Chuyển đại từ theo Luật số 12 bảng 4 |
| 4 | “anh ấy”:”Đại từ, 1”, “có lẽ”:”Trạng từ chỉ mức độ - Động từ - 3, 2”, “thích”:”Động từ, 3”, “cô ấy”:”Đại từ đối tượng, 4”, “Động từ chính”:”3”, “Thì”:”Hiện tại đơn” | “He”:”Đại từ, 1”, “probably”:”Trạng từ chỉ mức độ - Động từ - 3, 2”, “like”:”Động từ, 3”, “her”:”Đại từ đối tượng, 4”, “Động từ chính”:”3”, “Thì”:”Hiện tại đơn” | Dùng bảng dữ liệu 1 để ánh xạ sang tiếng anh với loại từ tương ứng  Trong bảng dữ liệu chỉ tồn tại tên và từ dịch  {  "word": "có lẽ",  "trans0": [  "probably",  "maybe"  ]  },  Không tồn tại kiểu loại nên sẽ lấy từ đầu tiên xuất hiện |
| 5 | “He”:”Đại từ, 1”, “probably”:”Trạng từ chỉ mức độ - Động từ - 3, 2”, “like”:”Động từ, 3”, “her”:”Đại từ đối tượng, 4”, “Động từ chính”:”3”, “Thì”:”Hiện tại đơn” | “He”:”Đại từ, 1”, “probably”:”Trạng từ chỉ mức độ - Động từ - 3, 2”, “likes”:”Động từ, 3”, “her”:”Đại từ đối tượng, 4”, “Động từ chính”:”3”, “Thì”:”Hiện tại đơn” | Dùng bảng dữ liệu 2 để ánh xạ sang đúng kiểu của thì (kiểu V-inf)  Dùng bảng dữ liệu 3 và 4 để tìm ngôi cho đại từ  Dùng luật chia động từ theo ngôi của bảng 3 |
| 6 | “He”:”Đại từ, 1”, “probably”:”Trạng từ chỉ mức độ - Động từ - 3, 2”, “likes”:”Động từ, 3”, “her”:”Đại từ đối tượng, 4”, “Động từ chính”:”3”, “Thì”:”Hiện tại đơn” | “He”:”Đại từ, 1”, “probably”:”Trạng từ chỉ mức độ - Động từ - 3, 2”, “likes”:”Động từ, 3”, “her”:”Đại từ đối tượng, 4”, “Động từ chính”:”3”, “Thì”:”Hiện tại đơn”” | Dùng các tri thức có trong bảng 5 để sắp xếp các loại từ:  - Chủ ngữ đứng đầu câu (tri thức 1), ở đây là giữ vị trí  - Động từ ở ngay sau chủ ngữ (tri thức 1), ở đây là giữ vị trí  - Trạng từ đứng ngay trước động từ (Do là trạng từ mức độ) (tri thức 3), ở đây là giữ vị trí (do đã đứng trước động từ số 3) |
| 7 | “He”:”Đại từ, 1”, “probably”:”Trạng từ chỉ mức độ - Động từ - 3, 2”, “likes”:”Động từ, 3”, “her”:”Đại từ đối tượng, 4”, “Động từ chính”:”3”, “Thì”:”Hiện tại đơn”” | He probably likes her | Ghép các từ theo thứ tự đã chuyển đổi. |

* 1. Kịch bản số 3

Mọi từ đều có trong từ điển, nhưng không tìm được luật sắp xếp từ phù hợp (Trường hợp này chưa thể ví dụ ra được do nếu đã tìm được một luật thì sẽ cập nhật luôn)

Giải pháp: Tương tự như kịch bản 1 nhưng giữ nguyên vị trí trong câu tiếng Anh.

* 1. Kịch bản số 4

Có xuất hiện từ không tồn tại trong từ điển (Trường hợp này chưa thể ví dụ ra được do nếu đã tìm được một từ mới thì sẽ cập nhật luôn)

Giải pháp: Tương tự như kịch bản 1 trên nhưng giữ nguyên từ và vị trí.

# Công cụ lập trình & Code

Sử dụng ngôn ngữ lập trình Python

Trình soạn thảo: Visual Studio Code

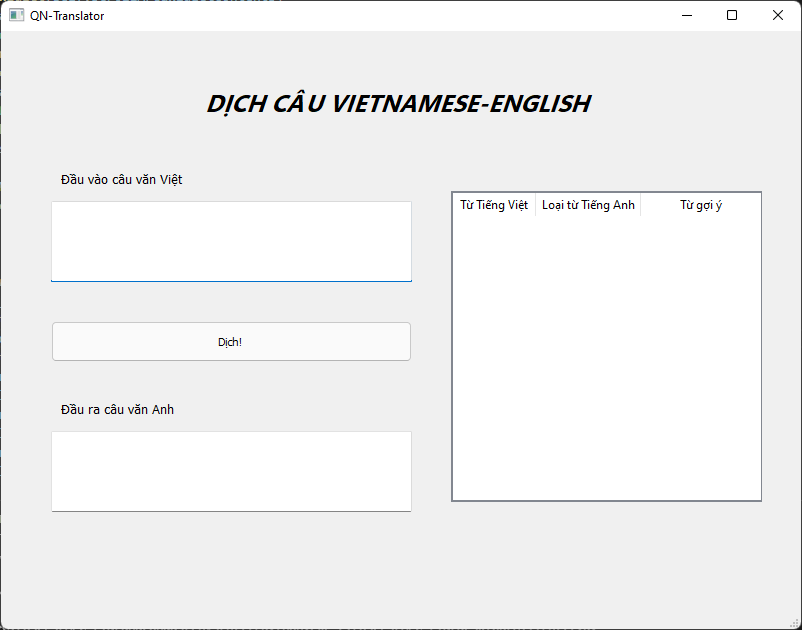
Quản lý môi trường: Conda/Anaconda

Các gói đi kèm:

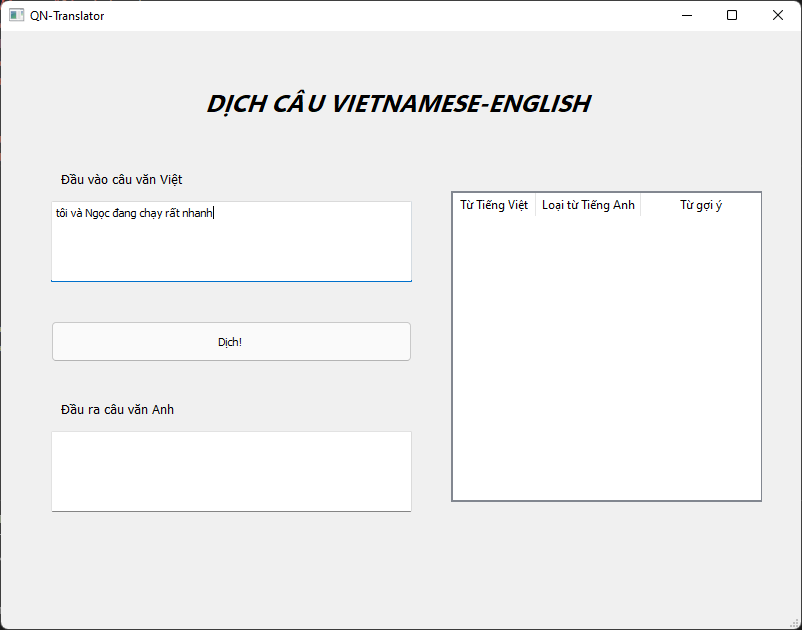
* JSON
* Underthesea
* Pandas
* Numpy
* PyQt5
* PyQt5-tools
* Source code & Data provided:

# Chạy hệ thống

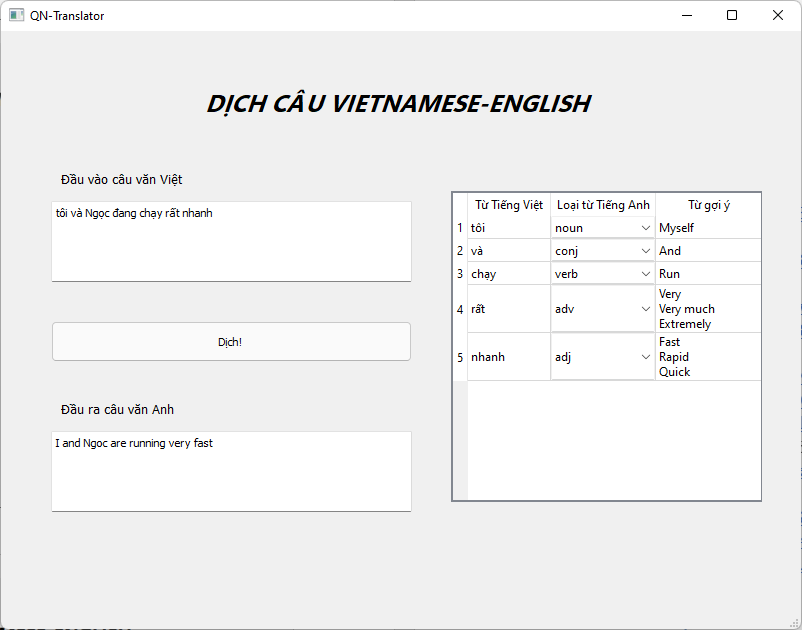
Chạy file ***httt.py*** để vào giao diện hệ thống



Nhập câu cần nhập vào ô “Đầu vào câu văn Việt”



Bấm nút dịch để nhận kết quả:



Nhận xét về kết quả:

* Câu đầu ra dịch được sẽ nằm ở ô “Đầu ra câu văn Anh”
* Bên cạnh là bảng gồm từ Tiếng Việt, mỗi từ lại có nhiều loại, và mỗi loại lại có nhiều từ đồng nghĩa.
* Bảng trên còn đóng vai trò như một nguồn thông tin thêm cho câu văn được dịch, giúp người dùng nhận biết chuẩn hơn

# Tài liệu tham khảo

## Tài liệu tiếng Việt

Ngữ văn 6 tập 1 <https://drive.google.com/file/d/1CeXcEgSGPTTb85TLMCYtljio1IE_MKx1/view>

Ngữ văn 6 tập 2

<https://drive.google.com/file/d/1G4Pv3729SOflVeQIvsJ3l-gSFE8Yz7nq/view>

Ngữ văn 7 tập 1 <https://drive.google.com/file/d/1erJRRw_mtejmilDb0wLdJSlq1xOJqRLh/view>

<https://vi.wikipedia.org/wiki/T%E1%BB%AB_lo%E1%BA%A1i>

Tập quy tắc cú pháp tiếng Việt

<http://www.jaist.ac.jp/~bao/VLSP-text/Mar2009/SP85_baocaokythuat2009thang3.pdf>

<https://link.springer.com/article/10.1007/s10579-017-9398-3>

<http://www.lrec-conf.org/proceedings/lrec2016/pdf/95_Paper.pdf>

Cách chia nhỏ từ loại Tiếng Việt có thể ánh xạ sang từ loại Tiếng Anh

<https://github.com/undertheseanlp/underthesea/blob/master/datasets/DI_Vietnamese-UVD/docs/2008%20Vu%20-%20Nghi%C3%AAn%20c%E1%BB%A9u%20v%C3%A0%20x%C3%A2y%20d%E1%BB%B1ng%20t%E1%BB%AB%20%C4%91i%E1%BB%83n%20ti%E1%BA%BFng%20Vi%E1%BB%87t%20cho%20m%C3%A1y%20t%C3%ADnh%20-%20ICT08-VLSP-SP72.pdf>

## Tài liệu tiếng Anh

Từ loại Tiếng Anh

<https://vinaenglish.com/ngu-phap-tieng-anh/>

<https://engbreaking.com/tu-loai-trong-tieng-anh>

Cách chia động từ “to be”

<https://elight.edu.vn/dong-tu-to-be/>

Cấu trúc ngữ pháp Tiếng Anh

<https://vinaenglish.com/ngu-phap-tieng-anh/>

<https://dictionary.cambridge.org/grammar/british-grammar/>

<https://tienganhmoingay.com/ngu-phap-tieng-anh/ngu-phap-tieng-anh-cot-loi-cau-truc-cau/>

Các thì trong Tiếng Anh

<https://topicanative.edu.vn/cac-thi-trong-tieng-anh/>

Chia động từ Tiếng Anh

<https://tienganhmoingay.com/ngu-phap-tieng-anh/tong-quan-hoa-hop-chu-ngu-dong-tu/>

Cách ánh xạ từ loại giữa Tiếng Việt và Tiếng Anh

<http://web.eecs.umich.edu/~mihalcea/wpt/papers/pdf/Dien.pdf>

<https://github.com/undertheseanlp/underthesea/wiki/M%C3%B4-t%E1%BA%A3-d%E1%BB%AF-li%E1%BB%87u-b%C3%A0i-to%C3%A1n-POS-Tag>

## Nguồn dữ liệu

Từ điển Tiếng Việt

<https://raw.githubusercontent.com/undertheseanlp/underthesea/master/datasets/DI_Vietnamese-UVD/corpus/underthesea_v170/underthesea_dictionary.yaml>

Từ điển loại từ Việt

<https://github.com/thanhphu/tudien>

Từ điển việt-anh bảng 1

<https://www.informatik.uni-leipzig.de/~duc/Dict/install.html>

Từ điển động từ bảng 2

<https://github.com/monolithpl/verb.forms.dictionary/tree/master/json>

Từ điển danh từ số ít - số nhiều bảng 3

<https://github.com/djstrong/nouns-with-plurals>

Từ điển đại từ ngôi ít, ngôi nhiều bảng 4

<https://github.com/sapics/pronoun-json>

## Công cụ

Underthesea dùng để tách câu thành các từ hoặc cụm từ có nghĩa

<https://github.com/undertheseanlp/underthesea>

JSON để cung cấp phương pháp truy cập vào hệ thống dữ liệu

Pandas để đọc file CSV dưới dạng từng cột

<https://github.com/pandas-dev/pandas>

Numpy để xử lý vector dữ liệu

<https://github.com/numpy/numpy>