

5. Soạn từ điển song ngữ

Khác với từ điển đơn ngữ, mục đích chính của từ điển song ngữ là cung cấp *từ ngôn ngữ đích* tương đương với *từ ngôn ngữ gốc*. Tính tương đương cần thỏa nhiều phạm trù: ngữ nghĩa, từ vựng, ngữ pháp, ...

Lấy ví dụ chữ "*anh*". Nếu dịch chữ này sang tiếng Anh, nhiều người sẽ liên tưởng ngay đến chữ "*you*", một tiếng xưng hô người đối diện. Thế nhưng nghĩ kỹ thì sẽ thấy phức tạp hơn nhiều.

Về **phạm trù ngữ nghĩa** (*semantic category*), "*anh*" không chỉ đồng nghĩa với "*you*" mà còn đồng nghĩa với "*I*", "*me*", "*he*", "*him*", "*man*", "*my*", "*your*", "*his*", "*elder brother*", "*elder cousin*".

Về **phạm trù từ vựng** (*lexical category*), "*anh*" là:

- một *nhân vật đại danh từ* (*personal pronoun*) tương đương với "*I*", "*you*", "*he*", "*one*", "*me*", "*him*".
- một *danh từ* (*noun*) tương đương với "*elder brother*", "*elder cousin*", "*man*".
- một *sở hữu tính từ* (*possessive adjective*) tương đương với "*my*", "*your*", "*his*".

Về **phạm trù ngữ pháp** (*grammar category*), "*anh*" tương đương với

- "*I*", "*you*", "*he*" trong vai trò *danh cách* (*nominative*), số ít.
- "*me*", "*you*", "*him*" trong vai trò *đôi cách* (*accusative*)/*tặng cách* (*dative*), số ít.

V.v.

Ví dụ mục từ "*anh*" được dịch sang tiếng Anh (theo Từ điển Anh-Việt Việt-Anh của Bùi Phụng^[1]):

- anh** 1. brother; elder brother; *anh ruột* blood brother; *anh họ* (frist) cousin;
2. you; *anh đi đâu thế?* where are you going?

Ngược lại là chiều Anh- Việt. Giả sử dịch từ "*you*" sang tiếng Việt. Tính tương đương cũng cần thỏa nhiều phạm trù giống ví dụ "*anh*" bên trên:

Về **phạm trù ngữ nghĩa** (*semantic category*), "*you*" không chỉ đồng nghĩa với "*anh*" mà còn đồng nghĩa với "*chị*", "*ông*", "*bà*", "*cô*", "*chú*", "*bác*", "*con*", "*em*", "*cháu*", ..., "*bạn*", "*mày*", "*bay*", "*mi*", "*cậu*", "*đấng ấy*", "*quý vị*", ..., "*các anh*", "*các chị*", "*các ông*", "*các bà*", "*các cô*", "*các chú*", "*các bác*", "*các con*", "*các em*", "*các cháu*", ..., "*các bạn*", "*bọn mày*", "*chúng mày*", "*tụi bay*", "*bọn mi*", "*các cậu*", "*mấy người*", Tiếng Việt có rất nhiều tiếng xưng hô tương đương với "*you*" cho từng trường hợp^[2].

Về **phạm trù từ vựng** (*lexical category*), "*you*" là một *nhân vật đại danh từ* (*personal pronoun*) tương đương với "*anh*", "*chị*", "*ông*", "*bà*", ... kể trên.

Về **phạm trù ngữ pháp** (*grammar category*), "*anh*" tương đương với "*anh*", "*chị*", "*ông*", "*bà*", ..., "*các anh*", "*các chị*", "*các ông*", "*các bà*" ... kể trên trong vai trò *danh cách* (*nominative*), *đôi cách* (*accusative*), *tặng cách* (*dative*), số ít lẫn số nhiều

Ví dụ mục từ "*you*" được dịch sang tiếng Anh (theo Từ điển Anh-Việt Việt-Anh của Bùi Phụng):

- you** [ju:] 1. ngôi thứ hai số ít hoặc số nhiều (bạn, các bạn, anh, các anh...). *You can swim*

fast anh có thể bơi nhanh. *Shall I get you a drink, John?* John, tôi lấy cho cậu đồ uống nhé.

2. mọi người nói chung. *It's not good for you to eat too much meat* ăn quá nhiều thịt thì không tốt. *You can't believe what politicians say* không nên tin những điều các chính trị gia nói. ^[3]

Những thông tin từ điển thường có giới hạn, chỉ nhằm cho biết ngữ nghĩa là chính. Tuy vậy cũng đủ dùng cho người tra cứu bình thường.

Vấn đề đặt ra ở đây – tương tự như phần soạn từ điển tiếng Việt đã đề cập - là làm sao có thể dựng từ điển song ngữ với độ phức tạp có thể tự định nghĩa, và xa hơn nữa sẽ dùng những thông tin đó cho mục đích dịch máy (sẽ nói ở phần sau).

5.1. Mô hình dịch thuật

5.1.1. Mô hình từ điển Việt-Anh

Tính đến nay, ta đã có kho dữ liệu từ vựng tiếng Việt, trong đó có phần dịch thuật sang tiếng Anh và tiếng Đức. Ví dụ mục từ "*anh*":

```
anh
...
LOCAL(1):
...
SUBCAT: personal pronoun
// phân dịch thuật
TRANS:
ENG:
...
LOCAL(3):
PERS: 3
SUBCAT: noun
// phân dịch thuật
TRANS:
ENG:
...
LOCAL(4):
SUBCAT: adjective
// phân dịch thuật
TRANS:
ENG:
...
```

Đó là cái nhìn toàn cảnh. Nay thử xem chi tiết. Xét một đoạn.

```
...
SUBCAT: personal pronoun
// phân dịch thuật
TRANS:
ENG:
  PERS == 1:
    CASE:
      NOM = "I";
      PRONUNC: → I.wav
      MAINNOTE: Phụ chú chính về phần NOM
      NOTE(1): Phụ chú phụ cho EXP(1)
      EXP(1): tiếng Việt / tiếng Anh
      ...
      NOTE(n): Phụ chú phụ cho EXP(n)
      EXP(n): tiếng Việt / tiếng Anh
    ACC = "me";
    PRONUNC: → me.wav
    MAINNOTE: Phụ chú chính về phần ACC
    NOTE(1): Phụ chú phụ cho EXP(1)
```

EXP(1): tiếng Việt / tiếng Anh
...

Lưu ý: MAINNOTE hoặc NOTE mang giá trị UNUSUAL có nghĩa là trường hợp không thông dụng. Các chọn lựa (option) này có thể dùng mọi chỗ cần ghi nhớ. MAINNOTE có ưu tiên cao hơn NOTE.

Cụ thể:

```
LOCAL(1):
  SUBCAT: personal pronoun
  TRANS:
    ENG:
      PERS == 1:
        CASE:
          NOM = "I";
          PRONUNC: → I.wav
          EXP(1): Anh đang về nhà. / I am going home.
          ACC = "me";
          PRONUNC: → me.wav
          EXP(1): Đưa anh cuốn sách. / Give me the buch.
      PERS == 2:
        CASE:
          NOM = "you";
          PRONUNC: → you.wav
          EXP(1): Anh đang làm gì? / What are you doing?
          ACC = "you";
          PRONUNC: → I.wav
          EXP(1): Tôi sẽ đến thăm anh. / I will visit you.
      PERS == 3:
        CASE:
          NOM = "he";
          PRONUNC: → he.wav
          MAINNOTE: UNUSUAL
          EXP (1): Anh (ấy) là sinh viên. / He is student.
          MORE → anh, anh ấy, anh ta
          ACC = "him";
          PRONUNC: → him.wav
          MAINNOTE: UNUSUAL
          EXP(1): Tôi biết anh (ta). / I know him.
          MORE → anh, anh ấy, anh ta

LOCAL(2):
  PERS: 3
  NUM: singular
  SEM:
    APPL(1): Người con trai trong gia đình lớn tuổi hơn các người con khác,
      // phân dịch thuật
    TRANS:
      ENG: "elder brother"
      PRONUNC: → elder_brother.wav
      EXP (1): Ông Nam là anh tôi. / Mr. Nam is my elder brother.
    APPL(2): hoặc trong họ hàng cùng thế hệ có vai vế cao hơn.
      // phân dịch thuật
    TRANS:
      ENG: "elder cousin"
      PRONUNC: → elder_cousin.wav
      EXP (1): Ông Bảo là anh họ tôi. / Mr. Bao is my elder cousin.

LOCAL(3):
  PERS: 3
  SEM: Người, người nam ngôi thứ 3.
  // phân dịch thuật
  TRANS:
    ENG:
      CASE: NOM = "one", "man";
```

EXP(1): Mỗi **anh** một việc. / One **man** one job.
MORE → người

LOCAL(4):

SUBCAT: adjective
VIETCAT: possessive adjective
PERS: 1, 2
GEN: true
SEM: Của anh (sở hữu từ dùng cho một người nam ngôi 1, 2, 3)
// phần dịch thuật
TRANS:
ENG:
PERS == 1:
CASE: NOM = "my";
PRONUNC: → my.wav
EXP(1): ...
PERS == 2:
CASE: NOM = "your";
PRONUNC: → your.wav
EXP(1): ...
PERS == 3:
CASE: NOM = "his";
PRONUNC: → his.wav
MAINNOTE: UNUSUAL
EXP(1): ...

5.1.1.1. Ví dụ từ điển Việt-Anh bản in

Tương tự từ điển Việt Nam, ta cũng có thể nhờ chương trình tự động lập từ điển Việt-Anh, ví dụ cho mục từ "anh". Độ phức tạp tùy theo lựa chọn.

Bản in 1 đơn giản nhất, không cần lựa chọn (trích phần LOCAL(1)):

anh 1. I, me, you, he, him 2. elder brother, elder cousin 3. one, man, 4. my, your, his

Bản in 1

Bản in 2 với hai lựa chọn CAT, SUBCAT:

anh 1. <pn>: I, me, you, he, him 2. <n>: elder brother, elder cousin 3. <n>: one, man, 4. <adj>: my, your, his

Bản in 2


(Ghi chú: pn = pronoun, n = noun, adj = adjective)

Bản in 3 cộng thêm lựa chọn EXP (minh họa phần 1):

anh
1. <pn>: I. **Anh** đang về nhà. / **I** am going home. me. Đưa **anh** cuốn sách. / Give **me** the book. you. **Anh** đang làm gì? / What are **you** doing? Tôi sẽ đến thăm **anh**. / I will visit **you**. he. **Anh** (ấy) là sinh viên. / **He** is student. him. Tôi biết **anh** (ta). / I know **him**.

Bản in 3


5.1.1.2. Ví dụ từ điển Việt-Anh bản điện tử

Cách thực hiện tự điển Việt-Anh điện tử cũng theo nguyên tắc từ điển tiếng Việt điện tử. Minh họa một phần theo bản in 3, cộng thêm lựa chọn PRONUNC (cách phát âm ):


anh 

1.

I . **Anh** đang về nhà. / **I** am going home.

me . Đưa **anh** cuốn sách. / Give **me** the buch.

you . **Anh** đang làm gì? / What are **you** doing? Tôi sẽ đến thăm **anh**. / I will visit **you**. d.

he . **Anh** (ấy) là sinh viên. / **He** is student.

him . Tôi biết **anh** (ta). / I know **him**.

5.1.2. Mô hình từ điển Anh-Việt

Cùng mô hình trên, muốn ứng dụng cho từ điển Anh-Việt, nghĩa là chiều ngược lại, phải làm hai công việc. Một là đổi mục từ "anh" (tiếng Việt) bằng một mục từ tiếng Anh tương ứng, ví dụ "you". Hai là thiết kế cho mục từ đó một cấu trúc với thuộc tính thích hợp.

Công việc thứ hai có hai giải pháp. Một là thêm cấu trúc mới vào bên cạnh cấu trúc cũ. Hai là tách rời.

Cách thứ hai tốt hơn vì là modul độc lập, dễ nhìn, dễ kiểm soát.

you

GLOBAL:

CAT: noun

PRONUNC: → you.wav

PERS: 2

LOCAL:

SUBCAT: personal pronoun

TRANS:

VIET:

NUM: singular

CASE:

NOM, ACC = "anh";

PRONUNC: → anh.wav

EXP(1): Where are **you** from? / **Anh** từ đâu đến?

EXP(2): See **you** again. / Hẹn gặp lại **anh**.

NOM, ACC = "chị";

PRONUNC: → chị.wav

EXP(1): How are **you**? / **Chị** khỏe không?

EXP(2): Happy birthday to **you**. / Chúc mừng sinh nhật **chị**.

...

NUM: plural

CASE:

NOM, ACC = "các anh";

PRONUNC: → cac_anh.wav

EXP(1): Who are **you**? / **Các anh** là ai?

EXP(2): I don't know **you**. / Tôi không quen **các anh**.

NOM, ACC = "các chị";

PRONUNC: → cac_chi.wav

EXP(1): What are **you** doing? / **Các chị** đang làm gì?

EXP(2): Can I help **you**? / Tôi có thể giúp **các chị**.

...

5.1.2.1 Ví dụ từ điển Anh-Việt bản in

Bản in 1 đơn giản nhất, không cần lựa chọn:

you anh, chị, ông, bà, cô, chú, bác, con, em, cháu, ..., bạn, mày, bay, mi, cậu, đấng ấy, quý vị, ..., các anh, các chị, các ông, các bà, các cô, các chú, các bác, các con, các em, các cháu, ..., các bạn, bọn mày, chúng mày, tụi bay, bọn mi, các cậu, mấy người, ...

Bản in 1




Bản in 2 với hai lựa chọn CAT, SUBCAT, EXP (minh họa một phần):

you <pn>: anh. *Where are you from? / Anh từ đâu đến? See you again. / Hẹn gặp lại anh.* chị. *How are you? / Chị khỏe không? Happy birthday to you. / Chúc mừng sinh nhật chị.*




Bản in 2

5.1.2.2 Ví dụ từ điển Anh-Việt bản điện tử

Minh họa một phần theo bản in 2:

you 
anh . *Where are you from? / Anh từ đâu đến? See you again. / Hẹn gặp lại anh.*
chị . *How are you? / Chị khỏe không? Happy birthday to you. / Chúc mừng sinh nhật chị.*

Bản điện tử có thể mở rộng tùy vào sự dồi dào của nó, ví dụ có thêm phần hyperlink (→) đến những từ liên hệ "tôi", "chị", "em", ...:

you 
anh . *Where are you from? / Anh từ đâu đến? See you again. / Hẹn gặp lại anh.*
→ *tôi, chị, em, ...*
chị . *How are you? / Chị khỏe không? Happy birthday to you. / Chúc mừng sinh nhật chị.*
→ *tôi, chị, em, ...*

(còn tiếp)

Chú thích:

[1] **Bùi Phụng** (2002) *Từ điển Anh-Việt Việt-Anh*. Hà Nội: Nxb Thế Giới.

[2] Sự phong phú của tiếng xưng hô tiếng là một ví dụ rất tốt cho thấy vấn đề của dịch máy: ngữ nghĩa sẽ nhập nhằng nếu kho từ vựng thiếu thông tin, vì máy không phân biệt được tiếng xưng hô nào dùng cho hoàn cảnh nào.

^[3] Thực ra phần 2 dùng cho trường hợp chung giống như tiếng xưng hô "*quý vị*" của tiếng Việt.