**BÁO CÁO KĨ THUẬT  
NGHIÊN CỨU XÂY DỰNG BỘ NHÃN VAI NGHĨA TIẾNG VIỆT**

Báo cáo này gồm 2 phần chính. Phần thứ nhất trình bày khảo sát về kho ngữ liệu có gán nhãn vai nghĩa cho tiếng Anh (English PropBank), trong đó nêu các nguyên tắc chú giải vai nghĩa và trình bày lại hướng dẫn chú giải vai nghĩa cho tiếng Anh được xây dựng trong dự án PropBank[[1]](#footnote-0). Trên cơ sở khảo sát kho ngữ liệu English PropBank, trong phần thứ 2 chúng tôi sẽ trình bày về việc xây dựng kho ngữ liệu có gán nhãn vai nghĩa cho tiếng Việt, bao gồm nguyên tắc xây dựng bộ vai nghĩa và hướng dẫn chú giải vai nghĩa cho tiếng Việt.

# Kho ngữ liệu gán nhãn vai nghĩa tiếng Anh (English PropBank)

## Các nguyên tắc chú giải động từ

* Tất cả các từ trong TreeBank có nhãn từ loại là động từ đều được chú giải tập vai nghĩa, ngoại trừ các động từ không bao giờ được sử dụng dưới dạng chủ động (ví dụ *collonaded* trong *collonaded house*) hoặc các động từ đóng vai trò giới từ (như based trong based on current research, given trong given the situation, accord trong according to our sources).
* Với các động từ đi kèm với tiểu từ thành một cấu trúc mang nghĩa xác định, cả cụm động từ + tiểu từ được coi là một quan hệ cần được gán nhãn vai nghĩa (ví dụ keep up).
* Mỗi động từ được chú giải đều có tham chiếu tới nghĩa tương ứng trong bộ ngữ liệu khung vị từ của động từ đó.
* Các vai nghĩa được chú giải cho mỗi động từ bao gồm: Arg0 (agent), Arg1 (patient), Arg2 (instrument, benefactive, attribute), Arg3 (starting point, benefactive, attribute), Arg4 (ending point), ArgM (modifier).
* Với các tham thể rỗng có đồng sở chỉ với một NP trong cùng câu thì được mô tả bằng một chuỗi đồng sở chỉ:
  + I made a decision [\*PRO\*] to leave
    - Rel: leave; Arg0: [\*PRO\*] \* I
* Với các động từ trong câu có 2 cấu trúc tham thể riêng biệt, dùng thêm nhãn DP (duplicate) để đánh dấu cho cấu trúc thứ 2. Ví dụ trong câu: I ate a sandwich and Cindy a banana.

## Arg0 và Arg1

* Arg0:
  + Tham dự có chủ ý vào một sự kiện hay một trạng thái
  + Gây ra 1 sự kiện hay thay đổi trạng thái của một tham thể khác
  + Chuyển động tương đối tới vị trí của một tham thể khác
* Arg1:
  + Thực hiện thay đổi trạng thái

* + Bị tác động với nguyên nhân từ một tham thể khác
  + Tĩnh so với chuyển động của một tham thể khác
* Với một số động từ ngoại động, có tham thể có thể gán nhãn Arg0 hoặc Arg1 đều được, khi đó nguyên tắc là gán nhãn có chỉ số thấp hơn.
* Tóm lại, Arg0 là các tham thể gây ra hành động tương ứng với động từ, có thể là tác nhân hay không, cũng như các tham thể mà truyền thống gọi là kẻ trải nghiệm (của các động từ trạng thái tình cảm như yêu, ghét). Arg1 là các tham thể bị thay đổi do các nguyên nhân bên ngoài, cũng như các tham thể dạng bị thể.

## Arg-A (Secondary Agent)

Tham thể tác nhân thứ hai dùng để chú giải thực thể cũng đóng vai trò tác nhân trong khi đã có một tác nhân khác gây ra sự kiện, ví dụ:

John walked his dog. (ARG-A: John, REL: walked, ARG-0: his dog)

## Chú giải các thành phần bổ nghĩa (modifiers) ArgM-

### COM: comitative

Thực thể cùng hành động với tác nhân, trừ các đối tượng đóng vai trò công cụ hành động

### LOC: Locative

Chú giải vị trí hành động được thực hiện. Có thể là vị trí vật lí hoặc vị trí trừu tượng (ví dụ: in his speech).

### DIR: Directional

Thành phần phụ chỉ hướng của hành động. Ví dụ: forward trong step forward, back trong snap back.

### GOL: Goal

Mục tiêu của hành động. Bao gồm đích đến của động từ chuyển động và các tham thể hưởng lợi hay các thành phần chỉ ra hành động của động từ được làm cho ai, cho cái gì. Ví dụ: The child fed the cat (for her mother)-ArgM-GOL. The walls crumbled (to the ground)-ArgM-GOL.

### MNR: Manner

Phó từ xác định cách thức hành động. Ví dụ well trong “work well”.

### TMP: Temporal

Thành phần phụ xác định thời gian xảy ra hành động, ví dụ “in 1987”, “last Wednesday”, “soon”, “immediately”, “often”, “always”, ngoại trừ “never” có nhãn phủ định NEG.

### EXT: Extent

Thành phần phụ xác định lượng thay đổi xảy ra do một hành động, thường là các định ngữ số lượng – ví dụ “by 15%” trong “raised prices by 15%”, lượng từ như “a lot”, từ so sánh như “more” trong “he raised prices more than she did”.

### REC: Reciprocals

Thành phần phản thân (ví dụ himself).

### PRD: Secondary Predication

Thành phần bổ ngữ của một vị từ mà chính nó có một cấu trúc vị từ. Ví dụ:

* Kết quả hành động: The boys pinched them DEAD; She kicked the locker lid SHUT.
* Mô tả: ROSY-CHEEKED, Santa came down the chimney.
* Ngữ bắt đầu bằng “as”: Pierre will join the board AS A NONEXECUTIVE DIRECTOR.

*Lưu ý thành phần này bổ nghĩa cho 1 tham thể của động từ (mô tả trạng thái của tham thể trong quá trình hoặc sau hành động) hơn là bổ nghĩa cho chính động từ đó.*

### PRP: Purpose

Thành phần mô tả động cơ của hành động.

### CAU: Cause

Thành phần mô tả nguyên nhân của hành động.

### DIS: Discourse

Bao gồm thành phần kết nối một câu với câu trước đó và các thán từ, từ đệm, lời gọi. Nguyên tắc đầu tiên để xác định một thành phần được gán nhãn này là nếu lược bỏ thành phần đó thì nghĩa của câu không đổi.

Lưu ý các từ nối 2 mệnh đề trong 1 câu chỉ được gán nhãn là ArgM-ADV. Ví dụ:

* The notification also clarifies the requirements of the evaluation. (Rel: clarifies, Arg0: The notification, Arg1: the requirements of the evaluation, **ArgM-DIS: also**)
* The notification recognizes the company and also clarifies the requirements of the evaluation. (Rel: clarifies, Arg0: The notification, Arg1: the requirements of the evaluation, **ArgM-ADV: also**)

Các ví dụ khác về ArgM-DIS:

* Vince trong “I aint kidding you, Vince”
* Ah, oh my God, damn, ...

Một ví dụ chú giải đầy đủ:

[**Your inability to report to my office this morning**]-Arg0 **has** [**not**]-ArgM-NEG [**ah**]-ArgM-DIS [**limited**]-Rel [**my knowledge of your activities**]-Arg1 [**as you may have hoped**]-ArgM-ADV.

### MOD: Modal

Các từ tình thái: will, may, can, must, shall, might, should, could, would.

### NEG: Negation

Các thành phần phủ định như “not”, “never”, “no longer”.

### DSP: Direct Speech

Dùng để chú giải một phát ngôn cho các động từ biểu đạt như nói, nghĩ, v.v., trong trường hợp các phát ngôn này ngắt quãng. Ví dụ về các trường hợp có và không dùng nhãn Arg-DSP:

Câu trong TreeBank:

(S

(PP (IN Among)

(NP (JJ other) (NNS things) ))

(PRN

(, ,)

(S

(NP-SBJ (PRP they) )

(VP (VBD said)

(SBAR (-NONE- 0)

(S (-NONE- \*?\*) ))))

(, ,) )

(NP-SBJ (NNP Mr.) (NNP Azoff) )

(VP (MD would)

(VP (VB develop)

(NP

(NP (JJ musical) (NNS acts) )

(PP (IN for)

(NP (DT a) (JJ new) (NN record) (NN label )))))

(. .) )

PropBank:

ARG1: [SBAR (-NONE- 0) S (-NONE- \*?\*)]

ARG0: they

ARGM-DSP: [-NONE-\*?\*] \* [Among other things , they said [\*?\*] , Mr. Azoff would develop musical acts for a new record label]

rel: said

Câu trong TreeBank:

(S ‘‘

(S-TPC-1 (NP-SBJ We)

(VP will

(VP win)))

,

’’

(NP-SBJ Mary)

(VP said

(S \*T\*-1))

.))

PropBank:

REL: said

ARG1: [\*T\*-1]

ARG0: Mary

### ADV: Adverbials

Các thành phần cú pháp làm biến đổi cấu trúc sự kiện của động từ, nhưng không thuộc bất kì loại phụ ngữ nào kể trên.

## Phạm vi chú giải

Các tham thể của một động từ cần tuân theo cấu trúc cú pháp của động từ đó

Không chú giải cho các quán từ hay liên từ, trừ khi chúng nằm ở vị trí đầu câu và có chức năng diễn ngôn. Không gán nhãn cho động từ phụ trợ.

Nguyên tắc xác định phạm vi chú giải trên cây cú pháp: Chú giải cho tất cả các thành phần cú pháp thuộc nút anh chị của nút động từ đang xét là quan hệ cần chú giải vai nghĩa cũng như thuộc nút anh chị của nút VP chứa động từ đó.

## Hiểu và chú giải cho các thành phần rỗng trong Penn TreeBank

Trong Penn TreeBank có các thành phần rỗng sau đây:

[\*PRO\*] đánh dấu chủ ngữ thiếu

[\*] lưu vết thành phần trong cấu trúc bị động, bao gồm cả các mệnh đề quan hệ rút gọn và các cấu trúc đề bạt

[\*T\*] Lưu vết tham thể bị dịch chuyển, gồm cả các thành phần trống phụ thuộc vào một thành phần bị lược bỏ khác. Ví dụ: **What book** did you review \_\_ without actually reading \_\_.

[(NP \*)] (arbitrary PRO, controlled PRO, and trace of A-movement)

[0] (null complementizer, including null wh-operator)

[\*U\*] (unit)

[\*?\*] Vị trí cho thành phần tỉnh lược

[\*NOT\*] (anti-placeholder in template gapping)

[\*RNR\*] (pseudo-attach: right node raising)

[\*ICH\*] (pseudo-attach: interpret constituent here)

[\*EXP\*] (pseudo-attach: extraposition)

### Chú giải câu bị động

Chú giải trong Penn TreeBank:

John-1 was hit [\*-1] by Mary.

PropBank: Rel: hit, Arg1: [\*-1], Arg0: by Mary

Penn TreeBank:

*(S (NP-SBJ-1 he)*

*(VP was*

*(VP accused*

*(NP-3 \*-1)*

*(PP-CLR of*

*(S-NOM (NP-SBJ \*-3)*

*(VP (VP conducting*

*(NP illegal business))*

*and*

*(VP possessing*

*(NP illegal materials))))))))*

PropBank:

*ARG1: [NP-3 \*-1]*

*REL: accused*

*ARG2: of [\*3\*] conducting illegal business and possessing illegal materials*

### Chú giải các tham thể đưa lên trước và dịch vị trí

Ví dụ 1: trường hợp tham thể vị trí của động từ “put” được đưa lên trước: There, I put the book.

TreeBank:

(S (ADVP-PUT-TPC-1 There)

,

(NP-SBJ I)

(VP put

(NP the book)

(ADVP-PUT \*T\*-1) ))

PropBank: Rel: put, Arg0: I, Arg1: the book, Arg2: [\*T\*-1]

Ví dụ 2: Thành phần phụ ArgM được dịch chuyển lên trước.

TreeBank:

*(S (SBAR-PRP-TPC-9 Because*

*(S (NP-SBJ I)*

*(VP ’m*

*(NP-PRD such a bad boy))))*

*(NP-SBJ I)*

*(VP think*

*(SBAR 0*

*(S (NP-SBJ I)*

*(VP wo n’t*

*(VP get*

*(NP a lollipop)*

*(SBAR-PRP \*T\*-9) )))))*

Chú giải PropBank:

*REL: get*

*ARG1: a lollipop*

*ARG0: I*

*ARGM-NEG: n’t*

*ARGM-MOD: wo*

*ARGM-CAU: [\*T\*-9]*

Ví dụ 3: Bổ ngữ là đại từ thay cho đối tượng được đưa lên phần đề của câu.

Chú giải TreeBank:

*(S (NP-TPC-1 John)*

*,*

*(NP-SBJ I)*

*(VP like*

*(NP (NP him)*

*(NP-1 \*T\*))*

*(NP-ADV a lot)))*

Chú giải PropBank:

*REL: like*

*ARG0: I*

*ARG1: [NP (NP him) (NP-1 \*T\*)]*

*ARGM-MNR: a lot*

### Câu hỏi và cụm Wh

Ví dụ 1: What do you like?

Chú giải TreeBank:

*(SBARQ (WHNP-1 what)*

*(SQ do*

*(NP-SBJ you)*

*(VP like*

*(NP \*T\*-1)))*

Chú giải PropBank:

*REL: like*

*ARG0: you*

*ARG1: [\*T\*-1]*

Ví dụ 2: Which day did you get there?

Chú giải TreeBank:

*SBARQ (WHNP-1 Which day)*

*(SQ did*

*(NP-SBJ you)*

*(VP get*

*(ADVP-DIR there)*

*(NP-TMP \*T\*-1)))*

Chú giải PropBank:

*ARG0: you*

*REL: get*

*ARG2: there*

*ARGM-TMP: [\*T\*-1]*

Ví dụ 3: How did you fix the car?

Chú giải TreeBank:

*(SBARQ (WHADVP-42 How)*

*(SQ did*

*(NP-SBJ you)*

*(VP fix*

*(NP the car)*

*(ADVP-MNR \*T\*-42)))*

*?)*

Chú giải PropBank:

*REL: ﬁx*

*ARG0: you*

*ARG1: the car*

*ARGM-MNR: [\*T\*-42]*

Ví dụ 4: *John didn’t know where-3 his parents had met [\*T\*-3].*

PropBank: *REL: met, ARG0: his parents, ARGM-LOC: [\*T\*-3]*

### Chú giải vết ICH

Lưu vết ICH trong TreeBank dùng để chú giải quan hệ giữa các thành phần bị phân tách bởi một thành can thiệp khác.

Ví dụ 1: A young woman entered whom she at once recognized as Jemima Broadwood.

Chú giải TreeBank:

*(S (NP-SBJ (NP a young woman)*

*(SBAR \*ICH\*-1))*

*(VP entered*

*(SBAR-1 (WHNP-2 whom)*

*(S (NP-SBJ she)*

*(PP-TMP at*

*(ADVP once))*

*(VP recognized*

*(NP \*T\*-2)*

*(PP-CLR as*

*(NP Jemima Broadwood)))))))*

Chú giải PropBank:

ARG0: a young woman [\*ICH\*-1]

REL: entered

Ví dụ khác:

*[Five \*ICH\*-1] ran, [out of the twenty-five that showed up]-1.*

ARG0: Five \*ICH-1\*

REL: ran

*[Some people in Paris]-1 want \*PRO\*-1 to hear more [\*ICH\*-2] from me [than those fellers*

*over at the conference house do]-2.*

ARG0: \*PRO\*-1

REL: hear

ARG1: more [ICH-2]

ARG2: from me

### Chú giải cho cấu trúc Right Node Raising

Ví dụ 1:

Chú giải TreeBank:

*(NP (NP (ADJP so many) enchained demons)*

*(VP straining*

*(PP-MNR in*

*(NP anger))*

*(S (NP-SBJ \*)*

*(VP to*

*(VP (VP tear*

*(NP \*RNR\*-1))*

*and*

*(VP gnaw*

*(PP-CLR on*

*(NP \*RNR\*-1)))*

*(NP-1 his bones))))))*

Trong ví dụ này, cụm danh từ “*his bones*” được hiểu là tham thể của cả động từ “*tear*” và động từ “*gnaw*”. Khi chú giải động từ “*tear*”, vết (NP \*RNR\*-1) là tham thể của động từ.

Chú giải PropBank:

*REL: tear; ARG1: [\*RNR\*-1]; ARG0: [NP-SBJ\*]*

*REL: gnaw; ARG1: on [\*RNR\*-1]; ARG0: [NP-SBJ\*]*

Ví dụ 2:

*I want \*RNR\*-1 and like \*RNR\*-1 [\* to eat ice-cream]-1.*

*ARG0: I; REL: want; ARG1: \*RNR\*-1*

Ví dụ 3:

*His dreams had revolved around her so much and for so long that…*

Chú giải TreeBank:

*(S (NP-SBJ His dreams)*

*(VP had*

*(VP revolved*

*(PP-CLR around*

*(NP her))*

*(UCP-ADV (ADVP (ADVP so much)*

*(SBAR \*RNR\*-1))*

*and*

*(PP-TMP for*

*(NP (NP so long)*

*(SBAR \*RNR\*-1)))*

*(SBAR-1 that...)))))*

Chú giải PropBank:

*ARG1: his dreams; REL: revolved; ARGM-LOC: around her; ARGM-EXT: so much [\*RNR\*]*

Ví dụ 4:

*But our outlook has been and continues to be defensive*

Chú giải TreeBank:

(S But

(NP-SBJ-2 our outlook)

(VP (VP has

(VP been

(ADJP-PRD \*RNR\*-1)))

,

and

(VP continues

(S (NP-SBJ \*-2)

(VP to

(VP be

(ADJP-PRD \*RNR\*-1)))))

,

(ADJP-PRD-1 defensive)))

Chú giải PropBank:

REL: continue

ARG1: [\*-2] to be \*RNR-1

### \*EXP\*

Các vị trí giả trong tiếng Anh như “it” hay “that” không thêm nghĩa cho câu. Trong ví dụ sau, chủ ngữ cú pháp của câu là từ “it”, chứa vết EXP-1 chiếu tới chủ ngữ ngữ nghĩa và logic của câu, được đánh dấu là SBAR-1.

Chú giải TreeBank:

(S (NP-SBJ (NP It)

(SBAR \*EXP\*-1))

(VP is

(ADJP-PRD clear)

(PP to

(NP me))

(SBAR-1 that

(S (NP-SBJ this message)

(VP is

(ADJP-PRD unclear))))))

Trong chú giải PropBank không chứa từ giả “it” và các vết EXP, không được gán nhãn chúng.

Chú giải PropBank:

REL: is

ARG1: that this message is unclear

ARG2: clear to me

Thay vào đó, chỉ gán nhãn các thành phần có nghĩa trong lời nói, thành phần mở. Do đó, nghĩa câu bên dưới có thể hiểu như sau: “*That this message is unclear is clear to me.*”. Từ ‘it’ chỉ đơn thuần được thêm vào đối với mục đích ngữ dụng để tránh việc có một thành phần dài như vậy ở phía trước cụm từ.

Ví dụ khác:

*It required an energy he no longer possessed to be satirical about his father.*

Chú giải PropBank:

ARG0: to be satirical about his father

ARG1: an energy he no longer possessed

REL: required

Trong các ví dụ dưới đây, các thành phần giả là bổ ngữ thay vì chủ ngữ. Như trong trường hợp của các chủ ngữ giả, chỉ các tham thể logic được gán nhãn, trong khi đại từ giả và các vết EXP không nằm trong các chú giải PropBank.

*Mrs. Yeargin was ﬁred [\*-1] and prosecuted [\*-1] under an unusual South Carolina law that-79 [\*T\*-79] makes it [\*EXP\*-2] a crime [\*] to breach test security.*

Chú giải PropBank:

ARG0: [\*T\*-79]

REL: makes

ARG2: a crime

ARG1: [\*] to breach test security

*Any raider would ﬁnd it [\*EXP\*-1] hard [\*] to crack AG ’s battlements*.

Chú giải TreeBank:

(S

(NP-SBJ (DT Any) (NN raider) )

(VP (MD would)

(VP (VB find)

(S

(NP-SBJ

(NP (PRP it) )

(S (-NONE- \*EXP\*-1) ))

(ADJP-PRD (JJ hard) )

(S-1

(NP-SBJ (-NONE- \*) )

(VP (TO to)

(VP (VB crack)

(NP

(NP (NNP AG) (POS ’s) )

(NNS battlements) )))))))

(. .) )

Chú giải PropBank:

ARG0: Any raider

ARGM-MOD: would

REL: ﬁnd

ARG3: hard

Arg1: [\*] to crack AG ’s battlements

Sai lầm phổ biến: Hãy chắc chắn để phân biệt từ giả “it” với đại từ “it” tham chiếu tới một cụm NP trước đó, một mệnh đề, hoặc một sự kiện. (Gợi ý: Đại từ thay thế không có vết EXP theo sau trong TreeBank). Mọi đại từ, bao gồm cả “it” sẽ được đánh dấu là tham thể trong PropBank.

*It sounds good.*

REL: sounds

ARG1: it

ARGM-MNR: good

*Italy ’s Foreign Ministry said [0] it is investigating exports to the Soviet Union*.

REL: investigating

ARG0: it

ARG1: exports to the Soviet Union

### Các vết khác

Chú giải TreeBank:

*(S (NP-SBJ I)*

*(VP believe*

*(SBAR 0*

*(S (NP-SBJ you)*

*(VP are*

*(ADJP-PRD smart))))))*

Chú giải PropBank:

*REL: believe*

*ARG0: I*

*ARG1: [[0] you are smart]*

## Kết nối và chú giải các phần tử rỗng

### Liên kết ràng buộc lựa chọn (liên kết SLC – SeLectional Constraint)

**Chú giải mệnh đề quan hệ**

**Các mệnh đề quan hệ điển hình** Các mệnh đề quan hệ là các mệnh đề bổ nghĩa cho một danh từ hay danh ngữ như trong “answers that we’d like to have”. Các mệnh đề quan hệ cũng bao gồm một vết được gắn cùng chỉ số với từ quan hệ trong TreeBank (‘that’, ‘which’, ‘who’). Ngoài ra, từ quan hệ có thể bị lược bỏ trong tiếng Anh: “answers we’d like to have.” Trong các trường hợp đó TreeBank vẫn chứa 1 vị trí cho từ quan hệ, nhưng giá trị ‘0’ sẽ xuất hiện tại vị trí mà từ quan hệ bình thường phải xuất hiện.

Ví dụ:

Chú giải TreeBank:

*(NP (NP answers)*

*(SBAR (WHNP-6 that/which/0)}*

*(S (NP-SBJ-3 we)*

*(VP ’d*

*(VP like*

*(S (NP-SBJ \*-3)*

*(VP to*

*(VP have*

*(NP \*T\*-6)))))))))*

Trong khi về mặt cú pháp vết được đánh chỉ số cùng với từ quan hệ, về mặt ngữ nghĩa lại có quan hệ giữa vết và cụm NP “answers” lại không được biểu diễn trong TreeBank. Quan hệ này hiện tại được nắm bắt thông qua hậu xử lí, nên người chú giải không cần tạo liên kết ở đây, nhưng cần nhận thức được làm thế nào nhận biết và hiểu các quan hệ giữa các thành phần trong một cấu trúc mệnh đề quan hệ.

### Liên kết đồng sở chỉ ngữ dụng (liên kết PCR – Pragmatic Coreference)

Liên kết PCR được chú giải bằng chức năng “\*”. Sau đó, trong quá trình hậu xử lý, chức năng này được chuyển đổi sang nhãn Link-PCR. Các liên kết quan hệ thu gọn được thâu tóm chặt chẽ trong khâu hậu xử lý, những cũng được chuyển đối thành cùng một loại liên kết.

**Chú giải quan hệ rút gọn**

Một mệnh đề quan hệ có thể được rút gọn ở thể bị động, dẫn đến cú pháp duy nhất của một mệnh đề quan hệ rút gọn. Ví dụ, một cấu trúc mệnh đề quan hệ rút gọn như “‘*The woman that was dressed in blue walked past the house*” có thể rút gọn thành “*The woman dressed in blue walked past the house*”. Do động từ trong trường hợp này luôn là bị đông, chú giải Treebank của quan hệ rút gọn sẽ chứa một vết đối tượng sau động từ. Tuy nhiên, không giống như cấu trúc bị động thông thường, vết này sẽ không bao giờ được đánh cùng chỉ số với chủ ngữ. Người chú giải có thể chỉ đơn giản gán nhãn vết là Arg1 và dựa vào phần hậu xử lý để nắm bắt mối quan hệ giữa vết và chủ ngữ.

*Chú giải TreeBank:*

(S

(NP-SBJ-1

(DT This)

(VP

(VBZ is)

(VP

(VBN considered)

(S

(NP-SBJ

(-NONE- \*-1)

(NP-PRD

(NP

(CD one)

(PP

(IN of)

(NP

(NP

(DT the)

(JJS biggest)

(NNS caches)

(VP

(VBN seized)

(NP

(-NONE- \*)

(PP-LOC

(IN in)

(NP

(DT the)

(NN district)

*Chú giải PropBank:*

*REL: seized*

*ARG1: [NP -NONE- \*]*

*ARGM-LOC: [PP-LOC in the district]*

**Chú giải của \*PRO\***

Nhiều vết có trong TreeBank phát sinh do việc dịch chuyển một thành phần khỏi vị trí thông thường của nó. Sự dịch chuyển này để lại một vết, biểu diễn bằng \* hoặc \*T\* trong TreeBank. Trong khi đó, \*PRO\* không phải là kết quả của việc dịch chuyển. Thay vào đó, \*PRO\* xuất hiện khi chủ ngữ của một động từ là dưới đặc tả hoặc không xuất hiện. Ví dụ, chủ ngữ c` ủa động từ “*leave*” trong câu “*she tried to leave*” không hiện thực hoá. Tuy nhiên, TreeBank sẽ biểu diễn chủ ngữ không hiện thực hoá của động từ “*leave*” bằng \*PRO\*.

*She-1 tried \*PRO\*-1 to leave*

*REL: leave*

*ARG0: \*PRO\*-1*

Trong trường hợp giống như ví dụ trên, phần tử \*PRO\* được đánh đồng chỉ số với chủ ngữ xác định đầy đủ bởi vì \*PRO\* được đặt trong một mệnh đề bị chi phối bởi mệnh đề cao hơn với cùng một chủ ngữ “*she*”. Như vậy, người chú giải không cần phải thêm bất kì liên kết nào giữa \*PRO\* và chủ ngữ hiện. Tuy nhiên, cũng có những trường hợp trong đó \*PRO\* phát sinh nhưng nó không bị chi phối bởi một mệnh đề cao hơn. Trong những trường hợp đó, nó không được đánh cùng chỉ số với một chủ ngữ được hiện thực hoá đầy đủ. Ví dụ:

*Chú giải TreeBank:*

*(NP-SBJ*

*(NP*

*(PRP it)*

*(S*

*(-NONE- \*EXP\*-1)*

*(VP*

*(VBZ is)*

*(ADJP-PRD*

*(JJS best)*

*(S-1*

*(NP-SBJ*

*(-NONE- \*PRO\*)*

*(VP*

*(RB not)*

*(TO to)*

*(VP*

*(VB use)*

*(NP*

*(PRP it)*

*(S-CLR*

*(NP-SBJ*

*(-NONE- \*PRO\*)*

*(VP*

*(TO to)*

*(VP*

*(VB cut)*

*(NP*

*(ADJP*

*(RB very)*

*(JJ hot)*

*(NN food)*

*Chú giải PropBank:*

*REL: cut*

*ARG0: [NP-SBJ -NONE- \*PRO\*]*

*ARG1: [NP very hot food]*

Khi chú giải \*PRO\* không được đánh chỉ số, nếu người chú giải chắc chắn rằng chủ ngữ được hiện thực hoá ở chỗ nào khác trong trường hợp này, thì cần tạo một liên kết giữa \*PRO\* và tham chiếu hiện. Nếu người chú giải không hoàn toàn chắc chắn rằng tham chiếu hiện và \*PRO\* cùng chỉ một tham chiếu thì người chú giải không nên tạo liên kết. Về cơ bản, trừ khi mối quan hệ này là hoàn toàn chắc chắn, chúng ta nên đi theo các chú giải của TreenBank và các chỉ số đã có.

*Chú giải TreeBank:*

(NP-SBJ

(NP

(NNP China)

(POS ’s)

(NN income)

(NNS taxes)

(VP

(VBD amounted)

(PP-CLR

(TO to)

(NP

(QP

(RB approximately)

(CD 180)

(CD billion)

(, ,)

(S-ADV

(NP-SBJ

(-NONE- \*PRO\*)

(VP

(VBG accounting)

(PP-CLR

(IN for)

(NP

(NP

(QP

(RB about)

(CD 6)

(SYM \%)

(PP

(IN of)

(NP

(NP

(DT the)

(NN state)

(POS ’s)

(JJ financial)

(NNS revenues)

*Chú giải PropBank:*

*REL: accounting*

*ARG0: [NP-SBJ -NONE- \*PRO\*] \* [NP China’s income taxes]*

*ARG1: [PP-CLR for about 6% of the state’s ﬁnancial revenues]*

Mục tiêu của chú giải này là cung cấp thông tin ngữ nghĩa bổ sung về các tham thể của động từ. Trong một số trường hợp, các mệnh đề đứng trước không phải là các thành phần cú pháp, hoặc có một dạng hình thái khác, như đại từ sở hữu “*your*” được minh họa dưới đây, đi trước liên kết trong các trường hợp này:

*On the issue of abortion , Marshall Coleman wants to take away your right [\*] to choose and give it to the politicians.*

*ARG0: [\*PRO\*]*

*REL: choose*

Ngoài ra, lưu ý rằng các phần tử null cần được liên kết với nút cao nhất có thể chứa tham chiếu của nó mà không chú giải đệ quy các tham thể khác hoặc chính thành phần REL (tức là, tạo ra các tham thể nhúng trong các tham thể khác). Nếu điều này là không thể, thì liên kết cần được bỏ qua.

### Kết nối nhiều nút vào một tham thể

Theo quy tắc, các chú giải phải được đặt ở nút cao nhất có thể bao hàm toàn bộ 1 thành phần (ví dụ mức NP hay PP trong TreeBank), và một thẻ phải tương ứng với một nút với mọi nút nằm trong vùng chú giải phù hợp. Tuy nhiên, trong một số trường hợp, thành phần quan hệ rel nằm trong một nút NP, và cách phân bố của TreeBank làm cho không thể tóm gọn toàn bộ một thành phần dưới một nút. Trong những trường hợp này, 2 lá của cây có thể được kết nối cùng nhau dưới cùng một nhãn tham thể. Ví dụ:

*Chú giải TreeBank:*

*...*

*(ADJP*

*(RB widely)*

*(VBN covered)*

*(NN skiing)*

*(NN competition)*

*Chú giải PropBank:*

*REL: covered*

*ARGM-MNR: [RB widely]*

*ARG1: [NN skiing] , [NN competition]*

Trong trường hợp tỉnh lược bị động của chủ ngữ và các động từ đề bạt chủ ngữ như “seem”, việc kết hợp có thể được yêu cầu để đặt chủ ngữ và mệnh đề đi sau động từ dưới cùng một nhãn tham thể.

### Các trường hợp đặc biệt của việc đề hoá

Đề hoá xảy ra khi một thành phần được chuyển từ vị trí cú háp thong thường tới một vị trí không điển hình để thu hút sự chú ý vào thành phần đó. Mặc dù phần lớn các trường hợp đề hoá chỉ yêu cầu chú giải một vết đã được đánh chỉ số tới thành phần được đề hoá, đôi khi thành phần đề hoá được lặp lại, thường dưới dạng một đại từ, và 2 thành phần này không được chỉ số hoá sẵn trong cây. Ví dụ: “*911, that’s the number you call in an emergency*”. Nếu nhãn REL được chú giải là “*is*”, thì người chú giải trước tiên phải chú giải đại từ “that” là ARG-1, sau đó cung cấp một liên kết đồng sở chỉ tới tham chiếu “911” của đại từ. Việc thực hiện liên kết này hoạt động giống hệt như tạo tham chiếu đồng sở chỉ cho \*PRO\* đã trình bày ở trên. Ví dụ:

*Chú giải TreeBank:*

(S

(NP-SBJ

(PRP I))

(VP

(VBD expected)

(NP

(PRP it))

(PP-LOC

(IN in)

(NP

(NP

(DT a)

(NN country))

(SBAR

(WHNP-1

(-NONE- 0))

(S

(NP-SBJ

(PRP we))

(VP

(VBP love)

(NP

(-NONE- \*T\*-1))

(, ,)

(NP-TPC

(PRP me)

(CC and)

(PRP you))))))))

*Chú giải PropBank:*

REL: love

ARG1: (NP (-NONE- \*T\*-1))

ARG0: [NP-SBJ we] \*[NP-TPC me and you]

### Trường hợp đặc biệt: tiểu cú và bổ ngữ câu

Phần này liên quan tới các kiểu bổ ngữ mệnh đề và các phụ ngữ, đặc biệt nhấn mạnh vào tỉnh lược bị động, động từ đề bạt chủ ngữ và các động từ về thể. Trong câu sau, mệnh đề S-CLR có một vết ở vị trí chủ ngữ của “asleep”, được đánh cùng chỉ số với chủ ngữ “I” của động từ “fell”.

*I fell asleep on the ﬂoor*.

*Chú giải TreeBank:*

*S (NP-SBJ-1 I)*

*(VP fell*

*(S-CLR (NP-SBJ \*-1)*

*(ADJP-PRD asleep))*

*(PP-LOC on*

*(NP the lobby floor))))*

Khi chú giải từ “*fell*”, tiểu cú (được đánh dấu S-CLR ở trên ) được gán nhãn là ArgM-PRD, và tham thể Arg1 là NP-SBJ: “*I*”. Chú ý rằng mặc dù nhãn rỗng NP-SBJ\*-1 có cùng chỉ số với “*I*”, vết này không phải là tham thể của “*fell*”, mà là chủ ngữ của “*asleep*”.

*Chú giải PropBank:*

REL: fell

ARG1: I

ARGM-PRD: [NP-SBJ \*-1] asleep

Những động từ như “*expect*” đuộc phân tích là có một mệnh đề là tham thể của nó (tương ứng với sự kiện được chờ đợi). Trong trường hợp này chú giải PropBank tuân theo phân tích của TreeBank, trong đó bổ ngữ mệnh đề được chọn là Arg1:

*John expected Mary to come.*

***Chú giải PropBank:***

REL: expected

ARG0: John

ARG1: Mary to come

Nếu những câu như vậy được bị động hóa như chỉ ra dưới đây, thì tham thể Arg1 là bổ ngữ mệnh đề của động từ. Song song với các vết ICH và RNR, ta giả theiets rằng vết [\*-1] được tái tạo, sao cho Arg1 trong trường hợp này tương ứng với vị trí “Mary to come”. Cần phải chú giải phần bị dịch chuyển (ví dụ NP Mary - 1) như một phần của Arg1 bằng cách sử dụng ‘,’ đã giới thiệu ở phần 1.7.3. Quá trình một chủ ngữ được đề bạt từ bổ ngữ mệnh đề thành chủ ngữ của động từ chính “expect” đôi khi được gọi là tỉnh lược bị động.

*Mary-1 is expected [\*]-1 to come*

REL: expected

ARG1: [Mary-1] , [\*-1 to come]

Một phân tích tương tự áp dụng cho các động từ “*seem*” và “*appear*” được biết đến như là các động từ đề bạt. Ví dụ, NP “Everyone” không phải là tham thể của động từ “seems”, mà câu này có thể được diễn giải như *“It seems that everyone dislikes Drew Barrymore*”.

*Everyone seems to dislike Drew Barrymore*

***Chú giải TreeBank:***

(S (NP-SBJ-3 Everyone)

(VP seems

(S (NP-SBJ \*-3)

(VP to

(VP dislike

(NP Drew Barrymore))))))

Trong chú giải PropBank, mệnh đề S và từ bị dịch chuyển ‘*everyone’* được chú giải là tham thể Arg1, cũng thông qua việc kết hợp:

REL: seems

ARG1: [NP-SBJ-3 Everyone], [NP-SBJ\*-3 to dislike Drew Barrymore]

Và cuối cùng, một lớp động từ khách tuân theo cách phân tích này bao gồm các động từ về thể như “*continue*” và “*start*”, trong đó các sự kiện là các tham thể của chúng. Cần xem xét tập vai nghĩa cho những động từ này một cách cẩn thận, thường có một sự tách biệt giữa nghĩa thuộc về thể và và nghĩa tác nhân của động từ (ví dụ: begin.01 và begin.02).

*[New loans]-4 continue [\*-4] to slow.*

Chú giải PropBank:

REL: continue

ARG1: [[New loans]-4] , [\*-4 to slow]

## Chú giải các động từ yếu

Những cách sử dụng động từ yếu là cách sử dụng trong đó các động từ được coi là giảm mất ngữ nghĩa so với nghĩa của động từ đó trong những ngữ cảnh khác. Ví dụ, một phiên bản “mạnh” của động từ “make” là trong câu “She made a pie out of fresh cherries and refrigerated dough.” Cách sử dụng này phản ảnh các vai nghĩa thong thường gắn với nghĩa tác tạo của “make”: tác nhân, sản phẩm và nguyên liệu. Phiên bản yếu của động từ “make” là trong câu: “She made an offer to buy the company”. Trong câu này động từ “make” không đặc tả nghĩa của sự kiện, mà bổ ngữ danh từ hành động hay vị từ thực “offer” mới đặc tả sự kiện. Vì thế, ta thường có thể đặt lại câu cho các cấu trúc động từ yếu với động từ thay thế của bổ ngữ danh từ hành động mà không làm mất nghĩa của câu: “she offered to buy the company”. Vì động từ yếu trong các trường hợp này không phải là thành phần đặc tả ngữ nghĩa của sự kiện hay chiếu cấu trúc tham thể, ta không thể xử lí các động từ yếu giống như cách xử lí các động từ mạnh. Do vậy ta dùng cách gán nhãn dưới đây, sử dụng bộ nhãn vai nghĩa .LV

Chú giải bước 1: Chú giải rel là động từ yếu với bổ ngữ danh từ hành động là ArgM-PRR (*PRedicating Relation*). Ví dụ:

*The lawmakers made a big mistake.*

REL: made

ARGM-PRR: mistake

Chú giải bước 2: Chú giải rel là danh từ có vai trò bổ ngữ cho động từ yếu ở trên. Ví dụ:

*Yesterday, Mary made an accusation of duplicity against John because she was enraged with jealousy*.

ARGM-TMP: Yesterday

ARG0: Mary

ARGM-LVB: made

REL: accusation

ARG2: of duplicity

ARG1: against John

ARGM-CAU: because she was enraged with jealousy.

Như vậy, tất cả các tham thể của vị từ phức (trong trường hợp này là make+accusation) được chú giải phù hợp với cấu trúc tham thể của vị từ, được đưa ra trong tập vai nghĩa của quan hệ danh từ.

## Chú giải quan hệ danh từ

Danh từ có thể được chú giải quan hệ giống như động từ. Tập vai nghĩa của một quan hệ danh từ có thể bao gồm:

* Vai chủ sở hữu của danh từ Arg0
* Vai định ngữ của danh từ Arg1
* Các vai phụ ngữ như ArgM-Adj, ArgM-Adv, ArgM-Tmp, ArgM-Neg.

Ví dụ:

*Channel Nine’s broadcast of the nightly news was praised for its quality*.

ARG0: Channel Nine’s

REL: broadcast

ARG1: of the nightly news

*The nightly news broadcast of Channel Nine was praised for its quality.*

ARG1: nightly news

REL: broadcast

ARG0: of Channel Nine

*Was your personal experience of having been on the cruise pleasurable?*

ARG0: your

ARGM-MNR: personal

REL: experience

ARG1: of having been on the cruise

*The mayor’s shocking abuse of public funds outraged citizens.*

ARG0: The mayor’s

ARGM-ADJ: shocking

REL: abuse

ARG1: of public funds

*Mary’s investigation and eventual condemnation of the local government made the news.*

ARG0: Mary’s

ARGM-TMP: eventual

REL: condemnation

ARG1: of the mayor

# Kho ngữ liệu có gán nhãn vai nghĩa tiếng Việt (Vietnamese PropBank)

Qua khảo sát chi tiết về dự án xây dựng English PropBank, chúng ta thấy rằng để xây dựng được một kho ngữ liệu tương đương với kho English PropBank, cần phải sở hữu các nguồn tài nguyên ngôn ngữ sau đây:r

* Kho ngữ liệu có gán nhãn cú pháp TreeBank
* Kho từ vựng có chú giải chi tiết thông tin về khung vị từ cho từng nghĩa của mỗi động từ VerbNet.

Tuy nhiên, nguồn tài nguyên ngôn ngữ hiện tại cho tiếng Việt mới chỉ có VietTreeBank, kho ngữ liệu đã chú giải cú pháp. Do vậy, trong khuôn khổ dự án này, chúng tôi chỉ có thể xây dựng một kho ngữ liệu có gán nhãn vai nghĩa tiếng Việt dựa theo bộ nhãn vai nghĩa được định nghĩa dưới đây, mà không thể có tham chiếu tới nghĩa tương ứng của động từ trong VerbNet như trong kho PropBank tiếng Anh.

***Cơ sở xây dựng bộ nhãn vai nghĩa:***

- Bộ nhãn vai nghĩa trong PropBank tiếng Anh

- Các vai nghĩa trong tài liệu bàn về vai nghĩa tiếng Việt

Mỗi câu có thể có vị ngữ do động từ, tính từ, số từ, danh từ hoặc các giới từ làm trung tâm. Với mỗi thành phần trung tâm vị ngữ, ta gán nhãn vai nghĩa cho các thành phần chính (tham thể -Arg) quây quần quanh trung tâm đó, và cũng gán nhãn vai nghĩa cho các thành phần phụ (modifier – ArgM) của nó. Mỗi vai nghĩa này có thể được chia thành các vai nghĩa con đối với các trường hợp phân biệt tường minh giữa các vai nghĩa với nhau.

Các tham thể Arg được đánh số từ 0 đến 4, theo thứ tự xuất hiện “tự nhiên” của chúng trong câu.

Chẳng hạn:

* Từ “tôi” trong “**tôi hát**” sẽ có vai nghĩa được đánh số là Arg0. Vị từ quan hệ là **hát(tôi)**
* Các từ « tôi », « cơm » trong “**tôi ăn cơm** » có vai nghĩa đánh số lần lượt là Arg0, Arg1. Vị từ quan hệ **ăn(tôi, cơm)**.
* Các từ « tôi », « hoa », « cô ấy » trong “**tôi tặng hoa cho cô ấy**» có vai nghĩa đánh số lần lượt là Arg0, Arg1, Arg2 và có cùng vai nghĩa trong câu « **tôi tặng cô ấy hoa** ». Vị từ quan hệ **tặng(tôi, hoa, cô ấy)**.

Sau đây là chi tiết danh sách vai nghĩa đi kèm chú thích và ví dụ minh hoạ.

**I. Gán nhãn vai nghĩa cho các thành phần câu**

**1. Vị ngữ do động từ đảm nhiệm:**

**1.1. Arg0: Tác thể**

**-** Biểu thị người, động vật gây ra hành động. Ví dụ:

1. *Nam***Arg0** đến trường.

2. *Đại tướng***Arg0**đứng dậy.

- Biểu thị đối tượng là nguyên nhân của kết quả được nêu trong câu. Ví dụ:

1. *Bão***Arg0**làm đổ cây.

2. *Gió***Arg0** thổi tắt đèn.

- Biểu thị chủ thể trải nghiệm một trạng thái nào đó. Ví dụ:

1. *Nó***Arg0** mệt.

2. *Nó***Arg0**buồn ngủ.

3. Và *tôi***Arg0** cũng yêu em, yêu em rộn ràng yêu em nồng nàn.

**1.2. Arg1Patient: Bị thể**

- Chỉ vật, người chịu sự tác động, dẫn đến một thay đổi nào đó, thay đổi về vật chất hoặc tinh thần.

- Ví dụ:

1. Bộ đội phá *cầu***Arg1-Patient**.

2. Nó doạ ma *cô ấy***Arg1-Patient**.

3. Nó đập vỡ *cái cốc***Arg1-Patient**.

**1.3. Arg2:**

**-** Biểu thị kẻ được hưởng thành quả từ một hành động do một ai đó thực hiện. Ví dụ:

1. Nó chữa cái xe cho *chị ấy***Arg2**.

2. Mẹ rửa chân cho *em***Arg2**.

**-** Chỉ công cụ được dùng để thực hiện hành động. Ví dụ:

1. *Xe hơi***Arg2** chở họ chạy đi.

2. Nó ăn cơm bằng *đũa***Arg2**.

- Biểu thị kẻ tiếp nhận hành động trong câu chứa vị từ trao tặng. Ví dụ:

1. Cô ấy tặng *Lan***Arg2** một quyển sách.

2. Ngân hàng cho *nhà Lan***Arg2**vay 50 triệu.

**1.4. Arg3:**

**-** Chỉ điểm xuất phát của chuyển động, hay nguồn của trạng thái.

- Ví dụ:

1. Nó từ *Hải Phòng***Arg3** đến.

2. Nó chết mê chết mệt *cô ấy***Arg3**.

- Chú ý: Đối với trường hợp Arg3 mang nghĩa là nguồn của trạng thái thì Arg3 không có khả năng cải biến thành câu bị động.

**1.5. Arg4Goal: Đích đến:**

**-** Chỉ đích đến của một chuyển động.

- Ví dụ: Hôm nay nó đến *trường***Arg4-Goal**.

**1.2. Vị ngữ do động từ “là” đảm nhiệm (Rel: là)**

**Quan hệ đồng nhất: D1 là D2 ⬄ D2 là D1**

**1.2.1. Arg0Identified: Bị đồng nhất thể**

**-** Yếu tố chỉ thực thể được đưa ra để nhận biết

**1.2.2. Arg1Identifier: Đồng nhất thể**

**-** Chỉ thực thể dùng để nhận diện thực thể kia

- Ví dụ:

1. *Hà Nội***Arg1-Identifier** là *thủ đô của Việt Nam***Arg0-Identified**.

2. *Cầu thủ giỏi nhất ở đây***Arg0-Identified** là *anh ấy* **Arg1-Identifier**.

**Quan hệ thuộc tính: D1 là D2 khác D2 là D1**

**1.2.3. Arg0Carrier: Vai đương thể**

- Yếu tố đứng trước nêu ra thực thể được đinh tính, tức là thực thể mang cái thuộc tính sẽ nói ở vế sau

**1.2.4. Arg1Attribute: Vai thuộc tính thể**

**-** Yếu tố đứng sau chỉ ra lớp thực thể mà yếu tố đứng trước thuộc vào

- Ví dụ:

1. *Hà Nội***Arg0-Carrier** là *một thành phố đẹp***Arg1-Atribute**.

2. *Cá voi***Arg0-Carrier** là *loài động vật có vú***Arg1-Attribute**.

**Đối với những trường hợp mà chủ ngữ của câu chứa vị từ “là” không phải là danh từ:**

**1.2.5. Arg0:**

- Các đại từ/cụm từ dùng để chỉ một sự vật, địa điểm, vị trí, thời gian, như: *kia, đây, đâu đó, nơi đây, chỗ này, chỗ kia, chỗ đó, nơi này, đó, trước đây, trước đó…* Ví dụ:

1. *Kia***Arg0** là bạn tôi.

2. *Nơi đây***Arg0** là nơi tôi sinh ra.

- Các từ chỉ định hướng về khoảng cách, như: *xa xa, 10 km, 10 mét,…* Ví dụ:

1. *Xa xa***Arg0** là cánh buồm căng gió.

2. *10 mét nữa***Arg0** là đến thành phố Yên Bái rồi.

- Các động từ giữ vai trò là chủ ngữ. Ví dụ:

1. *Sống***Arg0** là thay đổi.

2. *Khóc***Arg0** là nhục.

- Các cú giữ vai trò làm chủ ngữ. Ví dụ:

1. *Hay nói ầm ì***Arg0** là con vịt bầu.

2. *Hay hỏi đâu đâu***Arg0** là con chó vện.

**1.2.6. Arg1: Là vị ngữ tương ứng với chủ ngữ trong các câu mà chủ ngữ không phải là danh từ**

1. *Khóc***Arg0** là *nhục***Arg1**.

2. *Xa xa***Arg0** là *cánh buồm căng gió***Arg1**.

**1.3. Vị ngữ do tính từ đảm nhiệm (Rel = Adjective)**

**1.3.1. Arg0Carrier: Vai đương thể**

- Thực thể mang cái thuộc tính sẽ nói ở vế sau

- Ví dụ:

1. *Trăng***Arg0-Carrier** sáng.

2. *Cô ấy***Arg0-Carrier** thông minh.

**1.3.2. Arg1Location: Vai địa điểm**

- Ví dụ:*Nhà* **Arg0-Carrier** xa *trường***Arg1-Location**.

**1.3.3. Arg1Attribute: Vai thuộc tính thể**

- Yếu tố đứng sau chỉ ra lớp thực thể mà yếu tố đứng trước thuộc vào

- Ví dụ: *Nhà này***Arg0-Carrier** rộng *9 mét vuông***Arg1-Attribute**.

**1.4. Vị ngữ do số từ đảm nhiệm (Rel = dummy)**

**Arg0Carrier: Vai đương thể:** Thực thể mang thuộc tính biểu hiện ở sau

Ví dụ: *Ông ấy***Arg0-Carrier** 30**Arg1-Attribute**.

\* Trường hợp vị ngữ là số từ + danh từ, ví dụ: Ông ấy 30 tuổi => dummy (ông ấy, 30 tuổi)

**1.5. Vị ngữ do danh từ đảm nhiệm => dummy (Arg0; Arg1)**

**1.5.1. Arg0:**

**1.5.1.1. Arg0Carrier: Vai đương thể**

- Ví dụ: *Cái áo***Arg0-Carrier** màu xanh.

**1.5.1.2. Arg0Posseser: Vai chủ sở hữu**

- Chỉ chủ sở hữu của sự vật

- Ví dụ: *Phòng nào***Arg0-Possessor** cũng hai cửa sổ.

**1.5.2 Arg1Attribute: Vai sở thuộc thể**

- Chỉ thuộc tính của sự vật được nêu ra trong câu

- Ví dụ: *Đồng hồ* **Arg0-Possessor** *hai kim***Arg1-Attribute**.

**1.6. Vị ngữ do giới từ đảm nhiệm**

**1.6.1. Arg0:**

**1.6.1.1. Arg0Carrier: Vai đương thể**

- Thực thể mang cái thuộc tính sẽ nói ở vế sau

**-** Vai chung cho chủ ngữ của các giới từ: *để, do, bởi, trong, ngoài, trên, dưới,…*

- Ví dụ: *Ông ấy***Arg0-Carrier** ngoài vườn.

**1.6.1.2. Arg0Event: Vai sự việc (Rel = do, bởi, tại)**

- Biểu thị sự việc nào đó được nói đến trong câu

- Ví dụ: *Việc này***Event** do Bính.

**1.6.1.3. Arg0: Vai sở thuộc thể (Rel = của)**

**-** Chỉ sự vật có chủ sở hữu ở trong câu

- Ví dụ: *Cái xe này***Arg0** của *Giáp***Arg1**-**Possessor**.

**1.6.2. Arg1:**

**1.6.2.1. Arg1Material: Vai nguyên liệu (Rel = bằng)**

- Chỉ chất liệu nằm nên đối tượng được nói ở phía trước trong câu:

- Ví dụ: Cái áo nàybằng *lụa***Material**.

**1.6.2.2. Arg1Reason: Vai nguyên nhân (Rel = tại)**

- Chỉ nguyên nhân của sự việc được nêu ra trong câu

- Ví dụ: Việc nàytại *Bính***Reason**.

**1.6.2.3. Arg1Origin: Vai nguồn gốc (Rel = do, bởi)**

**-** Chỉ nguồn gốc của sự việc được nêu ra trong câu

- Ví dụ:*Bài thơ này***Arg0-Origin** do *một bạn trẻ làm***Arg1-Event**.

**1.6.2.4. Arg1Location: Vai địa điểm (Rel = trong, ngoài, trên, dưới,…)**

**-** Chỉ địa điểm của đối tượng được nêu ra trong câu

**-** Ví dụ: *Cái tủ***Arg0-Carrier** *trongnhà***Arg1-Location**.

**1.6.2.5. Arg1Value: Vai công dụng (Rel = để)**

- Chỉ công dụng của đối tượng được nêu ra trong câu

- Ví dụ:*Cái bàn này***Arg0-Carrier** để *uống nước***Arg1-Value**.

**1.6.2.6. Arg1Possessor:Vai chủ sở hữu (Rel = của)**

- Chỉ chủ sở hữu của sự vật trong câu

- Ví dụ: *Cái xe này***Arg0** của *Giáp***Arg1-Possessor**.

**II. Đối với định ngữ:**

**1.1. ArgMComitatives:Vai kẻ cùng hành động (ARGM-COM)**

- Chỉ người cùng hành động trong một hành động.

- Ví dụ: Nó chơi cờ với *bố*ArgM-Com.

**1.2. ArgMLocatitives: Vai địa điểm (ARGM-LOC)**

- Chỉ nơi chốn của sự tình, vị trí tồn tại của sự vật.

- Ví dụ: Tôi gặp chị ấy ngoài *ga*ArgM-Loc.

**1.3. ArgMDirectional: Vai hướng (ARGM-DIR)**

- Chỉ hướng của chuyển động

- Ví dụ: Bộ đội tiến quân dọc theo *bờ sông***ArgM-Dir**.

**1.4. ArgMGoal: Vai đích đến (ARGM-GOL)**

**-** Chỉ mục đích của hành động

- Ví dụ:

1. Nó chạy quanh *ao***ArgM-Gol**.

2. Con cá bơi *trên hồ***ArgM-Gol**.

**1.5. ArgMManner: Vai cách thức (ARGM-MNR)**

**-** Chỉ một hành động được thực hiện như thế nào, một sự việc xảy ra như thế nào

- Ví dụ: Rất may, Hải đã ứng biến *tốt***ArgM-Mnr***.*

**1.6. ArgMTemporal: Vai chỉ thời gian (ARGM-TMP)**

- Chỉ thời lượng kéo dài của hành động, trạng thái, hoặc chỉ thời gian diễn ra hành động, trạng thái đó…

- Danh sách các từ đảm nhiệm vai chỉ thời gian:

+ Phó từ chỉ thời gian là các từ: *đã, chưa, đang, sẽ, vừa, mới, sắp, từng*

*+* Phó từchỉ kết quả của hành động/hoạt động (hoàn thành hay chưa hoàn thành): *xong, rồi, chưa*

*+* Phó từ chỉ sự đồng nhất hay lặp đi lặp lại: *cũng, đều, vẫn, còn, mải, lại, cứ, mãi, nữa*

+ Phó từ chỉ tần số: *thường, hay, năng, thường xuyên, luôn, luôn luôn*

+ Các từ danh từ/cụm danh từ chỉ thời gian: *hôm qua, hôm nay, 2 ngày trước, tối qua, tối hôm trước, ngày mai,…*

+ Các cụm từ chỉ thời lượng: *2 tháng, 1 tuần, 2 ngày, 2 giờ,…*

- Ví dụ:

1. Năm ngoái, nó đi làm *bốn tháng***ArgM-Tmp**.

2. *Hôm qua*A**rgM-Tmp**, nó đến trường.

**1.7. ArgMDiscource: Vai diễn ngôn (ARGM-DIS)**

**-** Chỉ những dấu hiệu kết nối một câu với một câu trước đó hoặc các từ cảm thán.

- Ví dụ: Trước mặt chúng tôi là một ô cửa sổ, chắc là phòng trực, *nhưng***ArgM-Dis** bên trong không có ai.

**1.8. ArgMSecondary Predication: Vai vị từ thứ hai (ARGM-PRD)**

- Chỉ sự liên kết của một vị từ là tự nó có khả năng mang một số cấu trúc vị ngữ

- Ví dụ: Cũng như *số phận của vợ chồng anh Đại***ArgM-Prd**, những công nhân đứng ra tố cáo tiêu cực ở đường liên cảng A5 ngày xưa giờ phiêu bạt khắp nơi.

**1.9. ArgMAdverbials: Vai trạng ngữ (ARGM-ADV)**

**-** Biểu thị thành phần phụ cho cả câu

- Ví dụ: *Túng quá***ArgM-Adv**, họ rủ nhau đi buôn.

Trong hai ví dụ dưới đây, phần chữ in nghiêng của ví dụ 1 là ArgM-Adv, còn phần in nghiêng trong ví dụ 2 là ArgM-LOC.

*VD1: Cũng như anh Đại*, chúng tôi đều bị lão ta lừa lấy tiền.

*VD2: Trên cánh đồng lúa ngày xưa*, chúng tôi đã từng đụng độ nhau dăm ba lần.

**1.10. ArgMNegation: Vai phủ định (ARGM-NEG)**

**-** Thành phần phủ định trong câu, như: *không, chắng,*

- Ví dụ: Nó *không***ArgM-Neg** hiểu vấn đề đang gặp phải.

**1.11. ArgMModal: Tình thái (ARGM-MOD)**

**-** Biểu thị những yêu cầu bắt buộc, mong muốn, khả năng của đối tượng được nói đến trong câu

- *phải, nên, có thể, cần, tất,…*

- Ví dụ: Tôi *phải***ArgM-Modal**đi ngay.

**1.12. ArgMParticle: Tiểu từ tình thái**

**-** Đó là các từ có thể xuất hiện đâu đó trong phát ngôn, không nhất thiết là ở cuối, biểu thị nghĩa bổ trợ cho phát ngôn hoặc các tổ hợp đặc ngữ tương đương, chẳng hạn như: *à, ư, nhỉ, nhé, thôi, chứ, đi, mất, thật, cũng nên, lại còn, thì chết,…*

+ Tình thái từ nghi vấn: *à, ư, hả, hử, chứ, chăng, nhỉ, hở, cơ, nhé,…*

+ Tình thái từ cầu khiến: *đi, nào, với, đi thôi, nhé,...*

+ Tình thái từ cảm thán: *thay, sao, thật.*

+ Tình thái từ biểu thị sắc thái tình cảm: *ạ, nhé, cơ, mà, dạ, vâng, .*

- Ví dụ:

1. Con đi ngủ *à***ArgM-Particle**?

2. Nó ăn cắp tiền của cô ấy *cũng nên***ArgM-Particle**.

**1.13. ArgMPurpose: Mục tiêu, mục đích (ARGM-PRP)**

- Chỉ mục tiêu hoặc mục đích của hành động được nêu ra trong câu

- Ví dụ:

1. Nó đến thư viện *để mượn sách***ArgM-PRP**.

2. Cô ấy mua cá *để nấu canh chua***ArgM-PRP**.

**1.14. ArgMExtent: Vai chỉ mức độ (ARGM-EXT)**

**-** Chỉ sự mức độ hay sự thay đổi về số lượng

- Phó từ chỉ mức độ: *rất, quá, lắm, khá, khí, hơi,..*

- Các từ chỉ số lượng: *những, các, mọi, mỗi, từng, một, nhiều,…*

*-* Ví dụ:

1. Nó *rất***ArgM-Ext** chăm chỉ.

2. Nó tặng tôi *những***ArgM-Ext** bông hoa tươi thắm.

**1.15. ArgMInterjection: Cảm thán (ARGM-I)**

- Đó là những từ dùng để bộc lộ trực tiếp xúc cảm của người nói hoặc được dùng để hỏi đáp, chẳng hạn: *ôi, trời ơi, ô, ủa, kìa, ái, ối, than ôi, hỡi ôi, eo ôi, ôi giời ôi, ơi, vâng, dạ, bẩm...*

- Ví dụ:

1. *Ôi***ArgM-I**! Ngôi nhà mới đẹp làm sao!

2. *Ơ***ArgM-I**! Chị cứ tưởng là ai hóa ra là em.

Vai nghĩa này trong tiếng Anh được xếp vào nhóm vai nghĩa ArgM-DIS.

**1.16. ArgMDirect Speech: Lời thoại trực tiếp (ARGM-DSP):**

Nhãn vai nghĩa này dành cho lời thoại trực tiếp được viết trong cặp nháy kép.

1. Palmer M, Kingsbury P, Gildea D (2005). "The Proposition Bank: An Annotated Corpus of Semantic Roles". *Computational Linguistics* **31** (1): 71–106. [doi](http://en.wikipedia.org/wiki/Digital_object_identifier):[10.1162/0891201053630264](http://dx.doi.org/10.1162%2F0891201053630264) [↑](#footnote-ref-0)