



25 YEARS ANNIVERSARY  
SOICT

HA NOI UNIVERSITY OF SCIENCE AND TECHNOLOGY  
SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY

# PHÂN TÍCH VAI NGHĨA

Một số slide được tham khảo từ tutorial của **Scott Wen-tau Yih & Kristina Toutanova** (Microsoft Research)

# Giới thiệu

Phân tích cú pháp là một bài toán cơ bản trong NLP nhưng kiểu phân tích nào mới là thực sự tốt đối với NLP?

Phân tích cú pháp nhằm thực hiện phân tách câu ra thành các phần có nghĩa hay tìm ra các mối quan hệ có nghĩa mà có thể được sử dụng trong các bài toán tiếp theo về phân tích ngữ nghĩa:

- Gán nhãn vai trò ngữ nghĩa hay còn gọi là phân tích vai nghĩa (chỉ ra ai làm gì cho ai);
- phân tích ngữ nghĩa (chuyển 1 câu thành biểu diễn logic của câu);
- giải quyết nhập nhằng nghĩa từ (chỉ ra các từ trong câu mang ý nghĩa gì);
- xác định ngữ nghĩa hợp thành (tính ý nghĩa của 1 câu dựa trên ý nghĩa các phần của câu).

Trong chương này, chúng ta sẽ tìm hiểu bài toán phân tích vai nghĩa hay gán nhãn vai trò ngữ nghĩa.

# Giới thiệu

- Nhiệm vụ chính của gán nhãn vai trò ngữ nghĩa (**semantic role labeling - SRL**) là chỉ ra một cách chính xác các quan hệ ngữ nghĩa gì là đúng đắn giữa 1 vị từ và các thành phần kết hợp của nó, trong đó các quan hệ này được lấy ra từ một danh sách đã xác định các vai trò ngữ nghĩa có thể đối với vị từ này.

- Ví dụ:

[The girl on the swing]*Agent* [whispered]*Pred* to [the boy beside her]*Recipient*

# Giới thiệu

- Các vai trò điển hình được sử dụng trong SRL là các nhãn chẳng hạn như Agent, Patient, và Location đối với các thực thể tham gia trong 1 sự kiện, Temporal và Manner để đặc trưng các khía cạnh khác của sự kiện hay các quan hệ tham gia khác.
- Cách tiếp cận theo ngôn ngữ tính toán đối với bài toán SRL đòi hỏi xây dựng một từ điển ngữ nghĩa từ vựng và một bộ sưu tập các câu đã chú thích vai nghĩa.
- 2 kho ngữ liệu được xây dựng dựa trên Ngữ nghĩa khung (frame) có thể sử dụng đ/v bài toán SRL là FrameNet và PropBank.

# Các vấn đề đ/v các vai trò ngữ nghĩa

- Rất khó để đưa ra 1 định nghĩa hình thức cho vai trò
- Có các kiểu phân tách vai trò khác nhau tùy ý
- Các giải pháp đ/v vấn đề khó định nghĩa các vai trò ngữ nghĩa:
  - Không chú ý đến nhãn vai trò ngữ nghĩa, chỉ đánh dấu các vai trò/bổ ngữ của các động từ là 0, 1, 2
    - PropBank
  - Xác định các nhãn vai trò ngữ nghĩa đ/v một miền ngữ nghĩa đặc biệt.
    - FrameNet

# Frame

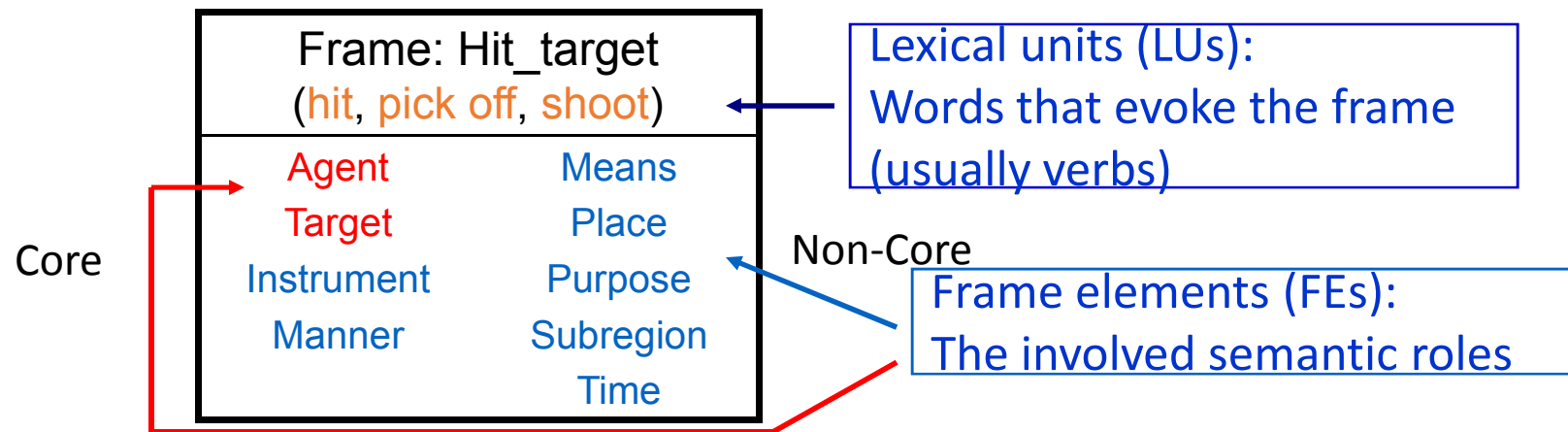
- **Các ngữ nghĩa khung (frame):**
  - *đề xuất bởi Fillmore (1976);*
  - *khung mô tả 1 trường hợp ở dạng nguyên mẫu;*
  - khung được xác định bởi 1 vị từ (predicate);
  - khung có thể bao gồm một số phần tử khung (arguments; sem. roles).

# Frame

- **Các tính chất của ngữ nghĩa khung :**
  - *cung cấp 1 phân tích ngữ nghĩa nông;*
  - *là mức trung gian giữa các vai trò tổng quát và các vai trò đặc biệt theo động từ;*
  - *tổng quát hóa tốt cho các ngôn ngữ khác;*
  - *có thể có lợi cho các ứng dụng NLP khác (IR, QA).*



# FrameNet [Fillmore et al. 01]



[Agent *Kristina*] **hit** [Target *Scott*] [Instrument *with a baseball*] [Time *yesterday*].

# Frames trong FrameNet

frame(TRANSPORTATION) frame_elements(MOVER(S), MEANS, PATH) scene(MOVER(S) move along PATH by MEANS)
frame(DRIVING) inherit(TRANSPORTATION) frame_elements(DRIVER (=MOVER), VEHICLE (=MEANS), RIDER(S) (=MOVER(S)), CARGO (=MOVER(S))) scenes(DRIVER starts VEHICLE, DRIVER controls VEHICLE, DRIVER stops VEHICLE)
frame(RIDING_1) inherit(TRANSPORTATION) frame_elements(RIDER(S) (=MOVER(S)), VEHICLE (=MEANS)) scenes(RIDER enters VEHICLE, VEHICLE carries RIDER along PATH, RIDER leaves VEHICLE )

Figure 1: A subframe can inherit elements and semantics from its parent

[Baker, Fillmore, Lowe, 1998]

FEG	Annotated Example from BNC
D	[ <sub>D</sub> Kate] <b>drove</b> [ <sub>P</sub> home] in a stupor.
V, D	A pregnant woman lost her baby after she fainted as she waited for a bus and fell into the path of [ <sub>V</sub> a lorry] <b>driven</b> [ <sub>D</sub> by her uncle].
D, P	And that was why [ <sub>D</sub> I] <b>drove</b> [ <sub>P</sub> eastwards along Lake Geneva].
D, R, P	Now [ <sub>D</sub> Van Cheele] was <b>driving</b> [ <sub>R</sub> his guest] [ <sub>P</sub> back to the station].
D, V, P	[ <sub>D</sub> Cumming] had a fascination with most forms of transport, <b>driving</b> [ <sub>V</sub> his Rolls] at high speed [ <sub>P</sub> around the streets of London].
D+R, P	[ <sub>D</sub> We] <b>drive</b> [ <sub>P</sub> home along miles of empty freeway].
V, P	Over the next 4 days, [ <sub>V</sub> the Rolls Royces] will <b>drive</b> [ <sub>P</sub> down to Plymouth], following the route of the railway.

Figure 2: Examples of Frame Element Groups and Annotated Sentences

# Các vấn đề đ/v FrameNet

- Các câu mẫu được chọn thủ công
  - Không lựa chọn ngẫu nhiên
  - Không gán nhãn toàn bộ câu
- Do TreeBank không được sử dụng
  - Không phân tích cú pháp hoàn hảo đ/v câu

# Phương pháp luận đối với xây dựng FrameNet

1. Định nghĩa 1 khung (eg DRIVING)
2. Tìm một số câu đối với khung này
3. Chú thích các câu

- Corpora
  - FrameNet I – British National Corpus only
  - FrameNet II – LDC North American Newswire corpora
- Size
  - >8,900 lexical units, >625 frames, >135,000 sentences

<http://framenet.icsi.berkeley.edu>

# Proposition Bank (PropBank) [Palmer et al. 05]

- Dựa trên Penn TreeBank
- Chú thích *mỗi tree* trong Penn TreeBank một cách hệ thống
  - Các thống kê trong corpus này là có ý nghĩa
- Giống FrameNet, dựa trên các lớp động từ của Levin (theo VerbNet)
- Hướng dữ liệu hơn & bottom up
  - Không có mức trừu tượng xa hơn nghĩa động từ
  - Chú thích mỗi động từ xuất hiện trong câu bất kể nó có thuộc khung hay không.

# Proposition Bank (PropBank) [Palmer et al. 05]

- Chuyển các câu thành các mệnh đề (propositions)
  - **Kristina** hit **Scott** → hit(**Kristina**,**Scott**)
- Penn TreeBank → PropBank
  - Thêm 1 tầng ngữ nghĩa trên Penn TreeBank
  - Xác định 1 tập các vai nghĩa đối với mỗi động từ
  - Các vai nghĩa của mỗi động từ được đánh số

...[**A0** the company] to ... *offer* [**A1** a 15% to 20% stake] [**A2** to the public]

...[**A0** Sotheby's] ... *offered* [**A2** the Dorrance heirs] [**A1** a money-back guarantee]

...[**A1** an amendment] *offered* [**A0** by Rep. Peter DeFazio] ...

...[**A2** Subcontractors] will be *offered* [**A1** a settlement] ...

# Proposition Bank (PropBank)

Xác định tập các vai nghĩa

- Rất khó để xác định được 1 tập các vai nghĩa chung đối với tất cả các kiểu vị từ (verbs).
- PropBank xác định các vai nghĩa và ý nghĩa của chúng đối với mỗi động từ trong frame files.
- Các arguments (core) được đánh nhãn bởi các con số.
  - A0 – Agent; A1 – Patient or Theme
  - Other arguments – no consistent generalizations
- Adjunct-like arguments – *universal* đối với tất cả verbs
  - AM-LOC, TMP, EXT, CAU, DIR, PNC, ADV, MNR, NEG, MOD, DIS

# Proposition Bank (PropBank)

## Frame Files

- hit.01 “strike”

- ❖ A0: agent, hitter; A1: thing hit;  
A2: instrument, thing hit by or with

[<sub>A0</sub> *Kristina*] **hit** [<sub>A1</sub> *Scott*] [<sub>A2</sub> *with a baseball*] *yesterday*.

AM-TMP  
Time

- look.02 “seeming”

- ❖ A0: seemer; A1: seemed like; A2: seemed to

[<sub>A0</sub> *It*] **looked** [<sub>A2</sub> *to her*] *like* [<sub>A1</sub> *he deserved this*].

- deserve.01 “deserve”

- ❖ A0: deserving entity; A1: thing deserved;  
A2: in-exchange-for

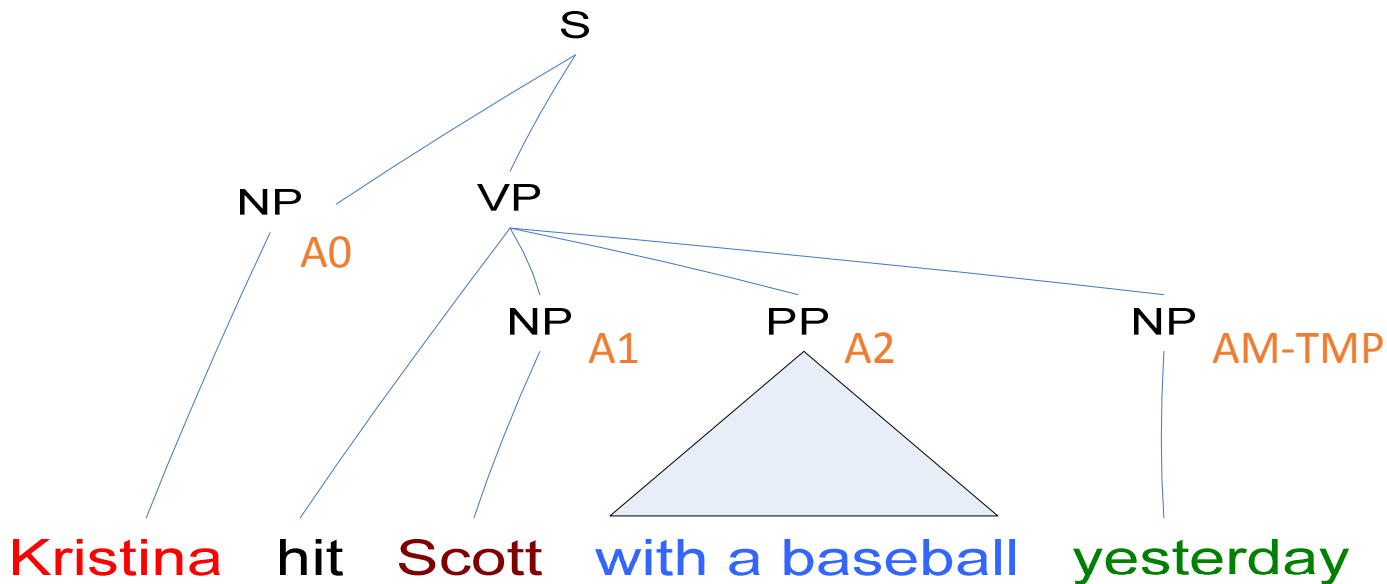
*It looked to her like* [<sub>A0</sub> *he*] **deserved** [<sub>A1</sub> *this*].

Proposition:  
A sentence and  
a target verb



# Proposition Bank (PropBank)

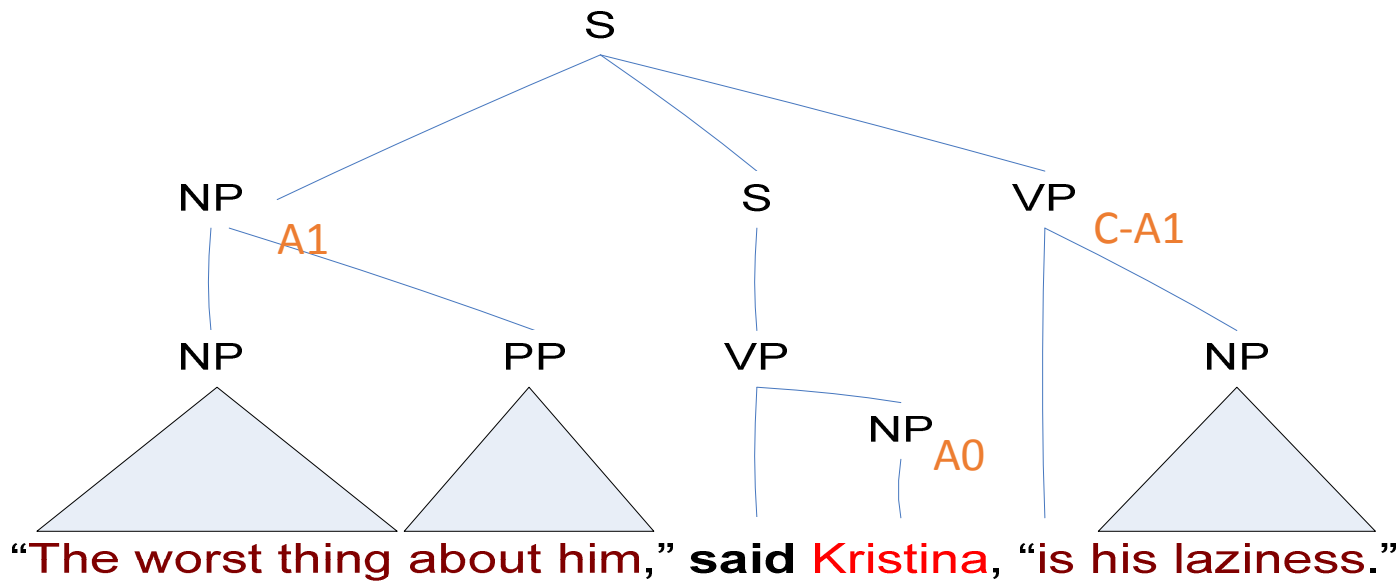
## Thêm 1 tầng ngữ nghĩa



[<sub>A0</sub> *Kristina*] **hit** [<sub>A1</sub> *Scott*] [<sub>A2</sub> *with a baseball*] [<sub>AM-TMP</sub> *yesterday*].

# Proposition Bank (PropBank)

## Thêm 1 tầng ngữ nghĩa– Continued



[<sub>A1</sub> *The worst thing about him*] **said** [<sub>A0</sub> *Kristina*] [<sub>C-A1</sub> *is his laziness*].



# Một số nghĩa động từ và “framesets” trong propbank

Frameset: **decline.01** “go down incrementally”

Arg1: entity going down

Arg2: amount gone down by, EXT

Arg3: start point

Arg4: end point

Ex: ...[Arg1 its net income] *declining* [Arg2-EXT 42%] [Arg4 to \$121 million] [ArgM-TMP in the first 9 months of 1989]. (wsj\_0067)

Frameset: **decline.02** “demure, reject”

Arg0: agent

Arg1: rejected thing

Ex: [Arg0 A spokesman<sub>i</sub>] *declined* [Arg1 \*trace\*<sub>i</sub> to elaborate] (wsj\_0038)

# FrameNet vs PropBank -1

FRAMENET ANNOTATION:

[Buyer Chuck] *bought* [Goods a car] [Seller from Jerry] [Payment for \$1000].

[Seller Jerry] *sold* [Goods a car] [Buyer to Chuck] [Payment for \$1000].

PROPBANK ANNOTATION:

[Arg0 Chuck] *bought* [Arg1 a car] [Arg2 from Jerry] [Arg3 for \$1000].

[Arg0 Jerry] *sold* [Arg1 a car] [Arg2 to Chuck] [Arg3 for \$1000].