# Lecture 2: Event Structure

Weiwei Sun

Department of Computer Science and Technology
University of Cambridge

Michaelmas 2024/25

The explosive **eruption** of *Hunga-Tonga Hunga-Ha'apai* **sent** *a shockwave* *around the world*.

The **event** literally touched every corner of the globe as the pressure wave spread out in all directions to complete a full circumnavigation.

WN S: (n) **event** (something that happens at a given place and time)

https://www.bbc.co.uk/news/science-environment-60029815

1. Events and participants
2. Subcategorisation, arguments and adjuncts
3. Semantic role labeling
4. Sisters, aunts, great-aunts, . . .
5. IKEAing annotations

# Events and Participants

https://www.bbc.co.uk/news/av/world-asia-60007163

# Computational lexicography

## Cambridge Dictionary *send* (verb)

- to cause something to go from one place to another, especially by post or email
- to cause or order someone to go and do something ...

# Computational lexicography

## Cambridge Dictionary *send* (verb)

- to cause something to go from one place to another, especially by post or email
- to cause or order someone to go and do something . . .

## FrameNet *send*.v SENDING

A Sender plans the Path (along with Source and Goal) of a Theme and places it in circumstances such that it travels along this Path under the power of some entity other than the Sender.

The explosive eruption of Hunga-Tonga Hunga-Ha'apai SENT a shockwave around the world.

FrameNet provides schematic representations of the conceptual structures and patterns of beliefs, practices, institutions, images, etc. FRAME ELEMENTS and LEXICAL UNITS.

# Frame Semantics

- **Assumption**: To understand the meanings of the words in a language we must first have knowledge of the semantic frames
- A semantic frame is a schematic representation of an **event**, object, situation, or relation providing the background structure against which words are understood

from https://framenet2.icsi.berkeley.edu/docs/allslides2.pdf

C. Fillmore (9 Aug 1929 – 13 Feb 2014)

# List as many verbs for eating that you know in English

*breakfast.v, consume.v, devour.v, dine.v, down.v, drink.v, eat.v, feast.v, feed.v, gobble.v, gulp.n, gulp.v, guzzle.v, have.v, imbibe.v, ingest.v, ingestion.n, lap.v, lunch.v, munch.v, nibble.v, nosh.v, nurse.v, put away.v, put back.v, quaff.v, sip.n, sip.v, slurp.n, slurp.v, snack.v, sup.v, swig.n, swig.v, swill.v, tuck.v*

# List as many verbs for eating that you know in English

*breakfast.v, consume.v, devour.v, dine.v, down.v, drink.v, eat.v, feast.v, feed.v, gobble.v, gulp.n, gulp.v, guzzle.v, have.v, imbibe.v, ingest.v, ingestion.n, lap.v, lunch.v, munch.v, nibble.v, nosh.v, nurse.v, put away.v, put back.v, quaff.v, sip.n, sip.v, slurp.n, slurp.v, snack.v, sup.v, swig.n, swig.v, swill.v, tuck.v*

## Frame: Ingestion

An Ingestor consumes food or drink (Ingestibles), which entails putting the Ingestibles in the mouth for delivery to the digestive system. This may include the use of an Instrument. Sentences that describe the provision of food to others are NOT included in this frame.

The wolves DEVOURED the carcass completely.

# List as many verbs for eating that you know in English

*breakfast.v, consume.v, devour.v, dine.v, down.v, drink.v, eat.v, feast.v, feed.v, gobble.v, gulp.n, gulp.v, guzzle.v, have.v, imbibe.v, ingest.v, ingestion.n, lap.v, lunch.v, munch.v, nibble.v, nosh.v, nurse.v, put away.v, put back.v, quaff.v, sip.n, sip.v, slurp.n, slurp.v, snack.v, sup.v, swig.n, swig.v, swill.v, tuck.v*

Lexical Units

## Frame: INGESTION

An Ingestor consumes food or drink (Ingestibles), which entails putting the Ingestibles in the mouth for delivery to the digestive system. This may include the use of an Instrument. Sentences that describe the provision of food to others are NOT included in this frame.

The wolves DEVOURED the carcass completely.

Frame Elements, , . . .

# FrameNet

A computational lexicography project based on the principles of Frame Semantics

- 1,224 frames
- 13,640 lexical units
- 10,542 frame elements
- 1,876 frame-to-frame relations
- 202,229 annotated sentences
- 14% “full-text” annotation

from `https://framenet2.icsi.berkeley.edu/docs/allslides2.pdf`

# Useful

- Provides a shallow semantic analysis (no modality, scope);
- generalises well across some languages;
- can benefit various NLP tasks (IR, QA).

**How much did Microsoft pay for Activision Blizzard?**

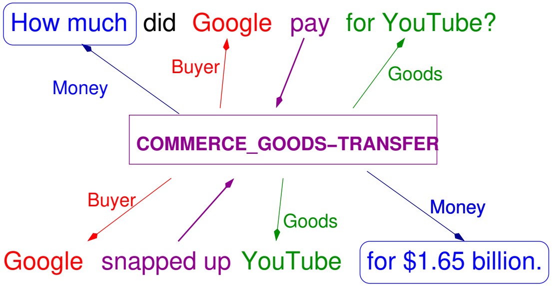
Microsoft Corp (MSFT.O) is buying "Call of Duty" maker Activision Blizzard (ATVI.O) for $68.7 billion in the biggest gaming industry deal in history as global technology giants stake their claims to a virtual future.

# Useful

- Provides a shallow semantic analysis (no modality, scope);
- generalises well across some languages;
- can benefit various NLP tasks (IR, QA).

**How much did Microsoft pay for Activision Blizzard?**

Microsoft Corp (MSFT.O) is buying "Call of Duty" maker Activision Blizzard (ATVI.O) for $68.7 billion in the biggest gaming industry deal in history as global technology giants stake their claims to a virtual future.



# Subcategorisation, Arguments and Adjuncts

# Subcategorisation

Fact 1: Verbs require a fixed configuration of required participants in the actions they denote:

(1) a. *I baked a cake*.

b. It is *raining*. ▷expletive

c. *I bet you five dollars I can spit further than you*.

(2) John ate the steak.

(3) a. John devoured the steak.

b. *John devoured.

(4) a. I dined.

b. *I dined pizza.

# python decorator

```
    @property
    @cache
    def stdev(self):
        return statistics.stdev(self._data)
```

# Argument vs adjunct (1)

Fact 2: There are also some optional participants (These can sometimes look surprisingly similar):

(5) a. I *waited* *for hours*.

b. I *waited* *for the bus*.

c. I *waited* *for the bus* *for hours*.

*ARGUMENT*; *ADJUNCT*

## Repetition of Adjuncts and Arguments

(6) a. I was waiting for the bus in July in the afternoon for a long time between church and tea at a bus station in the cold wind.

b. *I gave the cake to Kim to Sandy.

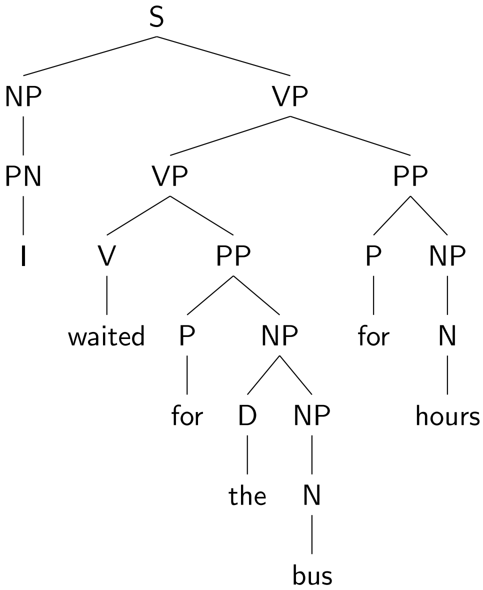
Arguments of the same kind can be applied exactly once in a clause and no more, whereas adjuncts of the same kind can be repeatedly applied in the same clause.

# Argument vs. Adjunct (2)

## Some Tests

- Adjuncts can be iterated and reordered. (Arguments can't).
- Arguments are located next to the head (in their canonical form); they can't be put anywhere else.
- When we want to use coordination on arguments and adjuncts, we can coordinate arguments with arguments and adjuncts with adjuncts, but we cannot mix the two.

# A syntactic tree



# Adjuncts are regular; arguments are idiosyncratic

Rule: everything that is lingustically idiosycratic needs to be remembered together with the lexical item.

Remembering:

- For instance, when a child learns the concept.
- We call items that need to be remembered with the lexical item *listemes*.
- We might sometimes use the phrase "it needs to be put into the lexicon".
- The "lexicon" is an abstract theoretical concept here.

Remembering is not necessary for regular combinations such as adjuncts.

# Uniform vs freely chosen clothes

# Not only verbs subcategorise

(7) a. the *book* *of poems* *with a brown cover*

b. *the *book* *of poems* *of fiction*

(8) a. *I* am *envious* of *you*

b. *I* am *angry* at/with *you*.

# Nominalisation and semantic roles

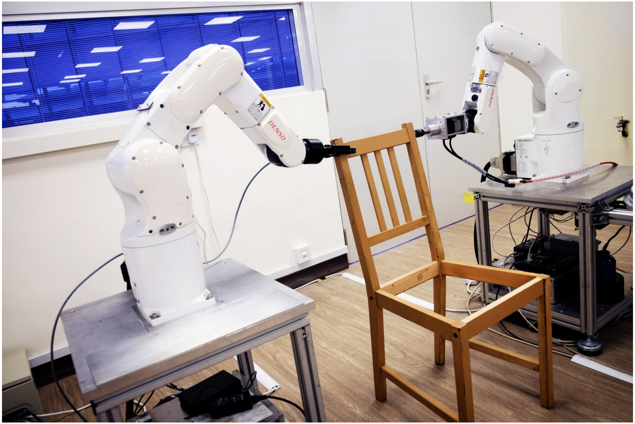
(9) a. abuse of cleaning lady

b. abuse by boyfriend/manager

c. abuse of boyfriend/manager

d. cleaning lady's abuse by boyfriend/manager

e. boyfriend/manager's abuse of cleaning lady




# Semantic Role Labeling

# Semantic Role Labeling

## Semantic role labeling (SRL)

- identifies arguments as well as adjuncts (harder)
- label their semantic roles (easier)

Supervised learning provides popular solutions.

## Constituency-based SRL

- CoNLL shared task 2004 and 2005
- AllenNLP Demo:
  `https://demo.allennlp.org/semantic-role-labeling`

## Dependency-based SRL

- CoNLL shared task 2008 and 2009

# VerbNet and Unified Verb Index for English

`https://uvi.colorado.edu`

## VerbNet: An extension of Levin (1993)

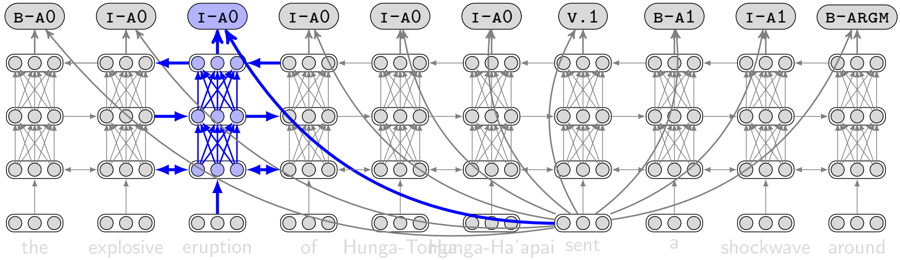
- Actor
- Agent
- Beneficiary
- Theme
- etc.

## PropBank: Annotations of semantic roles

- Arg0/A0: proto-Agent
- Arg1/A1: proto-Patient
- Arg2–6: verb-specific roles
- ArgM-Manner: adjuncts
- ArgM-...

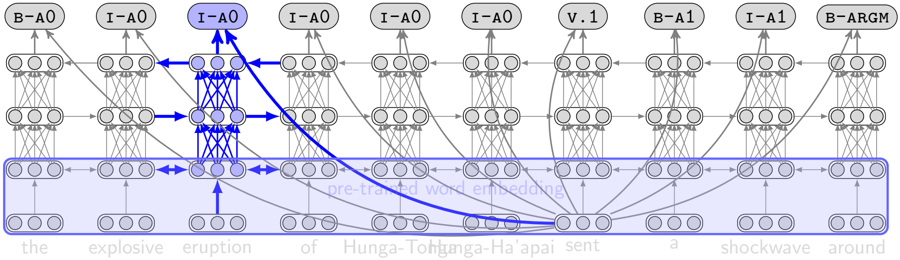
**PropBank is based on the Penn TreeBank trees**

# Syntax-agnostic SRL



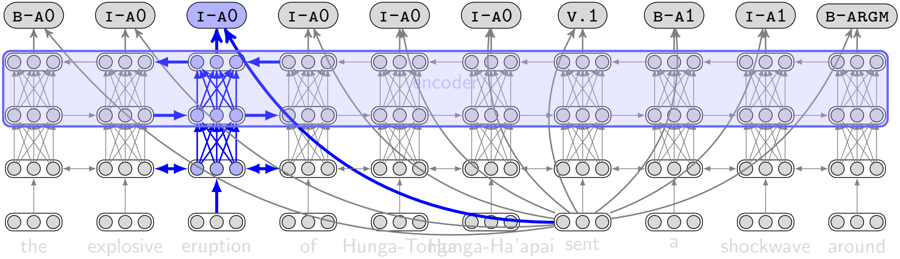

- Sequence labelling task
- One sequential classification per predicate
- Input: embeddings representing each word
- Pretrained language models, e.g. BERT
- Classification by LSTM, transformer…
- Output: BIO+ semantic labels

# Syntax-agnostic SRL



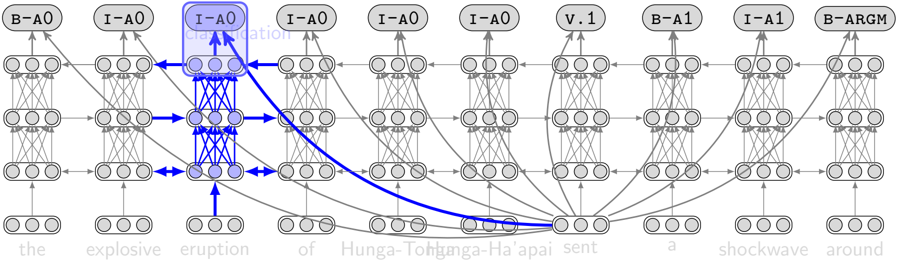

- Sequence labelling task
- One sequential classification per predicate
- Input: embeddings representing each word
- Pretrained language models, e.g. BERT
- Classification by LSTM, transformer...
- Output: BIO+ semantic labels

# Syntax-agnostic SRL




- Sequence labelling task
- One sequential classification per predicate
- Input: embeddings representing each word
- Pretrained language models, e.g. BERT
- Classification by LSTM, transformer. . .
- Output: BIO+ semantic labels

# Syntax-agnostic SRL




- Sequence labelling task
- One sequential classification per predicate
- Input: embeddings representing each word
- Pretrained language models, e.g. BERT
- Classification by LSTM, transformer...
- Output: BIO+ semantic labels

! You see this architecture everywhere

# Multilingual SRL

## Independently developed semantic resources

- German: SALSA, German FrameNet
- Chinese: Chinese PropBank, NomBank

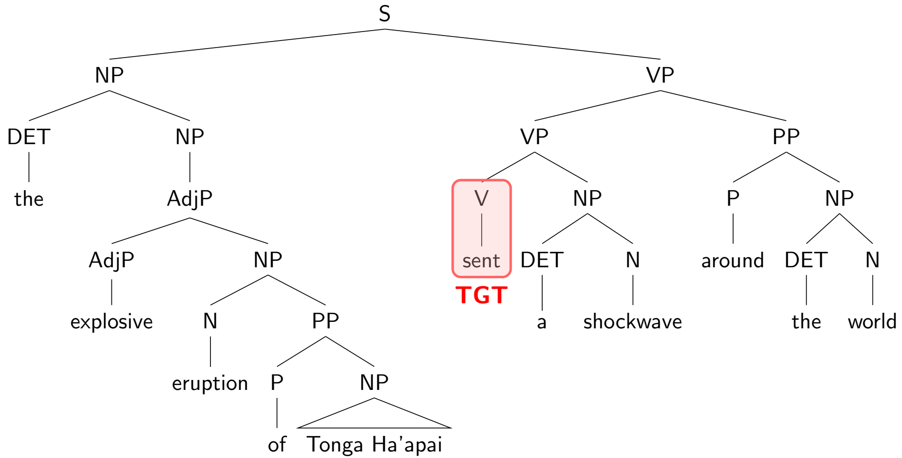
## Translation and annnotation projection

Angel Daza and Anette Frank. X-SRL: A Parallel Cross-Lingual Semantic Role Labeling Dataset

> *We apply high-quality machine translation to the English CoNLL-09 dataset and use multilingual BERT to project its high-quality annotations to the target languages. We include human-validated test sets that we use to measure the projection quality, and show that projection is denser and more precise than a strong baseline.*

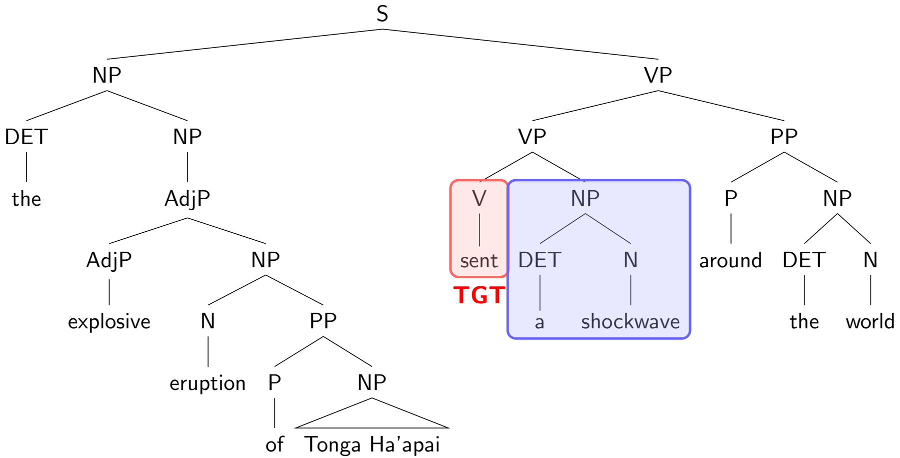
# Sisters, Aunts, Great-Aunts, . . .

# Where are arguments/adjuncts?



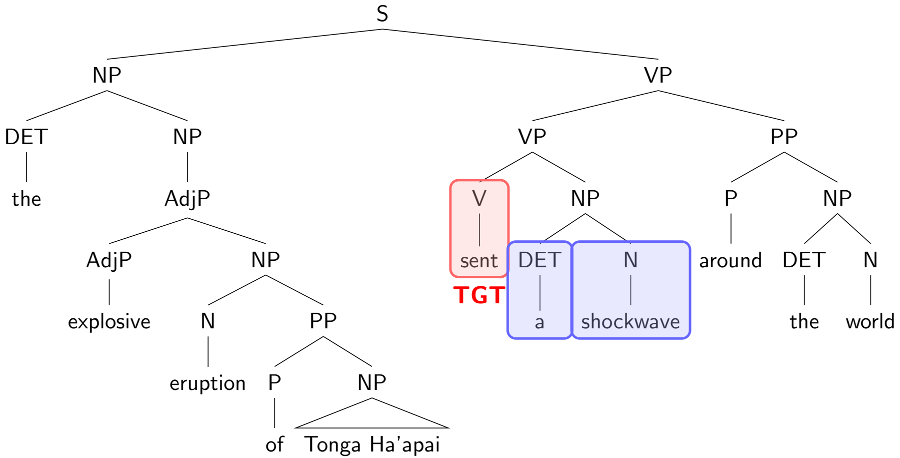

- Arguments/adjuncts should **c-command** a target verb.
- A node in a syntactic tree c-commands its sister node and all of its sister's descendants

# Where are arguments/adjuncts?



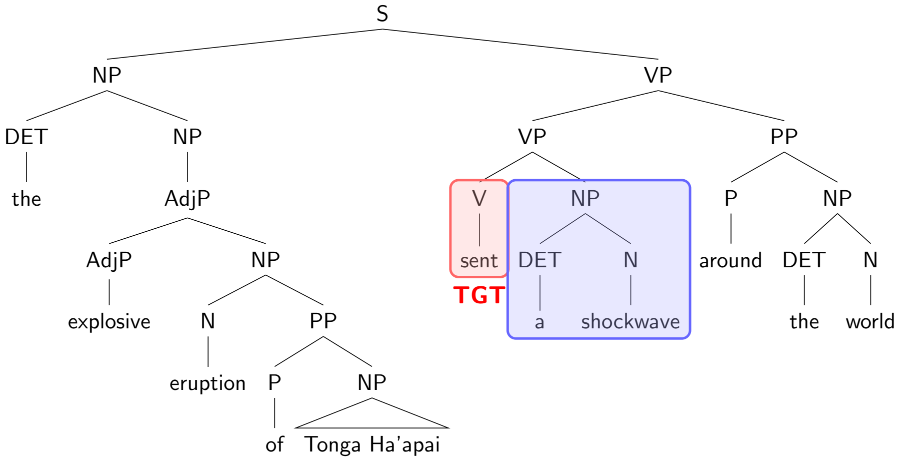

- Arguments/adjuncts should **c-command** a target verb.
- A node in a syntactic tree c-commands its sister node and all of its sister's descendants

# Where are arguments/adjuncts?



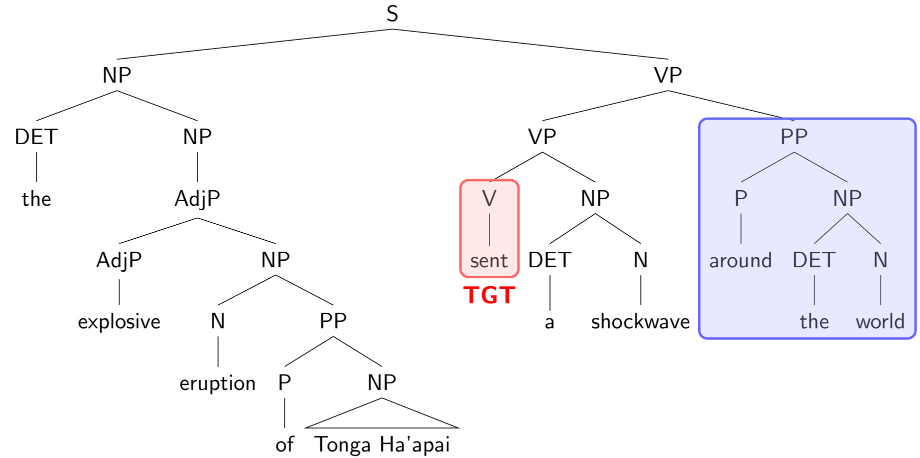

- Arguments/adjuncts should **c-command** a target verb.
- A node in a syntactic tree c-commands its sister node and all of its sister's descendants

# Where are arguments/adjuncts?



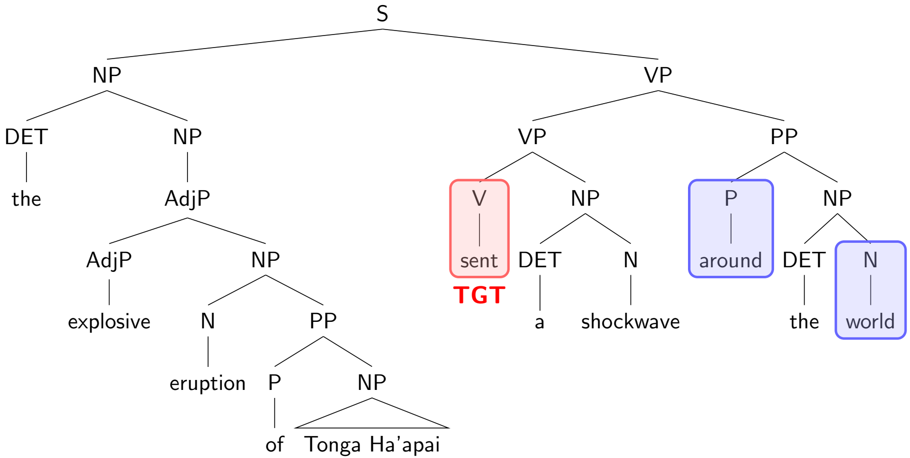

- Arguments/adjuncts should **c-command** a target verb.
- A node in a syntactic tree c-commands its sister node and all of its sister's descendants

# Where are arguments/adjuncts?



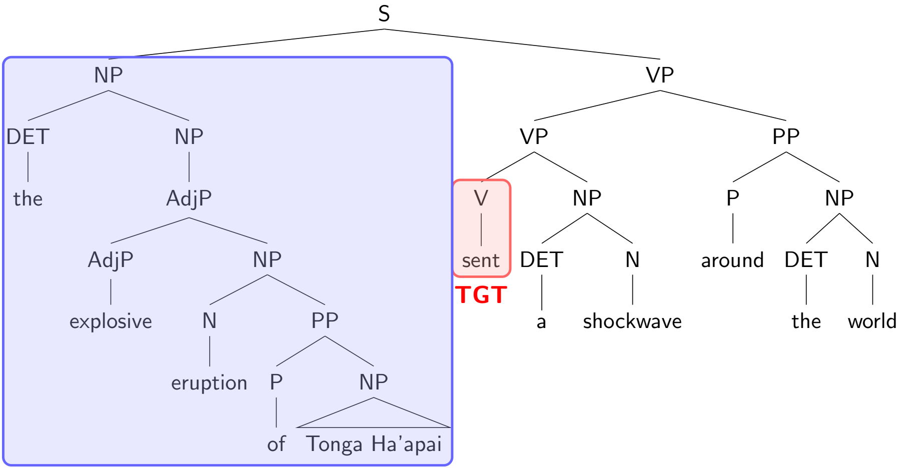

- Arguments/adjuncts should **c-command** a target verb.
- A node in a syntactic tree c-commands its sister node and all of its sister's descendants

# Where are arguments/adjuncts?



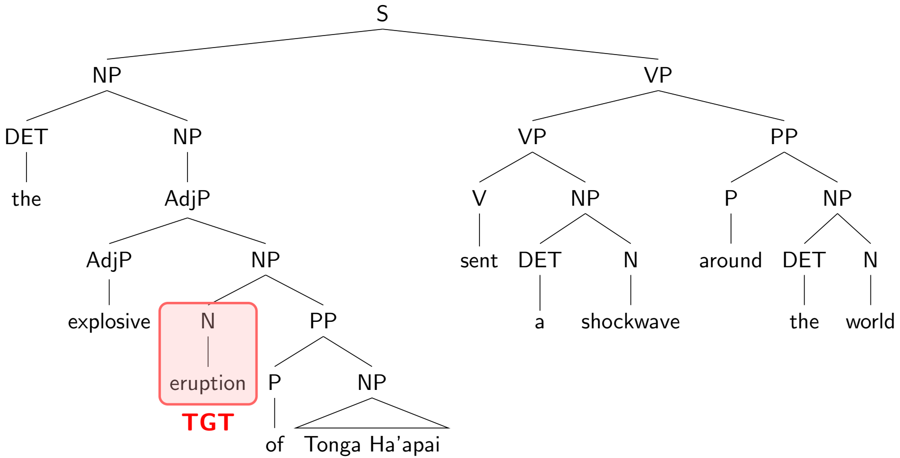

- Arguments/adjuncts should **c-command** a target verb.
- A node in a syntactic tree c-commands its sister node and all of its sister's descendants

# Where are arguments/adjuncts?




- Arguments/adjuncts should **c-command** a target verb.
- A node in a syntactic tree c-commands its sister node and all of its sister's descendants

# Where are arguments/adjuncts?




- Arguments/adjuncts should **c-command** a target verb.
- A node in a syntactic tree c-commands its sister node and all of its sister's descendants

# Syntax-based SRL

## Two-step pipeline

- Step 1: Employ a parser to find all relevant phrases/constituents/spans
- Step 2: Classify them
- Step 3*: Joint inference over all predicted semantic roles

## Challenges

- Error propagation due to syntactic parsing
- Encoding trees, which are complex discrete structures


IKEAing annotations

## from QA-SRL https://qasrl.org

*Our goal is to advance the state of the art in broad-coverage natural language understanding. We believe the way forward is with new datasets that are:*

- **Crowdsourced**: *modern machine learning methods require big training sets, which means scalability is a top priority.*
- **Richly structured**: *in order to improve over powerful representations learned from unlabeled data, we need strong, structured supervision signal.*
- **Extensible**: *annotation schemas should be flexible enough to accommodate new semantic phenomena without requiring expensive rounds of reannotation or brittle postprocessing rules.*

*[...] The common feature between our projects is* **using natural language to annotate natural language**. *This results in interpretable structures that can be annotated by non-experts at scale, which have the further advantage of being agnostic to choices of linguistic formalism.*

# Crowdsourcing

# Extensible

*annotation schemas should be flexible enough to accommodate new semantic phenomena without requiring expensive rounds of re-annotation or brittle postprocessing rules.*

Nothing as practical as a good theory!