PropBank

Topic-wise study notes | Natural Language Processing

Predicate-argument annotation, semantic role labeling, and argument labels ARG0-ARG5. Includes definitions, characteristics, architectures, worked examples, and applications.

1. Definition and Overview

PropBank (Proposition Bank) is a semantic annotation resource and framework that records predicates and the roles of their arguments. The original English resource added predicate-argument information to Penn Treebank syntactic analyses [1].

Predicate: an expression describing an event, action, or state.
Argument: a participant or complement associated with that predicate.
Semantic role: the participant's relation to the predicate, such as giver, thing given, or recipient.

Example: "Ravi gave Meena a book." The predicate is gave; its lemma is give. Ravi is the giver, a book is the thing given, and Meena is the recipient.

Characteristics

• Connects semantic labels to specific predicates and senses.

• Uses numbered argument labels and modifier labels.

• Preserves role relationships across different syntactic constructions.

• Annotates each predicate separately when a sentence contains several.

• Uses frame files to document role meanings and examples.

• Provides labelled data for training and evaluating semantic role labeling systems.

Resource versus task

PropBank	Semantic role labeling
An annotation scheme, frame lexicon, and annotated resources.	A task that predicts predicate-argument structures in text.
Defines the labels and predicate-specific role inventories.	Uses those definitions to identify and label arguments.

The original resource emphasized verbs; later PropBank work includes other predicate types. The examples here focus on verbal predicates for clarity.

2. Predicate-Argument Annotation

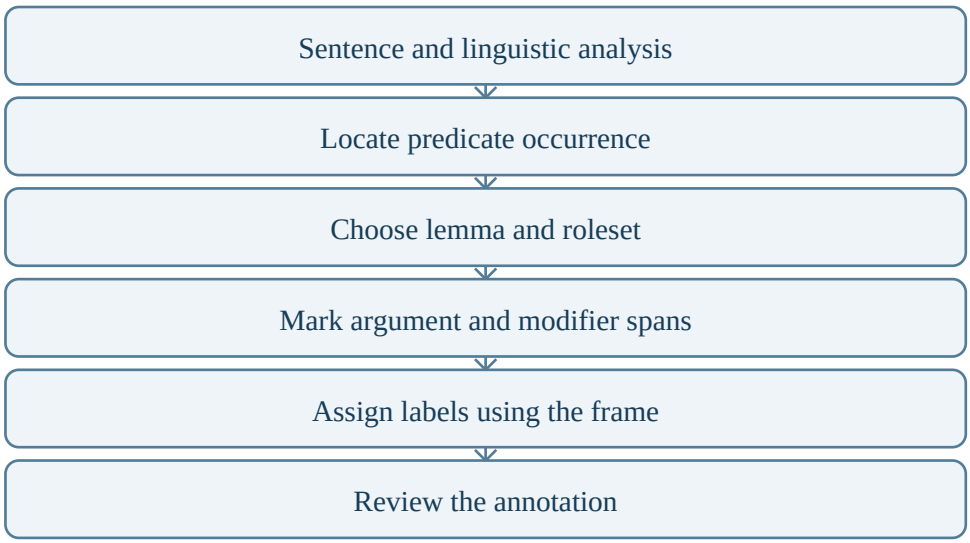
Definition

Predicate-argument annotation identifies a predicate occurrence, selects its sense or roleset, marks its argument spans, and assigns labels according to the chosen frame. Modifier spans can also be labelled.

Characteristics

- **Predicate-centred:** labels are assigned relative to one predicate at a time.
- **Sense-sensitive:** different senses can have different role inventories.
- **Span-aware:** the boundaries of phrases matter, not just their head words.
- **Syntax-independent meaning:** an argument can retain its label when voice changes.
- **Frame-guided:** role definitions come from the selected roleset.

Annotation architecture



Step-by-step example

Sentence: “Ravi gave Meena a book yesterday.”

- **1:** Locate gave as the predicate.
- **2:** Lemmatize gave to give.
- **3:** Choose give.01, the transfer sense [2].
- **4:** Identify the giver Ravi, thing given a book, and recipient Meena.
- **5:** Assign Ravi = ARG0; a book = ARG1; Meena = ARG2.
- **6:** Label yesterday as ARGM-TMP, a temporal modifier.

Annotated form: [ARG0 Ravi] [V gave] [ARG2 Meena] [ARG1 a book] [ARGM-TMP yesterday].

Applications

Annotation produces training corpora, supports linguistic comparison, and supplies structured event information for downstream extraction and question answering.

3. Semantic Role Labeling

Definition

Semantic role labeling (SRL) automatically identifies predicates and labels the associated participants and modifiers. It aims to represent “who did what to whom”, along with circumstances such as when and where.

Characteristics

• Predicts semantic relations rather than only grammatical categories.

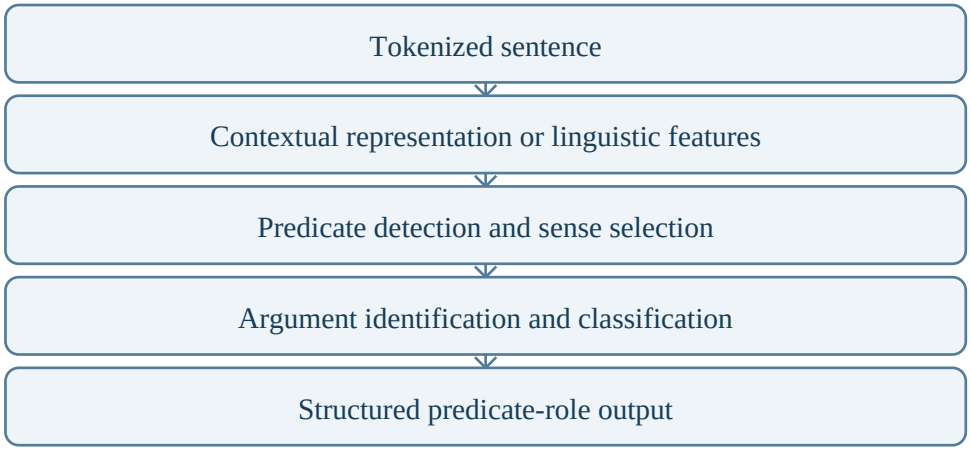
• Can use token spans or dependency-head representations.

• May predict predicate senses as well as role labels.

• Uses contextual information to identify argument boundaries and roles.

• Can be implemented with rules, statistical models, or neural models.

Illustrative automatic architecture



Systems may receive predicates as input or predict them; stages may be joint.

Step-by-step example

Input: “Ravi gave a book to Meena yesterday.”

Step	Result
Predicate detection	gave
Lemma and sense	give.01, transfer
Argument spans	Ravi; a book; to Meena
Role classification	ARG0; ARG1; ARG2
Modifier classification	yesterday → ARGM-TMP
Output record	give.01 with its labelled argument spans.

The recipient span includes to. Phrase position alone does not determine the role.

Applications

SRL supports event extraction, semantic search, question answering, summarization, and comparison of paraphrases. It supplies useful structure but does not by itself establish factual truth.

4. Argument Labels ARG0-ARG5

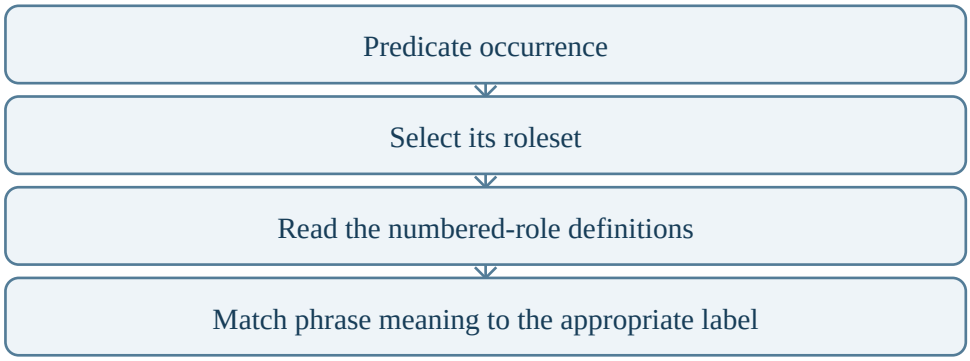
Definition

ARG0-ARG5 are numbered role identifiers. Their meanings are defined by the selected predicate roleset. They are not six universal semantic categories, and a sentence does not need to contain all six.

Characteristics and interpretation

Label	How to interpret it
ARG0	Often agent-like or initiator-like; consult the frame. Not automatically the grammatical subject.
ARG1	Often patient/theme-like or another central participant; consult the frame.
ARG2	An additional frame-defined role: recipient for give.01, seller for buy.01, destination for move.01.
ARG3	Another frame-defined role: price for buy.01; not always a source.
ARG4	Another frame-defined role: beneficiary for buy.01; not always a destination.
ARG5	A less common additional numbered role, used only when the roleset defines it. No universal meaning.

Architecture for label interpretation



Key examples

give.01: ARG0 = giver; ARG1 = thing given; ARG2 = recipient [2].
buy.01: ARG0 = buyer; ARG1 = purchase; ARG2 = seller; ARG3 = price; ARG4 = beneficiary [3].
move.01: ARG2 identifies the destination [4].

Important: A recipient is not always ARG2 across every predicate. ARG3 is not always source, ARG4 is not always destination, and ARG5 should not be invented to fill an unused slot.

Applications

Numbered roles make annotation compact and support consistent labeling within a predicate sense. Cross-predicate comparison requires interpreting each frame, rather than comparing numbers alone.

5. Worked Example: ARG0-ARG4

Sentence: “Ravi bought a laptop from Meena for 40,000 rupees for his daughter yesterday.”
Predicate: bought; lemma buy; roleset buy.01, purchase [3].

Step-by-step annotation

Step	Phrase	Role
1. Find buyer	Ravi	ARG0
2. Find purchased item	a laptop	ARG1
3. Find seller	from Meena	ARG2
4. Find price	for 40,000 rupees	ARG3
5. Find beneficiary	for his daughter	ARG4
6. Find time	yesterday	ARGM-TMP

Reasoning for each label

• Ravi performs the purchase, so the frame’s buyer role is ARG0.

• A laptop is the purchased item, so it receives ARG1.

• Meena supplies the item as seller, so from Meena receives ARG2.

• The monetary amount is the price, so it receives ARG3.

• The daughter benefits from the purchase, so for his daughter receives ARG4.

• Yesterday gives time rather than a numbered participant role.

Annotated form:
[ARG0 Ravi] [V bought] [ARG1 a laptop] [ARG2 from Meena] [ARG3 for 40,000 rupees] [ARG4 for his daughter] [ARGM-TMP yesterday].

What about ARG5?

The buy.01 frame shown here defines ARG0-ARG4. There is no ARG5 to assign in this example. Use ARG5 only for a predicate sense that explicitly defines that role.

Useful distinction

The same preposition for introduces both price and beneficiary here. A preposition alone does not determine the semantic role; the predicate frame and phrase meaning are needed.

6. Modifiers and Voice Alternation

ARGM: modifier labels

ARGM labels describe circumstances associated with a predicate. Numbered versus modifier status is frame-guided; it is not simply a distinction between grammatically mandatory and optional phrases.

Label	Meaning	Example phrase
ARGM-TMP	Time	yesterday
ARGM-LOC	Location	in the classroom
ARGM-MNR	Manner	carefully
ARGM-CAU	Cause	because of the rain
ARGM-NEG	Negation	not
ARGM-MOD	Modality	might

Location versus frame-defined destination

In “Ravi gave Meena a book in the classroom”, in the classroom can be ARGM-LOC. In “Ravi moved the box to the classroom”, the destination to the classroom is ARG2 under move.01 [4]. Both mention a place, but their role in the event differs.

Active and passive: step-by-step example

Active: “Ravi gave Meena a book.”
Passive: “A book was given to Meena by Ravi.”

Semantic participant	Active phrase	Passive phrase	Label
Giver	Ravi	by Ravi	ARG0
Thing given	a book	A book	ARG1
Recipient	Meena	to Meena	ARG2

- **1:** Choose the same transfer sense, give.01.
- **2:** Identify participant meanings in both sentences.
- **3:** Keep the semantic labels even though grammatical positions change.
- **4:** Notice that the passive subject A book is ARG1, not ARG0.

A numbered role may be absent from a sentence. For example, “A book was given to Meena” does not explicitly express the giver. Do not fabricate an ARG0 span.

7. Multiple Predicates and Workflow

One sentence, separate predicate analyses

Sentence: “Ravi bought a book and gave it to Meena.”

Predicate	Role links
bought / buy.01	ARG0: Ravi; ARG1: a book.
gave / give.01	ARG0: Ravi; ARG1: it; ARG2: to Meena.

The second clause has a shared subject understood through coordination; exact corpus encoding follows its guidelines. The table shows the semantic links. The pronoun it remains the expressed argument of gave; linking it to a book is a separate coreference task.

Step-by-step general annotation workflow

• **1. Read:** understand the sentence and context.

• **2. Identify:** find predicate occurrences.

• **3. Disambiguate:** choose the lemma and appropriate roleset for each.

• **4. Consult:** read the frame-specific role inventory.

• **5. Delimit:** identify the full argument spans.

• **6. Label:** assign numbered roles and modifiers.

• **7. Validate:** check frame compatibility and phrase boundaries.

• **8. Record:** store each predicate’s annotation separately.

Simple annotation record

Predicate token: gave
Roleset: give.01
ARG0: Ravi
ARG1: it
ARG2: to Meena

Manual annotation versus automatic prediction

Human annotation uses guidelines and review to create reference labels. Automatic SRL predicts labels using learned or rule-based methods. Evaluation compares predictions with reference annotations, accounting for the chosen representation and task setup.

8. Applications and Limitations

Application	Example use
Question answering	Who gave the book? → ARG0 of give.01.
Event extraction	Create records for buyer, item, seller, and price.
Semantic search	Find giving events even across active/passive wording.
Summarization	Preserve main predicate-participant relationships.
Paraphrase comparison	Compare semantic roles across different sentence forms.
Information organization	Support event or relation records for downstream systems.
NLP research	Train and evaluate SRL models using annotated examples.

Limitations and common errors

- The wrong predicate sense can lead to the wrong role assignments.
- Argument span boundaries may be ambiguous or structurally complex.
- Long-distance dependencies and coordination can complicate annotation.
- Domain and language changes can reduce automatic SRL performance.
- ARG0 is not always the subject; ARG1 is not always the object.
- Numbered roles must be interpreted through frames.
- Coreference, factuality, and full discourse reasoning require additional analysis.

Evaluation concepts

For span-based SRL, a common evaluation requires the predicted span boundaries and role label to match the reference. Precision measures the fraction of predicted arguments that are correct; recall measures the fraction of reference arguments recovered; F1 combines them. Predicate identification and sense prediction may be assessed separately.

Study checklist

For every example, ask: What is the predicate? Which sense is intended? What roles does the frame define? Which spans realize them? Which phrases are modifiers? Is any role unexpressed?

9. Revision and References

Practice questions with answer hints

1. Define PropBank. A resource and annotation framework for predicate-argument structures.
2. Define SRL. Predicting predicates and their labelled arguments and modifiers.
3. Is ARG2 always a recipient? No; it depends on the roleset.
4. Annotate "Ravi gave Meena a pen". Ravi ARG0; Meena ARG2; a pen ARG1; predicate give.01.
5. Why is the passive subject not automatically ARG0? Semantic roles describe participant relations, not grammatical position.
6. Must ARG5 appear? No. Use it only if the selected frame defines it and the sentence expresses that role.

Short exam answer

PropBank annotates predicates with sense-specific argument structures. Numbered labels ARG0-ARG5 denote roles defined in a predicate's frame, while ARGM labels describe modifiers. Semantic role labeling automatically predicts these structures. ARG0 is often agent-like and ARG1 often patient/theme-like, but role meanings, particularly higher-numbered arguments, must be checked in the selected roleset.

References and verified frames

[1] PropBank project overview.
[2] PropBank frame: give.01.
[3] PropBank frame: buy.01.
[4] PropBank frame: move.01.
[5] Official PropBank documentation and annotation guidelines.

Palmer, M., Gildea, D., and Kingsbury, P. (2005). *The Proposition Bank: An Annotated Corpus of Semantic Roles*. Computational Linguistics, 31(1), 71-106.

Frame examples use the linked version 3.4.0 pages. Sentences, architectures, and workflows are original teaching examples. Detailed corpus formats, special labels, and shared or discontinuous arguments should follow the relevant release guidelines.