

Guidelines for Propbank framers

Olga Babko-Malaya
September, 2005

1. Introduction	1
2. Selecting Examples	4
3. Distinguishing Framesets	5
3.1. Criteria for distinguishing framesets	5
3.2. Light Verbs	8
4. Numbering Arguments	9
4.1. Arg0 versus Arg1	9
4.2. Arg0 versus ArgA	11
4.3. Trends in Argument Numbering	11
4.4. Symmetric verbs	12
4.5. Psychological Verbs	14
5. Args and ArgMs.	14

1. Introduction

Frame files provide guidelines for Propbank annotators and include a list of framesets, or coarse-grained senses of the verbs. A frameset stands for a set of syntactic frames. Following Levin 1993, we assume that the set of syntactic constructions or frames that a verb can occur in is a direct reflection of the underlying semantic components that restrict allowable arguments. A frameset thus corresponds to a coarse-grained sense of the verb, which has a specific set of semantic arguments.

A description for each frameset includes the list of verb specific roles and examples of different syntactic realizations of the verb. The general procedure is to examine a number of sentences from the corpus and then select the roles which seem to occur most frequently and/or are semantically necessary. These roles are then numbered sequentially from Arg0 up to (potentially) Arg5, and each role is given a mnemonic label. An example of a frameset for the verb 'give' is given below:

Frameset give.01 "transfer":

Roles:

Arg0:giver

Arg1: thing given
Arg2: entity given to

Examples:

double object

The executives gave the chefs a standing ovation.

Arg0: The executives

REL: gave

Arg2: the chefs

Arg1: a standing ovation

ditransitive

She had given the answers to two low-ability geography classes.

Arg0: She

REL: had given

Arg1: the answers

Arg2-to: to two low ability geography classes

give way

The Beatles give way to baseball in the Nipponese version.

Arg0: The Beatles

REL: give

Arg1: way

Arg2-to: to baseball

discourse-level usage

Given the weakness in both the junk bond market and the stock market, traders fear that these transactions may be revised yet again.

REL: given

Arg1: the weakness in both the junk bond market and the stock market

Arg2: traders

For some verbs, it is impossible to provide one set of semantic roles for all senses of the verb. For example, the two senses of the verb 'leave' in the examples below take different arguments:

Mary left the room

Mary left her daughter-in-law her pearls in her will

In such cases, frame files distinguish two or more framesets, and define argument labels specific to each frameset:

*Frameset **leave.01** "move away from":*

Arg0: entity leaving

Arg1: place left

Frameset leave.02 "give":

Arg0: giver

Arg1: thing given

Arg2: beneficiary

In some cases, frame files define not only several framesets for each verb, but also several predicates. If a verb has a particle (marked as PRT in TreeBank), then it is being considered as a different predicate, and has a different set of semantic roles. For example, the frame file for the verb 'keep' defines three predicates: predicate 'keep' (which has 3 framesets), and predicates 'keep_up' and 'keep_on'. The following example illustrates the definition of the predicate 'keep_up'. Note that the relation (REL) in PB annotation includes both the verb and the particle. If the verb and the particle can be selected as one node, then this node is being marked as REL, otherwise, if one node is not available in Treebank, then the concatenated constituent (i.e. [keep][up]), should be labeled as REL.

Predicate keep_up:

keep.05 "keep up: maintain position":

Arg0: maintainer of position

Arg1: relative to what

John can't keep up with Mary's rapid mood swings.

Arg0: John

ArgM-MOD: ca

ArgM-NEG: n't

REL: keep up

Arg1: with Mary's rapid mood swings

These guidelines are intended to address different issues that may arise in the process of creating frame files, which include the following ones:

Can we ignore sense distinctions and assign one frameset to each verb?

Should argument labels have a specific semantic content?

How are arguments numbered? For example, should all framesets start with Arg0?

Can framesets be consistent across verb classes?

Can the same label be assigned to two different arguments?

How to distinguish arguments (Args) from modifiers (ArgMs)?

However, before we turn to these difficult issues, let us first describe the process of framing, which includes the following steps:

- examine examples from the corpus and annotate selected sentences as examples
- make a decision on how many framesets the verb has
- define each frameset

All these tasks are discussed in detail below.

2. Selecting Examples

While creating frame files, framers usually start by looking at the examples from the corpus and annotating those sentences which could be included as examples.

Frame files provide verb-specific description of all possible semantic roles, as well as illustrate these roles by examples of all possible syntactic constructions. In the case of alternating verbs, the examples should include all possible alternating structures. It is also important that examples illustrate all possible semantic roles. For example, in the case of the verb 'open' below, examples include a transitive/agentive sentence, intransitive/unaccusative one, as well as sentences with an instrument and beneficiary roles.

Frameset open.01 "open":

Roles:

Arg0: opener

Arg1: thing opening

Arg2: instrument

Arg3: beneficiary

Examples:

agentive (-)

Texas Instruments Inc. opened a plant in South Korea to manufacture control devices.

Arg0: Texas Instruments Inc.

REL: opened

Arg1: a plant

ArgM-LOC: in South Korea

ArgM-PNC: to manufacture control devices.

ergative (-)

The branch of the the Bank for Foreign Economic Affairs opened in July.

Arg1: The branch of the the Bank for Foreign Economic Affairs

REL: opened

ArgM-TMP: in July

instrumental (-)

Douglas Wilder opened his gubernatorial battle with Marshall Coleman with an abortion commercial.

Arg0: Douglas Wilder

REL: opened

Arg1: his gubernatorial battle with Marshall Coleman

Arg2-with: an abortion commercial

topicalized instrument (-)
This can opener opens bottles, too!
Arg2: This can opener
REL: opens
Arg1: bottles
ArgM-DIS: too

benefactive (-)
South Korea has opened its market to foreign cigarettes.
Arg0: South Korea
REL: opened
Arg1: its market
Arg3-to: foreign cigarettes

In many cases, it is also helpful to add passive and other syntactic structures which occur frequently in the corpus.

3. Distinguishing Framesets

After examining different sentences in a corpus, the question which should be addressed is whether the verb has one or more framesets.

3.1. Criteria for distinguishing framesets

The main principle for distinguishing framesets is that two verb meanings are distinguished as different framesets if they have distinct subcategorization frames.

For example, different framesets might have a different number of arguments, or the number of arguments might remain the same, but thematic roles are crucially different, as illustrated by the verb 'draw' below:

Frameset draw.01 'art'
Arg0:artist
Arg1:art
Arg2:beneficiary

He (Arg0) was drawing diagrams and sketches (Arg1) for his patron (Arg2)

Frameset draw.02 'pull'
Arg0:puller
Arg1:thing pulled
Arg2: source

The campaign (Arg0) is drawing fire (Arg1) from anti-smoking advocates (Arg2)

Different senses are distinguished as different framesets only if they have distinct syntactic behavior, which correlates with different types of allowable arguments. The distinctions made by Framesets are thus very coarse, and usually correspond to several standard dictionary entries for the lemma in question.

It is also important to note that not all roles must necessarily be present in each sentence. For example, the Frameset for the verb *serve*, shown below has three roles: Arg0, Arg1, and Arg2. Actual usages of the verb, on the other hand, do not require the presence of all three roles, either explicitly or implicitly. For example, the sentence below does not imply the existence of a job/project role.

Frameset serve.01 "act, work":

Arg0: worker

Arg1: job, project

Arg2: employer

*Each new trading roadblock is likely to be beaten by institutions seeking better ways *trace* to serve their high-volume clients.*

*Arg0: *trace* -> institutions*

REL: serve

Arg2: their high-volume clients

Most verbs in Propbank have one frameset, which generalizes over different senses of the verb. Only 20% of the verbs have more than one frameset. The strategy is to distinguish framesets only if it is impossible to provide a list of generally defined arguments which would apply to all the senses of the verb in question.

Verb senses are NOT distinguished as different framesets in the following cases:

1. Framesets are not distinguished in the case of argument structure alternations, such as unspecified object deletion or causative/inchoative. For example, the transitive and intransitive uses of the verb 'open' are analyzed as one frameset.

John (Arg0) opened the door (Arg1)

The door (Arg1) opened

Frameset open.01: cause to open

Arg0: agent

Arg1: thing opened

Arg2: instrument

Some of the tests which could be used when making a decision whether to distinguish framesets are as follows:

Entailment test: Verb senses are not distinguished if one sense entails the other one.

In the example above, the transitive sentence with the verb ‘open’ above entails the intransitive one.

The two senses are not distinguished if the set of roles for one sense is a subset of the set of roles for another sense.

In the case of the verb ‘open’, the agent argument is present only for the transitive use of the verb, so that the set of roles for the intransitive usages is a subset of the set of roles for the transitive one.

2. Framesets are not distinguished in the case of *Manner of Motion/Directed Motion* alternation. Verbs of manner of motion like ‘run’ or ‘walk’ are aspectually different from the verbs of directed motion such as ‘run to the store’ or ‘walk home’: manner of motion verbs are activities, whereas verbs of directed motion are accomplishments and imply a change of location. However, aspectual differences are not being distinguished as different framesets.

This analysis can be supported by the entailment test discussed above: ‘John walked to the store’ entails that ‘John walked’. Also, the set of roles for the verbs of manner of motion is a subset of the set of roles for the verbs of directed motion, since only verbs of directed motion include a direction (or Arg2) argument.

Frameset pull.01 ‘(try) to cause motion’:

Arg0: puller

Arg1: thing pulled

Arg2: direction or attribute of arg1

John (Arg0) pulled a sled (Arg1).

The van (Arg1) pulled up(Arg2).

3. And, finally, framesets are not distinguished if the two verb senses differ in the semantic type of the event, such as physical, concrete, mental, emotional, etc. For example, the physical sense of the verb ‘strike’ and the ‘attract’ sense are not distinguished as different framesets, since they have similar semantic types of arguments (such as a causer and a patient):

Frameset strike.01 ‘hit’

Arg0: causer

Arg1: thing hit

hit sense:

He (Arg0) struck the table (Arg1)

attract sense:

His behavior (Arg0) struck me (Arg1)

3.2. Light Verbs

An exception to the rules above are light verbs like ‘take a walk’ or ‘keep pace’. Light verbs are analyzed as having an argument structure parallel to ‘action’ verbs, so that ‘a walk’ and ‘pace’ are marked as Arg1. On the other hand, it is important to separate these idiomatic or light verb senses of the verbs for many applications. Therefore, light verb usages are being distinguished in Propbank by assigning them to a special frameset, even if their argument structure is not different from the other frameset of the verb. This is illustrated by the verb ‘keep’ below:

Verb keep:

Frameset keep.01 "maintain possession":

Arg0:Keeper

Arg1:thing kept

The Herald(Arg0) kept its old-time Hearst readership(Arg1).

Frameset keep.02 "aspectual":

Arg0: causer of continued action

Arg1: continuing action or state

*John(Arg0) kept *trace* annotating(Arg1).*

Frameset keep.03 "maintain some prepositional relationship":

Arg0:causer

Arg1: noun

Arg2: prepositional phrase

Growers (Arg0) can't always keep the worm(Arg1) from the apple (Arg2).

*Japanese demand (Arg0) has kept the U.S. currency (Arg1) from *trace* plunging (Arg2).*

Britain (Arg0) kept light tanks (Arg1) out of the talks (Arg2).

Keep your Foster Savings Institution (Arg1) off the federal budget deficit (Arg2).

The Fed (Arg0) kept inflation-fighting (Arg1) as its top priority (Arg2).

Frameset Keep 04. 'light verb'

Arg0:agent

Arg1:noun

Arg2:prepositional object

Newsweek(Arg0) keeps pace(Arg1) with its rival Time(Arg2).

The state(Arg0) keeps track(Arg1) of the achievement-test preparation booklets(Arg2).

The verb ‘keep’ has four framesets. Whereas the first three are distinguished based on different semantic roles, the third and the fourth frameset do not really differ with respect to their semantic arguments. Nevertheless, they are distinguished as two different framesets. The reason for making this exception is to allow Propbank users to be able to extract sentences with idiomatic expressions, if needed.

This frame file also illustrates another exception, specifically, the definition of Arg2 for the third and the fourth frameset is defined not as having a specific semantic role, but rather syntactically, as a prepositional phrase. Whereas in general the roles should be defined semantically, in some cases exceptions can be made. For example, in the case of the verb ‘keep’, the meaning of the verb depends on the preposition (e.g. ‘keep from’ as opposed to ‘keep as’), so that it seems plausible to generalize these different senses and define Arg2 argument as a prepositional phrase.

4. Numbering Arguments

The next question which a framer needs to decide is how to number the arguments. Whereas the arguments are being numbered sequentially, the choice has to be made whether the first argument is Arg0, Arg1, or ArgA.

4.1. Arg0 versus Arg1

The Arg0 label is assigned to arguments which are understood as agents, causers, or experiencers. The Arg1 label is usually assigned to the patient argument, i.e. the argument which undergoes the change of state or is being affected by the action. Semantically, Arg0 and Arg1 correspond to Dowty’s Proto-Agent and Proto-Patient roles, as discussed in detail in Propbank Annotation Guidelines. Following these assumptions, unaccusative verbs like ‘rise’ do not have an Arg0 argument, and numbering starts with Arg1:

Unaccusative Verbs: *rise*

Roles

Arg1 = Logical subject, patient, thing rising

Arg2 = EXT, amount risen

Arg3* = start point

Arg4 = end point

A question which arises is why do we distinguish Arg0 and Arg1 in such a way, instead of simply numbering the arguments starting from Arg0 in all the cases. One of the

reasons has to do with the fact that some unaccusative verbs, like break or open, have both transitive and intransitive variants:

John(Arg0) opened the door (Arg1)
The door(Arg1) opened.

Since we analyze transitive and intransitive uses of such verbs as one frameset, the subject of the intransitive verb ‘open’ is labeled as Arg1. In order to make annotations consistent across similar verb classes, we have to assume that verbs like ‘die’, which do not have a transitive alternant, also mark their subjects as Arg1.

Verb: kill
ArgO: killer
Arg1: corpse

Verb: die

Arg1: corpse

Another reason for distinguishing unaccusative verbs from unergatives is that unaccusative verbs that do not generally occur in a transitivity alternation sometimes can be used transitively. During the course of Propbank annotation, there were several verbs that have been analyzed as unaccusatives by framers (i.e. without Arg0 argument), but annotators had to add Arg0 argument because of a few unexpected transitive usages of those verbs encountered in the corpus.

Although this distinction seems to be linguistically motivated and practically useful, in some cases the choice between Arg0 and Arg1 is not very straightforward. For example, Propbank analysis of aspectual verbs, like start or continue, assumes that they have 2 arguments: an agent and an event which is being started or continued. In the case of agentive sentences, annotators need to mark the subject as Arg0, whereas in the case of non-agentive ones, the subject is part of Arg1.

Frameset continue.01 "aspectual":
Arg0: causer of continuation
Arg1: thing continuing

agent and event

*Investors continue *trace* to pour cash into money funds.*
Arg0:Investors
REL:continue
Arg1:*trace* to pour cash into money funds

event, no agent

New loans continue [-1] to slow; they were \$6.6 million in the quarter compared with \$361.8 million a year ago.*
Arg1:[New loans][[*-1] to slow]
REL: continue

There are some other verbs of variable behavior with respect to unaccusativity, which

might require distinguishing agentive (or internally-caused) usages of the verbs from non-agentive (or externally-caused) ones, as discussed in more detail in Propbank Annotation Guidelines.

4.2. Arg0 versus ArgA

ArgA is an external causer argument, which is being added to the set of semantic roles ONLY IF there is another agent or a causer argument. Verbs with an ArgA argument are unergative verbs that are being used transitively, as illustrated by the verbs of motion below:

The general (ArgA) marched the soldiers (Arg0) to the tents.
The rider (ArgA) jumped the horse (Arg0) over the fence.
We (ArgA) ran the mouse (Arg0) through the maze.

/Levin and Rappaport Hovav 1995, p. 111/

The agent of marching or jumping is the object, rather than the subject in this case. Transitive verbs of motion do not occur frequently in the corpus, however, the following sentence is from the WSJ:

*Colleagues today recall with some humor how meetings would crawl into the early morning hours [*T*-1] as Mr. Dinkins would march his staff out of board meetings and into his private office [*-2] to discuss, en masse, certain controversial proposals .*

ArgA: Mr. Dinkins
ArgM-MOD: would
REL: march
Arg0: his staff
ArgM-DIR: out of board meetings and into his private office
ArgM-PNC: [*-2] to discuss , en masse , certain controversial proposals

There are only a few instances of ArgA arguments in the WSJ corpus. Our approach is to include ArgA to the list of semantic roles only if there are actual examples in the corpus. Currently, many verbs of motion do not have an ArgA argument. However, it is quite possible that in the course of future propbanking annotators will encounter transitive usages of unergative verbs, in which case ArgA should be added to the frame files.

4.3. Trends in Argument Numbering

Arg0, Arg1, and ArgA are the only labels which are associated with a certain specific semantic content. The other numbered arguments vary across different semantic classes, as shown below:

Arg0 = external argument (Proto-Agent)
Arg1 = internal argument (Proto-Patient)
Arg2 = indirect object / beneficiary / instrument / attribute / end state

Arg3 = start point / beneficiary / instrument / attribute
Arg4 = end point
ArgA = external causer

Although in general argument labels are specific to a particular verb, a reasonable effort was made to make them consistent across semantic classes, including Levin/VerbNet classes. The following set of roles, for example, describes financial verbs, such as ‘jump’, ‘rise’, ‘fall’, and others:

Arg1: Logical subject, patient
Arg2: EXT, amount risen/jumped/fallen
Arg3: start point
Arg4: end point

Sales rose 4% to \$3.28 billion from \$3.16 billion.

Arg1: Sales
REL: rose
Arg2-EXT: 4%
Arg3-to: to \$3.28 billion
Arg4-from: from \$3.16 billion

Cray Research stock jumped \$2.875 yesterday to close at \$38.

Arg1: Cray Research stock
REL: jumped
Arg2: \$2.875
ArgM-TMP: yesterday
Arg4: to close at \$38

Profits fell 10% to \$118 million from \$130.6 million

Arg1: Profits
REL: fell
Arg2-EXT: 10%
Arg4-to: to \$118 million
Arg3-from: from \$130.6 million

Different VerbNet/Levin classes are also analyzed as having identical set of roles. These classes are illustrated below by symmetric and psych verbs.

4.4. Symmetric verbs

There are a few verbs in Propbank, where two different arguments are being marked as Arg1. These are symmetric stative verbs ‘match’, ‘tie’, and ‘attach’. All these verbs have the following set of roles:

Frameset match.01

Arg0: person performing match

Arg1: matching objects

The architect matched the paint and the wallpaper

Arg0: The architect

REL: matched

Arg1: the paint and the wallpaper

The architect matched the paint with the wallpaper

Arg0: The architect

REL: matched

Arg1: the paint

Arg1-WITH: with the wallpaper

The paint matches the wallpaper

Arg1: The paint

REL: matches

Arg1: the wallpaper

On the other hand, symmetric verbs like ‘meet’, ‘kiss’, ‘marry’ and others which are not stative, assign Arg0 and Arg1 labels to their arguments (consistent with the assumption that Arg0 label corresponds to a Proto-Agent role, see Dowty 1991 for discussion of symmetric verbs):

Symmetric verbs: ‘meet’

Arg0: *meeter*

Arg1: *person / entity / object being met*

Along the way he meets a solicitous Christian chauffeur who offers the hero God's phone number.

ArgM-ADV: Along the way

Arg0: he

REL: meets

Arg1: a solicitous Christian chauffeur who offers the hero God's phone number

John and Mary met at a cocktail party.

Arg0: John and Mary

REL: met

ArgM-Loc: at a cocktail party

As the verb ‘meet’ illustrates, the transitive and intransitive uses of symmetric verbs are not distinguished as different framesets: *John met Mary* and *John and Mary met* are analyzed as having the same set of roles. The Arg0 argument is certainly different for the two sentences, and can either refer to *John* or both *John and Mary*, dependent on the syntactic structure.

4.5. Psychological Verbs

The class of psychological verbs is also well-discussed in the linguistic literature, in relation to alternations of the following type:

Political concerns worried New England.
John's mother worries about her baby.

In the first sentence, the causer argument is the subject of the verb 'worry', whereas the 'worrier' is the object. In the second sentence, the 'worrier' is the subject, whereas the object is a topic of worry.

Our approach to the analysis of these verbs follows Dowty 1991 and Pesetsky 1995 among many others, who argued that the meaning of the verbs in the two constructions is different: whereas the first sentence is causative and the subject is the cause of 'worry', the second sentence is stative and the object is the topic of emotion rather than the causer (see Pesetsky 1995, for example, for evidence supporting this claim). Along the lines of this analysis, we assume that the two interpretations of 'worry' have different framesets, which are distinguished by different semantic arguments:

Frameset worry.01 "worry with direct object":

Arg0:cause of worrying
Arg1:worrier
Arg2:instrument

Political concerns(Arg0) worried New England(Arg1)

Frameset worry.02 "prepositional usage":

Arg0:worrier
Arg1:topic of worry

John's mother(Arg0) worries constantly about her baby(Arg1).

One of the advantages of adopting this analysis is that it makes annotation consistent with non-alternating psych verbs, such as frighten/fear. According to the present guidelines, verbs like 'frighten' mark the causer argument as Arg0. Verbs like 'fear', on the other hand, also tag their subjects as Arg0, since they are experiencer arguments of stative verbs. In this respect, worry.01 patterns with 'frighten'-type verbs, whereas worry.02 has the same roles as 'fear'-type verbs.

5. Args and ArgMs.

And, finally, the last question which we need to address is how to distinguish numbered arguments, or ArgNs, from modifiers, or ArgMs. The problem of distinguishing arguments from modifiers is not an easy one. There are different syntactic, semantic, and other types of criteria which could be used, however, they do not usually agree on where

to draw the line. Our approach is a practical one: a semantic role is being marked as an argument, if it frequently occurs in a corpus and is specific to a particular class of verbs.

In most cases, this assumption over generalizes the use of arguments, as in the case of financial verbs, shown below, where the start point, extent, and end point are being marked as Args, rather than ArgMs:

Frameset increase.01 "go up incrementally":

Arg0: causer of increase
Arg1: thing increasing
Arg2: amount increased by
Arg3: start point
Arg4: end point

Net income increased to \$274 million from \$130 million.

Arg1: Net income
REL: increased
Arg4-to: to \$274 million
Arg3-from: from \$130 million

This decision has been made based on the fact that most examples in the corpus have Arg2, Arg3, and Arg4 arguments.

On the other hand, in the case of verbs of motion, the 'direction' is being marked as ArgM-DIR, rather than Arg2, given that directed verbs of motion do not frequently occur in the corpus.

Frameset run.02 "walk quickly":

Arg0: runner
Arg1: course, race, distance

A good number decide it's not worth it and run for home.
Arg0: A good number
REL: run
ArgM-DIR: for home