Knowledge Representation: The structure of the world and its reflection in the language

Wojciech Jaworski

Institute of Informatics
University of Warsaw

Motto

At any moment in history we have some image of the nature of reality which always turns out to be partially false and leads to consequent more adequate images, and so on; But for now, at this moment, each option may prove true, the world remains describable only by the competing theories.

Jacek Dukaj

Bibliography and resources

- The formalism of the conceptual graphs and the presented ontology originate in the book

  John F. Sowa
  „Knowledge Representation
  Logical, Philosophical, and Computational Foundations”

- On the website

  `http://www.jfsowa.com/krbook`

  you will find some material for the book, among other:
  - a description of the ontology
  - a definition of the conceptual graphs with examples
  - a description of the thematic roles and the other relations used in the graphs

- Conceptual graphs editor CharGer

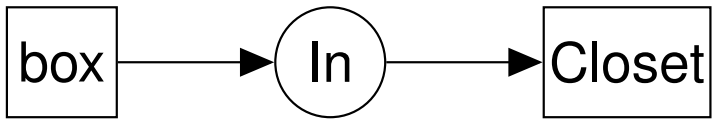
  `http://charger.sourceforge.net`

Physical entities and Space-time

- Everything is an **Entity**.
- The entities immersed in space-time are called **Physical**.
- **Object** is a material entity immersed in space-time which is also **Continuant** and **Independent**.
- Objects enter into spatial relations such as **In**, **On**, **Under**, . . .

I box is in a closet.

$$(\exists x : \text{Box})(\exists y : \text{Closet})\ \text{in}(x, y)$$



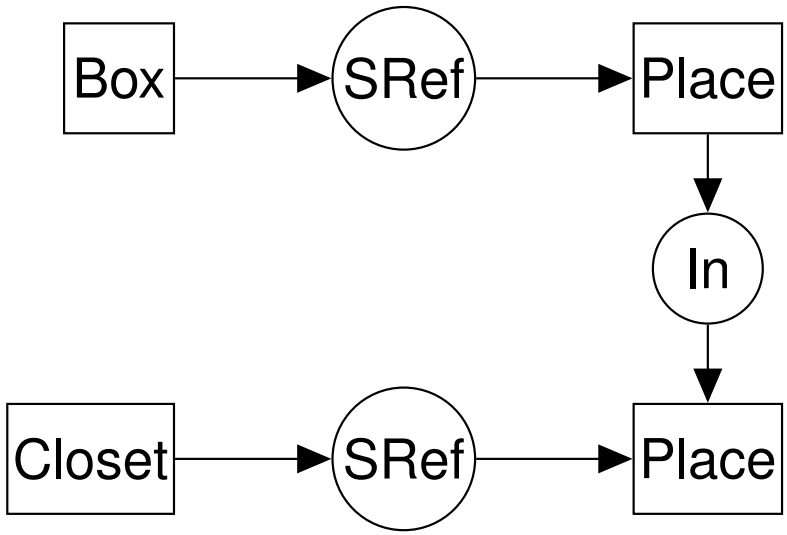
Physical entities and Space-time

- Alternatively we may define a spatial reference which connects an object with a part of space-time, i.e.**Place**.
- In that case it's places that are in spatial relations, not objects.

A box is in a closet.

$$(\exists x : \text{Box})(\exists y : \text{Closet})(\exists a_1, a_2, b : \text{Place})$$

$$\text{sref}(x, a_1) \wedge a_1 = a_2 \wedge \text{in}(a_2, b) \wedge \text{sref}(y, b)$$



Junctures

- Let us consider the entities:
  - *a knot on a rope*,
  - *a joint between bones*,
  - *a connection of car parts*.
- They are Continuants.
- They are not independent (a knot cannot exist without the rope) but **Relative**.
- We will call them **Junctures**.

Phenomenal types

- A phenomenal type describes en entity by its inherent structure.

a rubber ball

- *ball* — a name of a geometric concept — may refer to both a geometric shape and to a material thing.
- *rubber* describes the realization of the form as an object made of rubber.
- In order to determine whether an entity belongs to a phenomenal type we need to know the entity and a classifying procedure.
- Phenomenal types reflect the phenomenal world view, seeing the world *as it appears*, considering *what is immediately given*.
- The type Object and its subtypes characterize individuals with respect to their phenomenal type.
- Phenomenal type corresponds to a class in object modeling.

Role

- The type **Role** is a role played by an entity (an actor) in a relation with another entity
- A role is a specification of a typical behavior, skills, social relations or competence of the entity 'playing' the role.
- For example, *mother, employee, passer-by, student*
- Assigning roles to entities is subjective, requires an observer and depends on their intention (it defines the difference between a role and a connection).
- For example,*a car* can be taken as *a vehicle* or *a determinant of a social status*.
- Sometimes the role is given by the context: *a carp* in a restaurant menu plays the role of *food*, not of *living being*.
- Role classification does not depend on the structure of the object.
- The same role can be assumed by radically opposite phenomena, eg. both *a car* and *a scooter* are vehicles.
- Roles correspond to interfaces in object modeling.

Role

- Sometimes a form can be derived from the role, e.g. a student is a human but there is no strict implication.
- Even if all the entities that can play the given role for the time being belong to a certain phenomenal type, it can change in the future.
- Example role types:
  - **Function** — a role whose actor is not a person.
  - **Social Role**, eg. a father.
  - **Organization Role**, a role played by an organization member, eg. head, student, officer, applicant.

Prehension

- Roles and junctures are cases of **prehension**,
- The actor of the role (prehending entity) is connected with the role reference (prehended entity) by the connecting intention (subjective form).
- An example prehension which is not a role is describing a person with regard to a property: *fatty*, *ginger*, etc.
- The connecting intention does not have to belong to the connected entity: The red-headed person does not decide on how people perceive them.
- The intention determines among others how does the actor play their role.
- For a given prehension complement we can create a role name:
  - A man who *writes* is a *writer*.

Has-test

- determine the prehending entity and the prehended entity of a prehension.
  - For two given words X and Y say the sentence “X has Y”.
  - If the sentence feels natural, X prehends Y.
- Eg. *A car has an engine. A car has a color.*
  but not: *An engine has a car. A color has a car.*
- If the referent is abstract we can say “knows about”, eg. *A physicist knows about physics.*
- For interdependent entities the relation is symmetric:

  A mother has a child. A child has a mother.

  An employer has an employee. An employee has an employer.
- Possession is the basic relation between nouns:

  I have virtues, clothes, knowledge, house, lecture, wife etc.
- We will take prehension as identical to passing the has-test and we will represent it with the relation **Has**.
- Has is a basic conceptual relation, all other in the graphs are its subtypes.

Representation of roles (prehension)

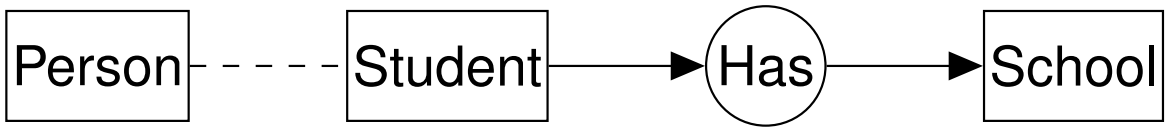
- In natural language we focus on the actor and leave behind the reference of the role.

$$\textit{student}$$

- In logic we represent them as dyadic relations.

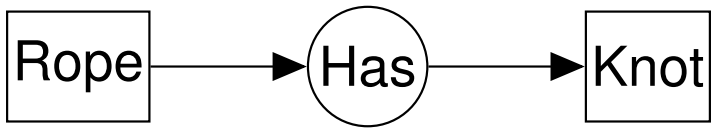
$$\text{student}(x, y)$$

- We represent them in conceptual graphs as concepts (boxes).
- The referent of the role is at the same time an object with its own phenomenal type.

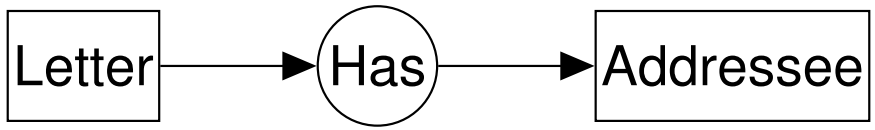


Prehension examples

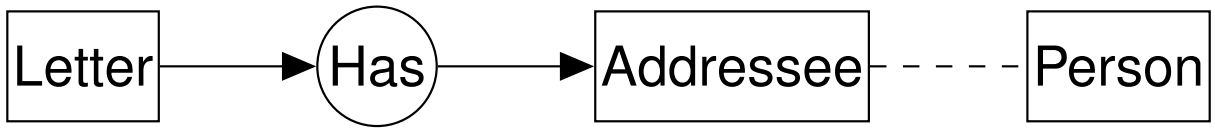
A rope has a knot.



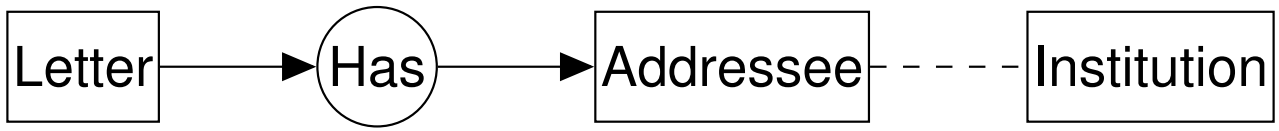
A letter has an addressee.



A letter has an addressee (*a person*).

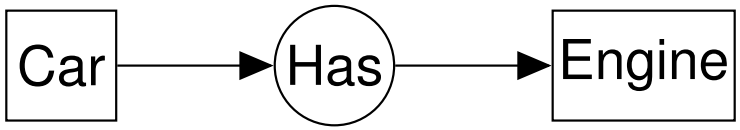


A letter has an addressee (*an institution*).

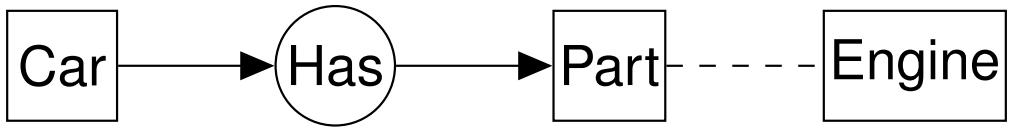


The Part relation

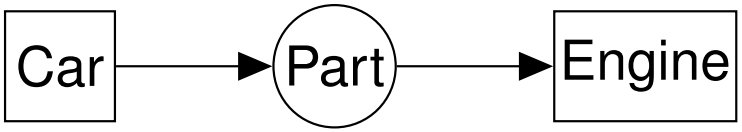
A car has an engine.



Engine is a part of a car.
A car has an engine as a part.



- Being a **Part** is a subtype of a role.
- We define a conceptual relation **Part** for it.



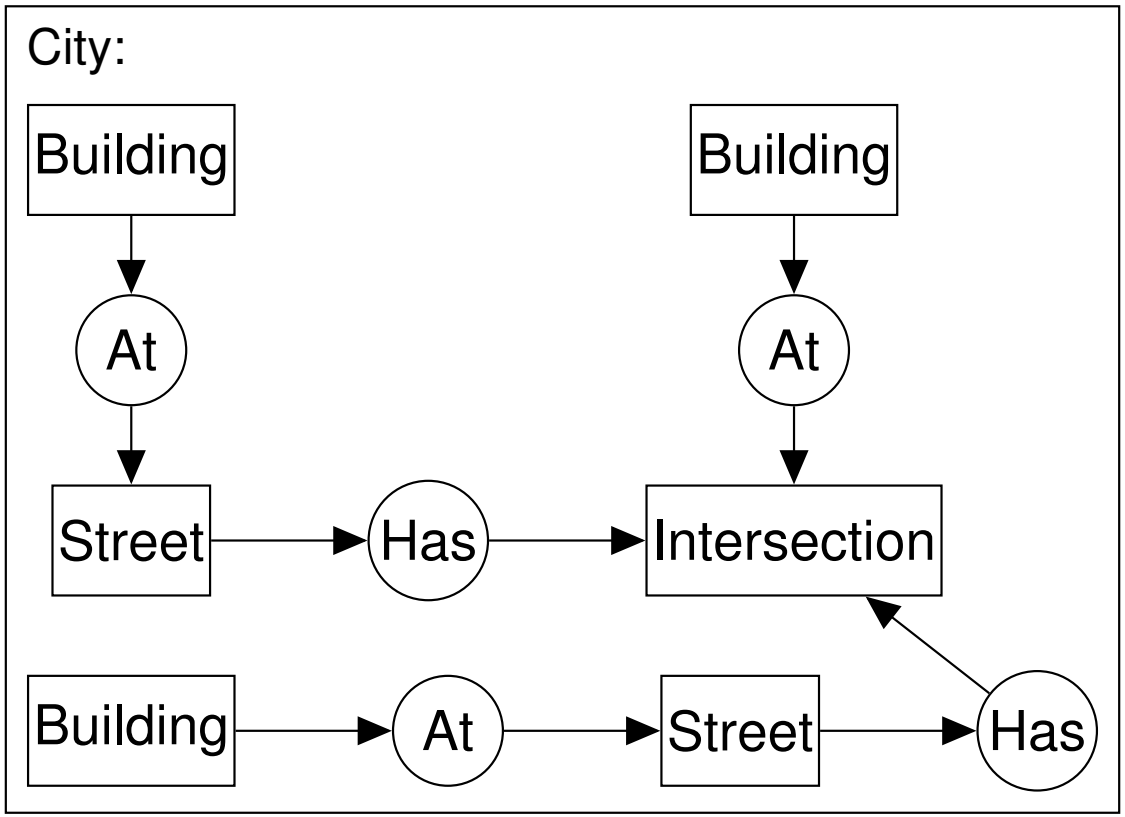
- The Part relation is such a subtype of Has that its second argument is in the role of being a part of the first argument.

Structures

- **Structures** are physical entities continuous in time (Continuants).
- A structure is a **Mediating** entity.
- A structure mediates multiple objects whose junctures constitute the structure.
- The objects are organized to perform a function.
- Each structure has a reason to exist.
- Examples:
  - an arch consists of conjoined stone blocks
  - a nest is a combination of twigs, leaves etc.
  - a city consists of houses, streets

Contexts

- We represent structures as contexts.
- The referent of a structural concept is a graph that shows its structure.



Contexts

A context is given in logic by a metapredicate dscr:

$$(\exists m : \text{City})\text{dscr}\Big(m, (\exists b_1, b_2, b_3 : \text{Building})$$

$$(\exists u_1, u_2 : \text{Street})(\exists s : \text{Intersection})$$

$$\text{at}(b_1, u_1) \wedge \text{at}(b_2, u_2) \wedge \text{at}(b_3, s) \wedge \text{has}(u_1, s) \wedge \text{has}(u_2, s)\Big)$$

Time structure

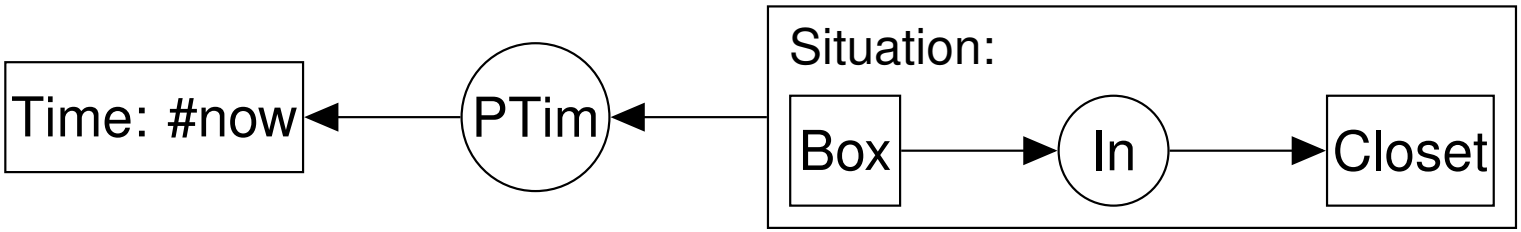
- Time is described by use of mereology, ie. it consists of intervals (time periods) bound by temporal relations.
- Eg. finishes(t_1, t_2) says that t_1 finishes t_2 which means that t_2 begins before t_1 and both end at the same time point.
- Moments - the points in time - are taken to be abstract entities: limits of decreasing interval sequences
- In conceptual graphs we will use:
  - the type **Time** consisting of intervals;
  - the relacji **PTim** binding **Occurrents** with intervals;
  - the relation **Succ** binding succeeding intervals and occurents
  - the indicator **#now** that gives the time point dependent on the context in which it was used.
- Dates are time constants, individuals of the type time.

Physical entities changing in time

- **Processes** are physical entities in changing time and independent.
- Eg. fire, preparing a dinner, ...
- The distinction between physical objects and processes is motivated by the time scale and the granularity (eg. a glacier observed for 5 minutes or for 5 years)
- **Participation** is a relation between a physical object constant in time called **Participant**, eg. a barking dog.
- A **Situation** is a state of affairs, a configuration of entities that takes place in time.

A present situation

- The sentence *The box is in the closet* describes a situation in time



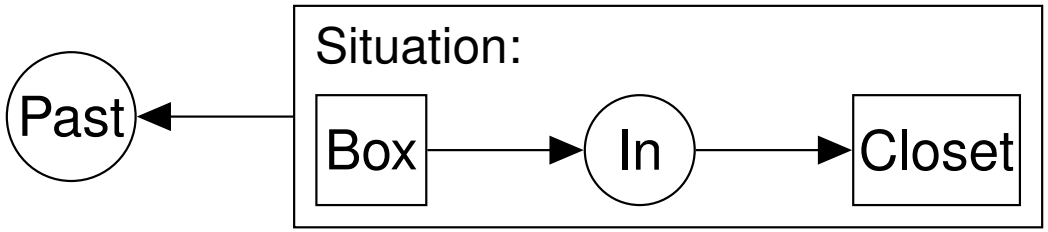
$$\text{Time}(06.11.2012) \wedge (\exists s : \text{Situation})\text{pTim}(s, 06.11.2012) \wedge$$

$$\text{dscr}\big(s, (\exists x : \text{Box})(\exists y : \text{Closet})\text{in}(x, y)\big)$$

- The graph inside the situation box is a graphical representation of the situation, not the very situation.

A past situation

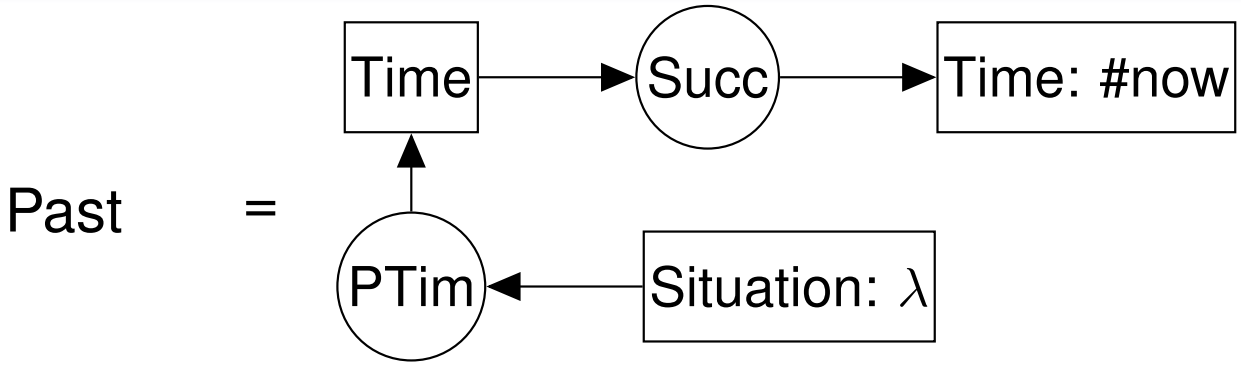
Zdanie *A Box was in a closet* describes a different situation.



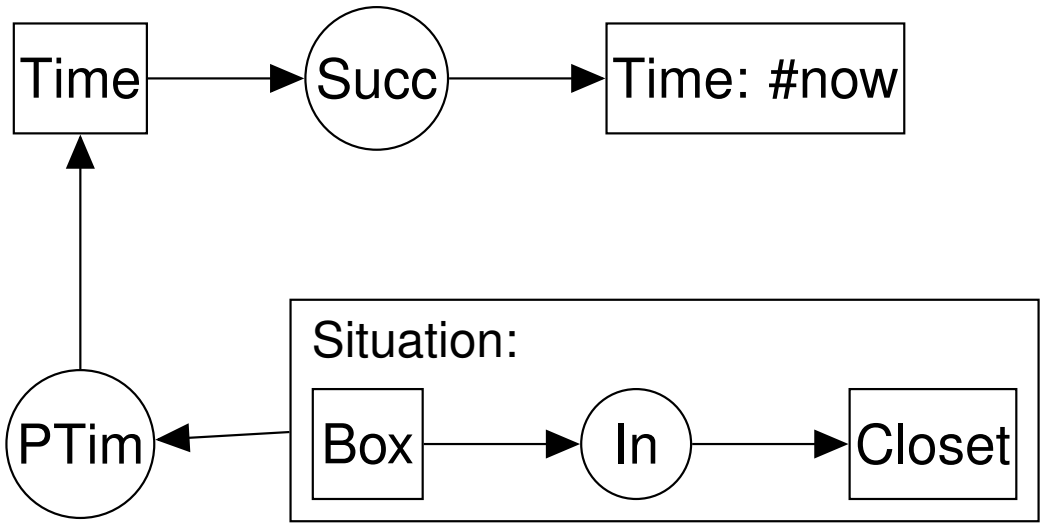
$$(\exists s : \text{Situation})\text{past}(s) \wedge$$

$$\text{dscr}\big(s, (\exists x : \text{Box})(\exists y : \text{Closet})\text{in}(x, y)\big)$$

Definition of the past



The situation description after definition expansion:

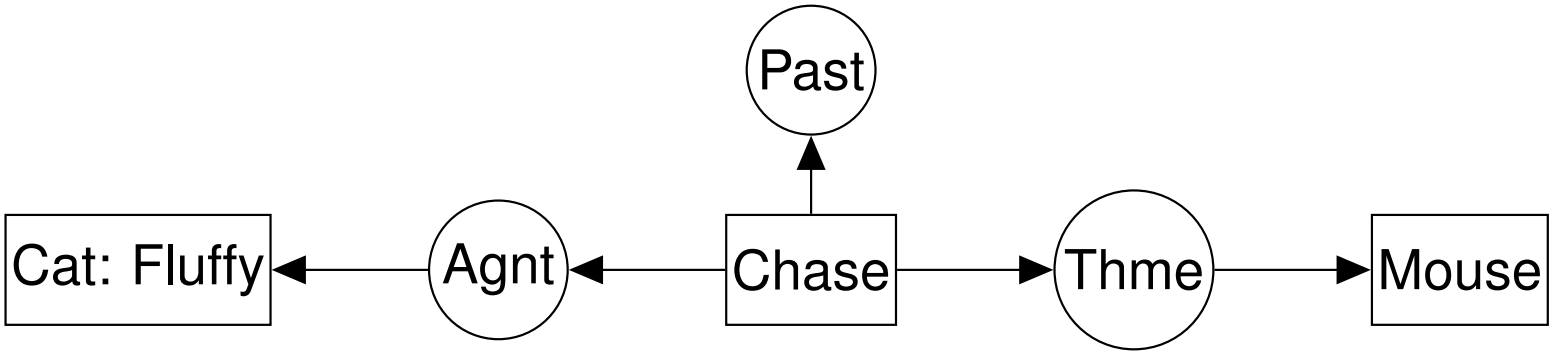


Varieties of processes

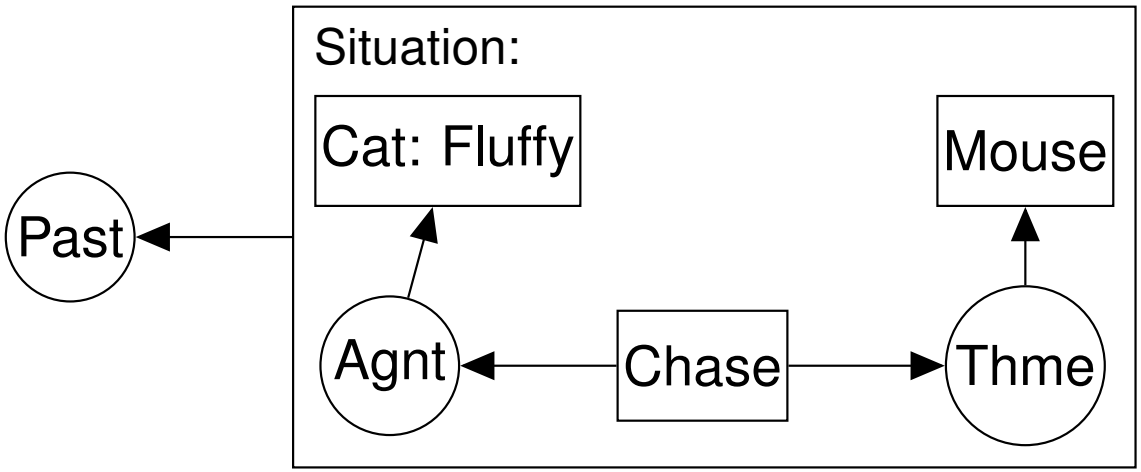
- The change brought about by a process can be continuous in time (eg. physical processes) or discreet (eg. program executions).
- Discreet processes may be divided into **Events** when a change occurs and **States** otherwise.
- On the other hand an **Action** is a change caused by some intentional action of an agent.
- **Agent** is an entity capable of intentional action.
- An agent and the action are bound together by the relation **Agnt**.
- Example agents are people, animals, organizations, computer programs.
- Pronouns *I* and *you* indicates the agents dependently on the context.
- The analyses of cause and intentional action will follow in the lecture.
- Caution: the words *process*, *event* and *action* have no fixed meaning in the literature.

Location of an action in time

- The sentence *Fluffy chased a mouse* may be translated in two different ways



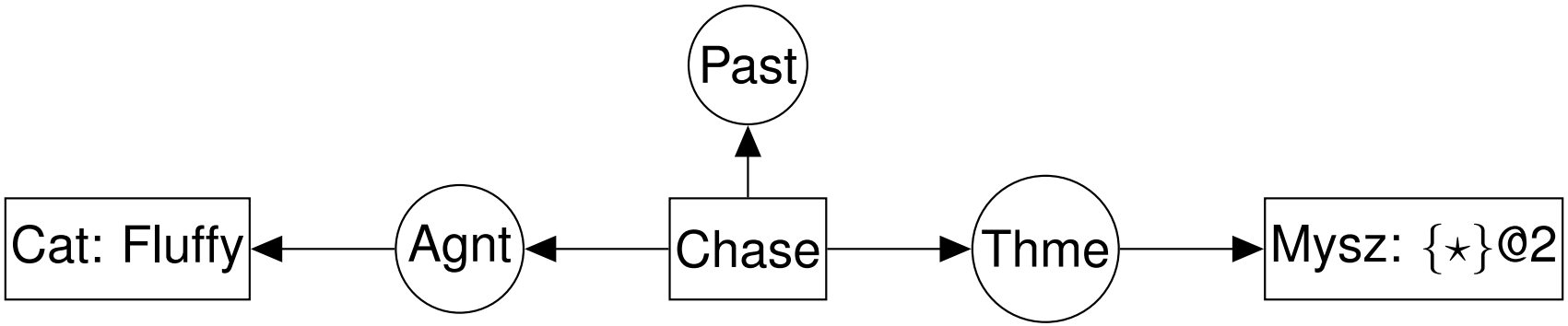
- In the interpetation above the temporal location of Fluffy and the mouse is not directly given.



- The second interpretation can be inferred from the first.

Sentences true in the past only

- The sentence *Fluffy chased two mice* may be translated to



- But there was no situation when Fluffy chased two mice at once.
- $\{*\}$ indicates a set of anonymous mice and @2 gives its cardinality.
- If we add **Dist** in the graph before $\{*\}$@2 we will state that the relation between mice and the rest of the sentence is meant to be distributive, so there were two chasing actions.
- But then if we add **Col** instead we will state that the plural number is collective, so Fluffy indeed chased a set of two mice.

The scope of the verb arguments

- The modifiers can refer to the situation as a whole or to a particular entity participating in it.

  He studied at a university in Warsaw

- This is a syntactic ambiguity, but semantically *in Warsaw* refers to the whole situation.

  He studied at a university in Warsaw taking online classes,

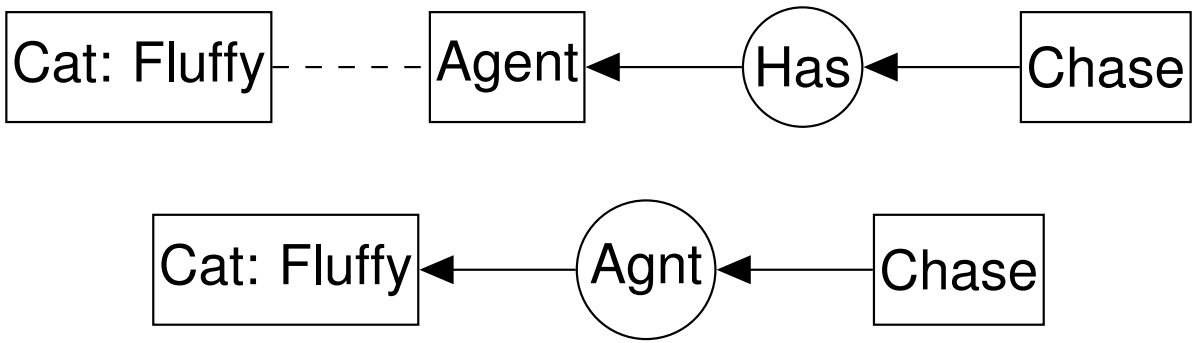
- *taking online classes* suggests that it's only the university that was located in Warsaw but the studying action is undefined.

  He studied in Warsaw taking online classes at a university.

- The position of the modifier *in Warsaw* indicates that it describes the action (the location of the agent) and the university can be anywhere.

Thematic roles

- Thematic roles are relations that classify entities - actions and processes.
- They are approximate correlates of syntactic relations and syntactic categories.
- There is no fixed set of thematic roles and different sets are employed for different objectives.
- Thematic roles are represented in conceptual graphs as relational concepts (circles); they are subrelations of the relation Has.



Thematic roles by John F. Sowa

- http://www.jfsowa.com/ontology/thematic.htm
- Sowa's proposal is a decision tree.
- We ask a sequence of questions for a given argument and determine its role depending on the answers.
- The construction of the decision tree ensures that the set of roles will not need to be extended and it facilitates their ascription.

Decision tree:

Does the participant determine the direction of the process, either from the beginning as the initiator or from the end as the goal?

- Yes $\rightarrow$ Must be present at the beginning of the process, but need not participate throughout the process?
  Must be present at the end of the process but need not participate throughout the process?
  - Yes,No $\rightarrow$ An active animate entity that voluntarily initiates an action?
    - Yes $\rightarrow$ Agent (**Agnt**)
    - No $\rightarrow$ An active determinant source?
      Yes $\rightarrow$ Effector (**Efct**) No $\rightarrow$ Origin (**Orgn**)
  - No,Yes $\rightarrow$ Is participant animate?
    - Yes $\rightarrow$ Is participant an active animate goal of an experience?
      Yes $\rightarrow$ Experiencer (**Expr**) No $\rightarrow$ Recipient (**Rcpt**)
      (Beneficiary (**Benf**))
    - No $\rightarrow$ Result (**Rslt**)
  - Yes,Yes $\rightarrow$ We choose one or both of the above-mentioned options.
  - No,No $\rightarrow$ We assume such arguments do not occur.

Decision tree:

Does the participant determine the direction of the process, either from the beginning as the initiator or from the end as the goal?

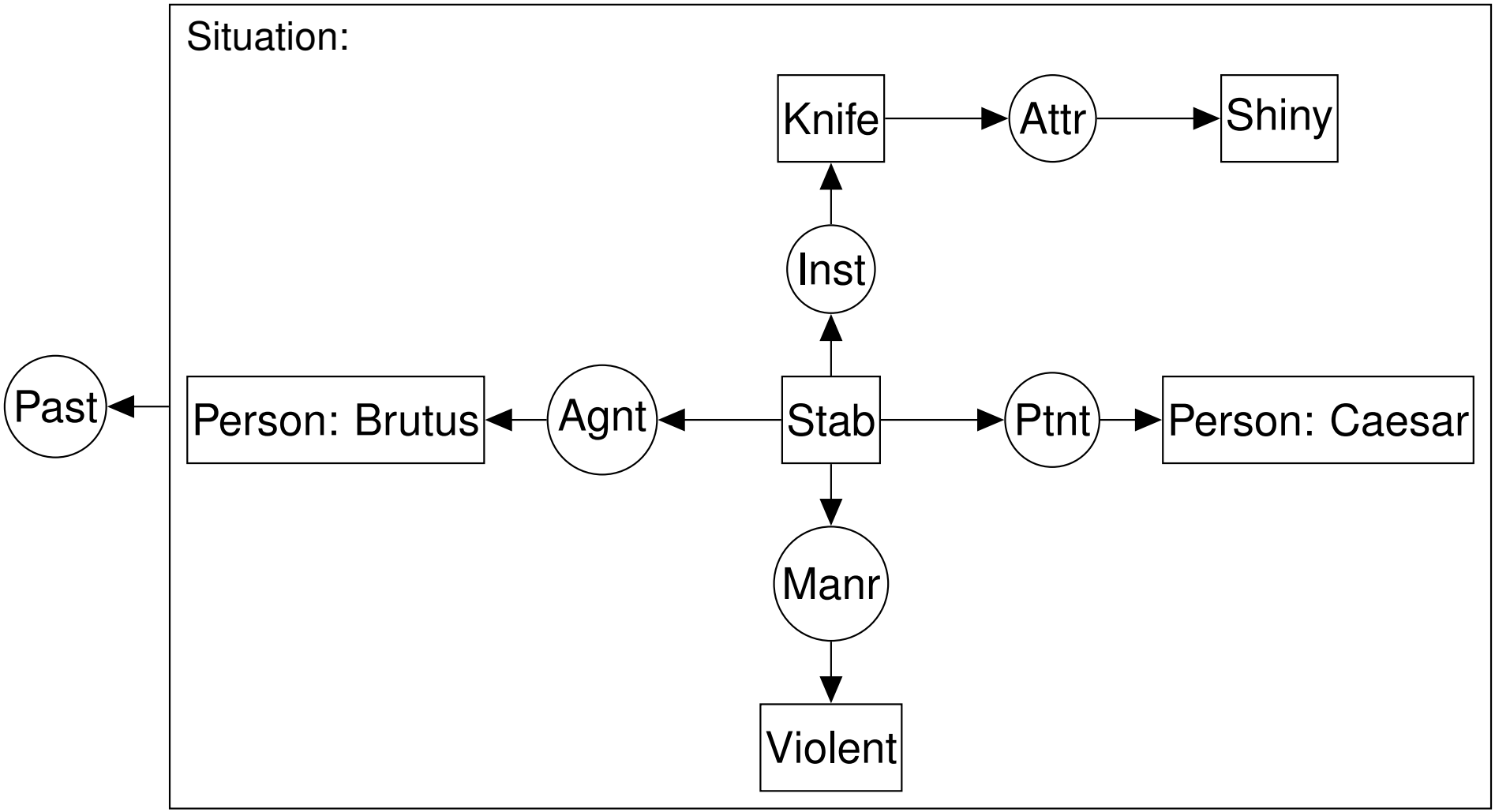
- No $\rightarrow$ Must be present at the beginning of the process, but need not participate throughout the process?
  Must be present at the end of the process but need not participate throughout the process?
  - Yes,No $\rightarrow$ Is participant changed by the event?
    - Yes $\rightarrow$ Matter (**Matr**)
    - No $\rightarrow$ Instrument (**Inst**)
  - No,Yes $\rightarrow$ Does the participant undergo some structural change as a result of the event?
    - Yes $\rightarrow$ Patient (**Ptnt**)
    - No $\rightarrow$ Theme (**Thme**)
  - Yes,Yes $\rightarrow$ We choose one or both of the above-mentioned options.
  - No,No $\rightarrow$ We assume such arguments do not occur.

Next to the ones before we also have temporal and spacial roles:

- Origin (**Orgn**), Path (**Path**), Destination (**Dest**), Location (**Loc**)
- Start(**Strt**), Duration(**Dur**), Completion(**Cmpl**), PointInTime(**PTim**).

Brutus stabbed Caesar violently with a shiny knife.

Brutus stabbed Caesar violently with a shiny knife.



The role Manner (**Manr**) indicates how the action is conducted.

VerbNet

- A list of 5257 verb meanings with thematic roles of their arguments and with their semantic preferences.
- The verbs are grouped into 274 classes after the thematic roles of their arguments.
- 1396 sentences tagged with thematic roles.
- http://verbs.colorado.edu/ mpalmer/projects/verbnet.html
- Verbnet roles can be divided in two categories:
  - General roles: Agent, Experiencer, Recipient, Beneficiary, Product, Material, Instrument, Patient, Theme, Location, Source, Destination, Time;
  - VerbNet specific roles: Actor, Asset, Attribute, Cause, Extent, Predicate, Stimulus, Topic.
- The official examples also feature roles not mentioned on the list, like Co-Agent.

FrameNet

- A set of 1164 semantic frames.
- A semantic frame corresponds to a concept.
- One frame includes
  - a list of arguments (and modifiers) of its concept given by their thematic roles (summing up to 10011)
  - a list of lexemes (or multi-word expressions) expressing the concept (summing up to 12714)
  - relations between the frames, like inheritance or being a sub-frame.
- Each lexeme is associated with a list of semantic types of its arguments and each type is given together with different possible syntactic realizations.
- Lexemes that realize concepts are verbs(4880), nouns (5177), adjectives (2270) and other (387), eg.
  - The frame “Death” is for the lexemes “die”, “death”, “perish”, “pass away”,
  - The frame “Documents” is for different types of documents: “license”, “certificate”, “confirmation”, “permit”, “subpoena”, “visa”.

Semantic Frames

- Schematic representations of situations.
- A frame consists of a set of verbs describing a situation and of names of the roles of their arguments.
- Eg. the situation *Becoming_a_member* can be expressed among others by the verbs *join*, *enroll* and *enlist*.
- One verb can play multiple roles that correspond to its different meanings.
- Eg. *join* occurs in *Attaching*, *Cause_to_amalgamate* and *Becoming_a_member*.
- The sets of roles depend on the project.
  - FrameNet takes roles specific for a particular situation, eg. *Group*, *New_member*.
  - VerbNet uses thematic roles similar to those described previously.
  - PropBank has generic roles *Arg0*, *Arg1*, *Arg2*,. . .
- SemLink integrates all 3.

Actions as roles

- Verbs can express the form (structure) of an action or describe a role.

  Mary hid a ball by putting it in a box.

- Mary performed one action, described by both 'hide' and 'put in'.
- An external observer could have seen that Mary put something in the box but does not know that she 'hides' (wants to hide) without knowing her intentions.

  Mayor spent a lot of time presenting his projects. He spoke for almost three hours.

- Three verbs express the same action:
  - *spoke* gives the form of an action (firstness);
  - *presenting* ignores the form, indicates its effect (secondness);
  - *spent time* refers only to the duration which is an aspect of the effect of the action.
- One action can be described through its form (firstness), result (secondness) and agent's intentions (thirdness).

Classification of adjectives

- **Qualitative (characterizing) adjectives** indicate definite (ie. incidental, non-constant) qualities, eg. shape, size, color, temperature etc.
- Qualities referenced by qualitative adjectives can occur to a lesser or greater extent, which is indicated by degree of comparison of the adjective.

- **Possessive adjectives** tell us to whom a thing belongs, eg. paternal advice, avian beak, feline tail, Polish language.

Classification of adjectives

- **Varietal (classifying, distinctive) adjectives** denote indefinite (ie. constant) features and qualities of an object.
- They indicate a characteristic feature of a kind: the defining one, and not just one of many features that the object has.
- They denote features concerning eg. the material out of which the thing was made (*wooden sculpture*), character of a chemical substance (*sulfuric acid*), type of vehicle (*electric car*), kind of action indicated by the noun (*cashless transaction*), time (*daily exercise*), place (*urban church*).
- These features do not possess degrees of comparison.

A methodological remark

- This classifiation does not constitute a comprehensive and fully accurate description of phenomena.
- There is no unambiguous assignment of words to classes, and the set of classes doesn't cover all possibilities.
- The true merit of this classification lies in its statistical accuracy, decided by the proportion of cases where the classifier gave an answer and the fraction where the answers were correct.

Representing qualities of objects

- Qualities are colors, shapes, sizes, names etc.
- Representations of the phrase *a red ball*:
  - a simple but impractical one

$$(\exists x : \text{Ball})\text{red}(x)$$

  - a representation coherent with semantics of natural language:

$$(\exists x : \text{Ball})(\exists y : \text{Redness})\text{attr}(x, y)$$

    Each word has a corresponding concept. The relation of being an attribute (‘attr‘) links an object to its quality, which is itself an object.
  - a representation in the spirit of relational database:

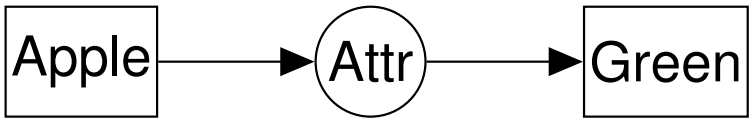
$$(\exists x : \text{Ball})(\text{color}(\text{Red}) \wedge \text{chrc}(x, \text{Red}))$$

    There is a table ‘Ball‘ with a column named ‘color‘. x is an object in that table and ‘Red‘ is the value of attribute ‘color‘ for that object. ‘chrc‘ is a predicate that links an object in table to its quality – a type.
- The type ‘Color‘ is of the type ‘Characteristic‘, which a third-order type.

Relationship of qualitative adjectives to nouns

- We denote the relation of being a quality by **Attr** (Attribute).

a green apple



- We can think of Attr as of an identity relation.
- In other words, we mean here objects of the type indicated by the noun that have the quality indicated by the adjective.
- This construct helps us describe qualitative adjectives.
- This is enough for varietal adjectives when the variety is prescribed by intersection of types of the noun and the adjective, eg. *a wooden sculpture*.
- But this does not indicate the varietal nature of the relationship.
- Varietal adjective can be used with a qualitative meaning, as in *a sulfuric smell*.

Relation of adjectives *sulfur*, *sulfuric* to the noun

- The adjectives *sulfur*, *sulfuric* denote objects that contain sulfur or are otherwise related to it: *sulfur soap, sulfur shampoo, sulfur cream, sulfur pumice, sulfur concrete, sulfur fertilizer, sulfur wick, sulfur industry*.
- When meanings are related to chemistry: (*sulfuric acid, sulfuric amino acid, sulfuric electrolyte*) the nature of sulfur's involvement is precisely indicated by the noun.
- In each of the above examples we can infer the meaning of the phrase from the noun and the adjective. We can write this down using types

  Sulfur(Soap) Sulfuric(Acid) itd.

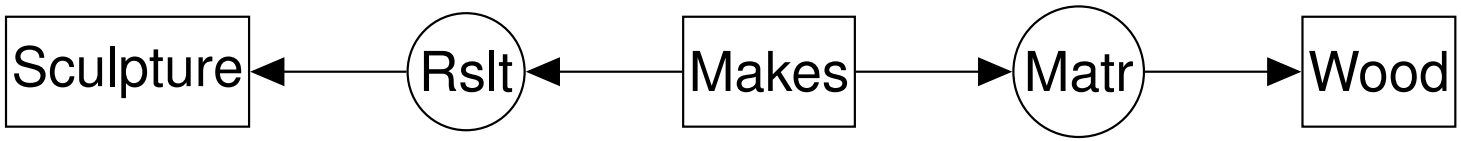
- The 'Sulfuric(Acid)' notation can be expanded to a graph that describes how the algorithm detecting molecules of acid is narrowed, because of the adjective, to an algorithm detecting only sulfuric acid molecules.

Names of species

- In the case of *sulfur polypore* (a kind of fungus), it is not related to sulfur (except for the color), but is a subtype of a *polypore*.
- But the meaning cannot be inferred from meanings of the individual words.
- *Guinea pig* is an extreme example: it is not a *pig* and isn't from *Guinea*.
- In mathematics, *outer measure* is a more general concept than *measure*.

Default processes

- When a noun denotes a process, we interpret the adjective as some thematic role in that process.
- In some cases the noun as well as the adjective are in some thematic roles of some unmentioned process.
- For example, *a wooden sculpture* is *a sculpture made out of wood*, which is given by the graph



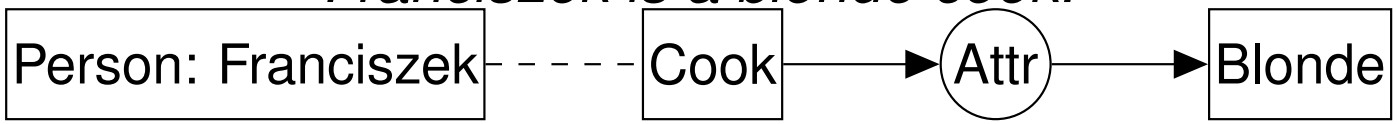
- On the other hand, *an Italian book* can be *a book wrote in Italy*, *a book bought in Italy*, etc.

- **Prtc** (Participant) is a relation encompassing all thematic roles.

Role-modifying adjectives

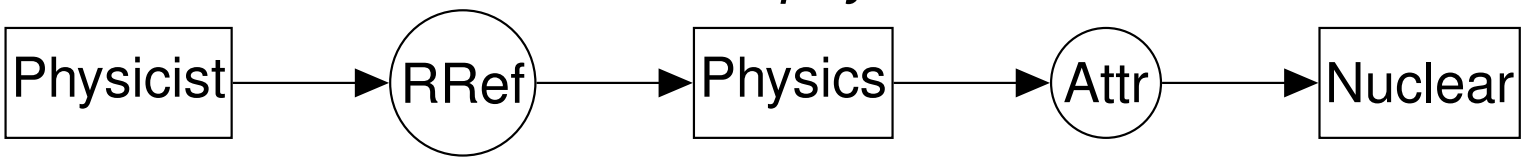
- When noun describes an entity on the basis of its structural type, this entity is the only thing a feature of which we can express.
- When noun descrives a role, an adjective can modify the actor of this role, its reference or intention which describes how the actor performs its role.
- *A blonde cook* is *a blonde person*.

Franciszek is a blonde cook.



- But *a bad cook* need not be *a bad person*.
- Similarly, *a nuclear physicist* is not *a nuclear human*.

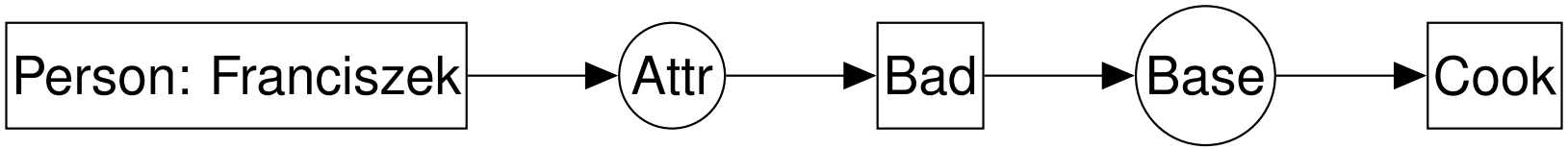
a nuclear physicist



- We are assuming that *a nuclear physicist* differs from *a physicist* only by the role's reference.

Role-modifying adjectives

Franciszek is a bad cook.



- We can think of the role as of a set of skills, abilities, qualities possessed by the actor.
- For example a cook possesses the skill of cooking.
- In this convention the adjective *bad* becomes a modificator of the type *cook* that deprives it of professional skills.
- Similarly *former senator, theoretical computer scientist, university teacher, senior specialist, habilitated doctor*.
- In English a noun can also be used for that purpose, eg. *assistant professor*.

Adjuncts

- Possesive adjunct tends to signify prehension: the fact of owning something or being in a relation to something.
- It can be interpreted with the relation Has (*John's father*), but it can also be a modifier (*butcher's knife*).
- Substantive adjunct tends to introduce proper name of the object, the type of which is given by the main noun (*The Hoover Building*).
- Apposition (*driver salesman, king Kazimierz*) denotes coreference: the existence of two names for the same object.
- Adverbial adjuncts can be interpreted as arguments of a presumed verb: *a university in Warsaw* is *a university located in Warsaw*.
- This lets us replace the Has relation in the graph with the appropriate thematic role, or add a presupposed process.
- The task of discovering appropriate relations makes natural language understanding hard for humans and computers.