

Endomorphic metalanguage and abstract planning for real-time intent recognition

Antoine Gréa

Abstract

Human-machine interaction is among the most complex problems in the field of artificial intelligence. Indeed, software that cooperates with dependent people must have contradictory qualities such as speed and expressiveness, or even precision and generality. This requires a radical paradigm shift from standard approaches, which, while effective, are often too rigid. The objective is then to design tools capable of generating intermediate solutions that sometimes get very useful for certain problems. These solutions make it possible to expand the possibilities of adaptation in a fluid and continuous way. With this idea, the role of explainability in the resolution process is a decisive criterion for obtaining flexible systems. Thus, the search at all costs for a complete and optimal response has overshadowed the usefulness of these models.

Mathematics offers the best tools for formalization. However, it appears that the foundation of mathematics is defined by an implicit circularity. This defect reveals a lack of rigor in describing the basic concepts used in mathematics. It is therefore sufficient to recognize the need for a formalism based on a precursor language. In the absence of alternatives, natural language is a precursor to a new formalism. This is inspired by lambda calculus as well as type theory. From basic concepts, defined by explicit circularity, it is then possible to reconstruct classical mathematical concepts as well as other tools and structures very useful what is following.

By using this formalism, it becomes possible to axiomatize an endomorphic metalanguage. This one manipulates a dynamic grammar capable of adjusting its semantics to exploitation. The recognition of basic structures allows this language to avoid using keywords. This, combined with a new model of knowledge representation, supports the construction of a very expressive description logic.

With this language and this formalism, it becomes possible to create frameworks in fields that were previously heterogeneous. For example, in automatic planning, the classic state-based model makes it impossible to unify the representation of planning domains. The interest of such tools on planning and its multiple paradigms is illustrated. This results in a general planning framework that allows all types of domains to be expressed.

Concrete algorithms are then created that show the principle of intermediate solutions. Two new approaches to real-time planning are presented and evaluated. The first is based on a usefulness heuristic of planning operators to repair existing plans. The second uses hierarchical task networks to generate valid plans that are abstract and intermediate solutions. These plans allow for a much shorter execution time for any use that does not require a detailed plan.

We then illustrate the use of these plans on intent recognition by reverse planning. Indeed, in this field, the fact that no plan libraries are required makes it much easier to



~~design recognition models~~. By exploiting abstract plans, it becomes possible to create a system theoretically more efficient than those using complete planning.

Résumé

L'interaction personne-machine fait partie des problèmes les plus complexes dans le domaine de l'intelligence artificielle. En effet, les logiciels qui coopèrent avec des personnes dépendantes doivent avoir des qualités contradictoires telles que la rapidité et l'expressivité, voire la précision et la généralité. Cela exige un changement radical de paradigme par rapport aux approches standards, qui s'avèrent souvent efficaces, mais rigides. On a alors, pour objectif, la conception d'outils capables de générer des solutions intermédiaires qui ont parfois une grande utilité pour certaines problématiques. Ces solutions permettent d'élargir les possibilités d'adaptation de manière fluide et continue. Dans l'idée, le rôle de l'explicabilité dans le processus de résolution constitue un critère déterminant pour obtenir des systèmes flexibles. Aussi, la recherche à tout prix d'une réponse complète et optimale n'a que trop occulté l'utilité de ces modèles.

Les mathématiques offrent les meilleurs outils de formalisation. Cependant, il s'avère que la fondation des mathématiques est définie par une circularité implicite. Ce défaut agit comme un révélateur d'un manque de rigueur pour décrire des notions de base utilisées en mathématique. Il suffit dès lors de reconnaître la nécessité d'un formalisme établie sur un langage précurseur. À défaut d'alternatives, le langage naturel fait office de précurseur pour un nouveau formalisme. Celui-ci est inspiré par le lambda-calcul ainsi que par la théorie des types. À partir de concepts de base, définis par circularité explicite, il est alors réalisable de reconstruire les notions mathématiques classiques ainsi que d'autres outils et structures très utiles pour la suite.

En se servant de ce formalisme, il devient possible d'axiomatiser un métalangage endomorphique. Celui-ci manipule une grammaire dynamique capable d'ajuster sa sémantique à l'usage. La reconnaissance des structures de base permet à ce langage de ne pas utiliser de mot-clés. Ceci, combiné à un nouveau modèle de représentation des connaissances, supporte la construction d'une logique de description très expressive.

Avec ce langage et ce formalisme, il devient envisageable de créer des cadriels dans des champs jusqu'alors hétéroclites. Par exemple, en planification automatique, le modèle classique à état rend l'unification de la représentation des domaines de planification impossible. On illustre l'intérêt de tels outils sur la planification et ses multiples paradigmes. Il en résulte un cadriel général de la planification permettant d'exprimer tout type de domaines en vigueur.

On crée alors des algorithmes concrets qui montrent le principe des solutions intermédiaires. Deux nouvelles approches à la planification en temps réel sont présentées et évaluées. La première se base sur une euristique d'utilité des opérateurs de planification afin de réparer des plans existants. La seconde utilise la planification hiérarchique pour générer des plans valides qui sont des solutions abstraites et intermédiaires. Ces

plans rendent possible un temps d'exécution bien inférieur pour tout usage ne nécessitant pas le plan détaillé.

On illustre alors l'usage de ces plans sur la reconnaissance d'intention par planification inversée. En effet, dans ce domaine, le fait de ne pas nécessiter de bibliothèques de plans rend la création de modèles de reconnaissance bien plus aisée. En exploitant les plans abstraits, il devient possible de réaliser un système théoriquement plus performant que ceux utilisant de la planification complète.

Table of Content

Acknowledgements	14
Remerciements	15
Preface	16
Text format	16
Citations	16
Quotes	16
1 Introduction	17
1.1 Motivations	17
1.2 Problem	17
1.3 Contributions	18
1.4 Plan	19
2 Foundation and Tools	20
2.1 Properties	20
2.1.1 Abstraction	20
2.1.2 Formalization	21
2.1.3 Circularity	22
2.2 Functional theory of mathematics	23
2.2.1 Formalism definition	23
2.2.2 Literals and Variables	25
2.2.3 Functional algebra	25
2.2.4 Properties	28
2.3 First Order Logic	29
2.4 Set Theory	30
2.4.1 Base Definitions	30
2.4.2 Set Operations	30
2.4.3 The ZFC Theory	32
2.5 Graphs	33
2.5.1 Adjacency, Incidence and Connectivity	33
2.5.2 Digraphs	34
2.5.3 Path, cycles and transitivity	34
2.5.4 Trees	35
2.5.5 Quotient	35
2.5.6 Hypergraphs	36
2.6 Sheaf	37

3 Knowledge Representation	40
3.1 Grammar and Parsing	41
3.1.1 BNF	41
3.1.2 Dynamic Grammar	41
3.2 Description Logics	42
3.3 Ontologies and their Languages	43
3.4 Self	44
3.4.1 Knowledge Structure	45
3.4.1.1 Consequences	45
3.4.1.2 Native Properties	46
3.4.2 Syntax	48
3.4.3 Dynamic Grammar	50
3.4.3.1 Containers	50
3.4.3.2 Parameters	52
3.4.3.3 Operators	53
3.4.4 Contextual Interpretation	53
3.4.4.1 Naming and Scope	54
3.4.4.2 Instanciation identification	54
3.4.5 Structure as a Definition	55
3.4.5.1 Quantifiers	55
3.4.5.2 Inferring Native Properties	58
3.4.6 Extended Inference Mechanisms	59
3.4.6.1 Type Inference	59
3.4.6.2 Instanciation	59
3.5 Example	60
3.5.1 Modality of Statements	60
3.5.2 Higher order knowledge	60
4 General Planning Framework	64
4.1 Illustration	64
4.2 Formalism	66
4.2.1 Planning domain	67
4.2.1.1 Fluents	67
4.2.1.2 States	67
4.2.1.3 Verification and binding constraints	68
4.2.2 Application	68
4.2.2.1 Actions	69
4.2.2.2 Domain	70
4.2.3 Planning problem	70
4.2.3.1 Solution to Planning Problems	70
4.2.3.2 Planning Problem	71
4.3 Planning search	71
4.3.1 Search space	72
4.3.2 Solution constraints	73
4.3.3 Temporal constraints	73
4.4 General planner	74
4.5 Classical Formalism	75
4.5.1 State-transition planning	75

4.5.2	Plan space planning	75
4.5.3	Case based planning	76
4.5.4	Probabilistic planning	76
4.5.5	Hierarchical planning	77
4.6	Existing Languages and Frameworks	77
4.6.1	Classical	77
4.6.2	Temporality oriented	80
4.6.2.1	PDDL+	80
4.6.3	Probabilistic	81
4.6.3.1	PPDDL	81
4.6.3.2	RDDL	82
4.6.4	Multi-agent	83
4.6.4.1	MAPL	83
4.6.4.2	MA-PDDL	83
4.6.5	Hierarchical	84
4.6.5.1	UMCP	84
4.6.5.2	SHOP2	85
4.6.5.3	HDDL	85
4.6.5.4	HPDDL	86
4.6.6	Ontological	87
4.6.6.1	WebPDDL	87
4.6.6.2	OPT	88
4.7	Color and general planning representation	89
4.7.1	Framework	89
4.7.2	Example domain	89
4.7.3	Differences with PDDL	90
5	Online and Flexible Planning Algorithms	92
5.1	Existing Algorithms	92
5.1.1	Plan Space Planning	93
5.1.1.1	Definitions	93
5.1.1.2	Classical POCL Algorithm	96
5.1.1.3	Existing PSP Planners	97
5.1.2	Plan Repair & Reuse	97
5.1.3	HTN	98
5.2	LOLLIPOP	99
5.2.1	Operator Graph	99
5.2.2	Negative Refinements	102
5.2.3	Usefulness Heuristic	104
5.2.4	Algorithm	106
5.2.5	Theoretical and Empirical Results	109
5.2.5.1	Proof of Soundness	109
5.2.5.2	Proof of Completeness	110
5.2.5.3	Experimental Results	111
5.3	HEART	113
5.3.1	Domain Compilation	113
5.3.2	Abstraction in POCL	113
5.3.3	Planning in cycle	115

5.3.4	Properties of Abstract Planning	116
5.3.5	Computational Profile	118
6	Toward Intent Recognition	120
6.1	Domain problems	120
6.1.1	Observations and inferences	120
6.1.2	Cognitive inconsistencies	121
6.1.3	Sequentiality	121
6.2	Existing approaches	121
6.2.1	Constraint	122
6.2.1.1	Deductive Approach	122
6.2.1.2	Algebraic Approach	122
6.2.1.3	Grammatical Approach	123
6.2.1.4	Linear programming approach	123
6.2.1.5	Markovian Logic Approach	124
6.2.2	Networks	124
6.2.2.1	And/Or trees approach	124
6.2.2.2	HTN Approach	125
6.2.2.3	Hidden Markovian Approach	126
6.2.2.4	Bayesian Approach	126
6.2.2.5	Markovian network approach	126
6.2.2.6	Bayesian Theory of Mind	127
6.3	Inverted planning	127
6.3.1	Probabilities and approximations	128
6.4	Intent recognition Using Abstract Plans	130
6.4.1	Linearization	131
7	Conclusion and Perspectives	132
7.1	Fondation	132
7.1.1	Issues with current fondations	132
7.1.1.1	Set theory	132
7.1.1.2	Type theory	132
7.1.1.3	Category theory	133
7.1.1.4	Proof theory	133
7.1.2	Reverse mathematics	133
7.1.3	Possible paradox in proposed theory	133
7.2	Knowledge representation	133
7.2.1	Literal definition using Peano's axioms	134
7.2.2	Advanced Inference	135
7.2.3	Queries	135
7.3	General Automated Planning	135
7.4	Planning Improvements	135
7.4.1	Heuristics using Semantics	135
7.4.2	Macro-Action learning	135
8	References	136
Apendix		146
8.1	Chapter 1	146

List of Figures

2.1	Alfred Korzybski	20
2.2	Illustration of basic functional operators and their properties.	26
2.3	Dependency graph of notions in the functional theory	28
2.4	Example of an upgraded Venn diagram to illustrate operations on sets. .	31
2.5	Example of the recursive application of the transitive cover to a graph. .	34
2.6	Example of and/or tree.	35
2.7	Example of graph quotient.	36
2.8	Example of hypergraph with total freedom on the edges specification. .	37
2.9	Example of a seed, a section and a stalk.	38
2.10	Example of sheaves.	39
3.1	Noam Chomsky 2017	40
3.2	Projection of a statement from the SELF to RDF space.	46
3.3	Venn diagram of subsets of \mathbb{U} along with their relations. Dotted lines mean that the sets are defined a subset of the wider set.	48
3.4	Aristotle's square of opposition	56
3.5	Hierarchy of types in SELF	58
3.6	Example of modal logic propositions: Alice gossips about what Beatrice said about Claire	61
4.1	The block world domain setup.	65
4.2	An example of a solution to a planning problem with a goal that requires three blocks stacked in alphabetical order.	66
4.3	Example of a state encoded as an and/or tree.	68
4.4	Example of verification and/or tree with binding constraints	68
4.5	Example of application of a state with the updated binding constraints .	69
4.6	Venn diagram extended from the one from SELF to add all planning knowledge representation.	70
4.7	Structure of a partially ordered plan.	72
4.8	Venn diagram extended with general planning formalism.	74
4.9	Dependencies and grouping of PDDL requirements.	79
5.1	Planning phases for online planning	92
5.2	Example of partial plan having flaws	94
5.3	Example of resolvers that fixes the previously illustrated flaws	95
5.4	Example of the side effects of the application of a resolver	96
5.5	Venn diagram of the expressivity class of HTN paradigms.	100
5.6	Diagram of the operator graph of example domain. Full arrows repre- sent the domain operator graph and dotted arrows the dependencies added to inject the initial and goal steps.	101

5.7	Schema representing flaws with their signs, resolvers and side effects relative to each other	103
5.8	Domain used to compute the results. First line is the initial and goal steps along with the useful actions. Second line contains a threatening action t , two co-dependent actions n and l , a useless action u , a toxic action v , a dead-end action w and an inconsistent action x	111
5.9	In full lines the initial safe operator graph. In thin, sparse and irregularly dotted lines respectively a subgoal, alternative and threat caused causal link.	112
5.10	Illustration of how the cyclical approach is applied on the example domain. Atomic actions that are copied from a cycle to the next are omitted.	116
5.11	Evolution of the quality with computation time.	118
5.12	Impact of domain shape on the computation time by levels. The scale of the vertical axis is logarithmic. Equations are the definition of the trend curves.	119
6.1	The lattice formed by observations (top), matching plans, possible hypothesis and problem (bottom)	123
6.2	The valued grammar used for intent recognition	124
6.3	The HTN and decision tree used for intent recognition	125
6.4	Representation of the encoding of constraints using extra fluents.	128
6.5	Illustration of how plan costs are affecting the resulting probabilities.	130
6.6	Example of linearization of partial order plan.	131

Acknowledgements

Ehlo, this is the author of this thesis. I would like to thank a few people without whom none of this would have been possible.



Actual picture of the time

It was a beautiful morning of June that I first walked the Doua wooded driveways to meet my supervisor team: Samir Aknine and Laetitia Matignon. They have given me their trust in the success of this thesis project. Without their advice and support none of this would have been possible. I would like to thank Samir in particular whom, as a mentor and friend, has listened to me and supported me even in the most difficult moments with compassion and understanding.

I would also like to thank Christine Gertosio, who offered me unfailing support in my teaching department and granted me the opportunity to pursue an ATER offer at  POLYTECH Lyon. It was a real pleasure to teach classes of passionate and enthusiastic students.

I would also like to offer my gratitude to the researchers who helped me in this adventure, in no particular order: Fabrice Jumel, Jacques Saraydaryan, Shirin Sohrabi, Damien Pelier. I would also like to thank Linas Vepstas, who greatly helped me in the mathematical part of this thesis.

Without funding, this work wouldn't have been possible. I would like to thank the  arc 2  (Academic Research Community) research allocation courtesy of the Auvergne-Rhône-Alpes  region in France. This organism is acting toward an improvement of the quality of life and of aging.

This work was done with the  Lyon 1 university in the  LIRIS (Laboratory of computeR science in Image and Information Systems). I was part of vibrant and wonderful  sma  team (Multi-Agent System). Part of the work was funded by teaching at the university  Lumière Lyon 2 and the  POLYTECH engineering school.

Finally, I would like to thank my family for supporting me despite my periods of avolition. I would like to thank in particular my mother Charlotte and my sister Sarah for having done their best to keep me company and pretend to listen to me talk about the technicalities of my thesis in order to make me progress.

Thank you again.



Sarah Gréa helping me do the rubber duck method.

Remerciements

Je voudrais remercier quelques personnes sans qui rien de cela n'aurait été possible.

C'était une belle matinée de juin que j'ai parcouru pour la première fois les allées boisées de la Doua afin de rencontrer mon équipe encadrante: Samir Aknine et Laetitia Matignon. Ces derniers m'ont accordé leur confiance dans la réussite de ce projet de thèse. Sans leur conseil et soutien rien de cela n'aurait été possible. Je remercie particulièrement Samir, qui en tant qu'encadrant, et ami, a su m'écouter et me soutenir même dans les moments les plus pénibles avec compassion et compréhension.



Photo de l'époque

Je souhaite également remercier Christine Gertosio, qui m'a offert une aide sans faille dans mon service d'enseignement et qui m'a donné l'opportunité d'une offre d'ATER à **POLYTECH Lyon**. C'était un véritable plaisir de donner des cours à des promotions d'élèves passionnés et enthousiastes.

Je souhaite également offrir ma gratitude aux chercheurs qui m'ont aidé dans cette aventure, sans ordre particulier: Fabrice Jumel, Jacques Saraydaryan, Shirin Sohrabi, Damien Pelier. J'aimerais également remercier Linas Vepstas, qui m'a grandement aidé dans la partie mathématique de cette thèse.

Sans financement, ce travail n'aurait pas été possible. Je tiens à remercier l'allocation de recherche du **arc 2** (Academic Research Community) gracieuseté de la région de Auvergne-Rhône-Alpes en France. Cet organisme agit pour l'amélioration de la qualité de vie et du vieillissement.

Ce travail a été réalisé avec l'université **Lyon 1** dans le **LIRIS** (Laboratoire de calcul scientifique en Image et Systèmes d'Information). Je faisais partie d'une équipe dynamique et conviviale **sma** (Multi-Agent System). Une partie du travail a été financée par l'enseignement à l'université **Lumière Lyon 2** et à l'école d'ingénieurs polytech.

Enfin, je tiens à remercier ma famille de m'avoir supporté malgré mes périodes d'avolition. Je voudrais remercier tout particulièrement ma mère Charlotte et ma sœur Sarah pour avoir fait de leurs mieux pour me tenir compagnie et faire semblant de m'écouter déblatérer des technicités de ma thèse afin de me faire avancer.

Encore merci.



Sarah Gréa m'a aidant en utilisant la méthode du canard en plastique.

Preface

In this section we present a quick guide to the presentation of the information in this document. This will give its global structure and each of the type of formatting and its meaning.

Text format

The text can be emphasized **more** or *less* to make a key word more noticeable.

Citations

In text citations will be in this format: Author *et al.* ([year](#)) to make the author part of the text and ([Author *et al.* year](#)) when simply referencing the work.

Quotes

Sometimes, important quotes needs emphasis. They are presented as:

Gréa ([2019](#)) “*Don’t quote me on that!*”

1 Introduction

In antiquity, philosophy, mathematics and logic were considered as a single discipline. Since Aristotle we have realized that the world is not just black and white but full of nuance and colour. The inspiration for this thesis comes from one of the most influential philosopher and scientist of his time: Alfred Korzibsky. He founded a discipline he called *general semantics* to deal with problems of knowledge representation in humans. Korzibsky then found that complete knowledge of reality being inaccessible, we had to abstract. This abstraction is then only similar to reality in its structure. In these pioneering works, we find notions similar to that of modern descriptive languages.

It is from this inspiration that this document is built. We then start off the beaten track and away from computer science by a brief excursion into the world of mathematical and logical formalism. This makes it possible to formalize a language that allows to describe itself by structure and that evolves with its use. The rest of the work illustrates the possible applications through specific fields such as automatic planning and intention recognition.

1.1 Motivations

The social skills of modern robots are rather poor. Often, it is that lack that inhibits human-robot communication and cooperation. Humans being a social species, they require the use of implicit social cues in order to interact comfortably with an interlocutor.

In order to enhance assistance to dependent people, we need to account for any deficiency they might have. In our case we chose to address cognitive deficiencies. The main issue is that the patient is unable or unwilling to express their needs. That is a problem even with human caregivers as the information about their intents needs to be inferred from their past actions.

These aspect of social communication often eludes the understanding of Artificial Intelligence (AI) systems. This is the reason why intent recogniton is such a complicated problem. The primary goal of this thesis is to address this issue and create the formal foundations of a system able to help dependent people.

1.2 Problem

First, what exactly is intent recognition ? The problem is simple to express: finding out what other agents want to do before they do. It is important to distinguish between

several notions. *Plans* are the sequence of actions that the agent is doing to achieve a *goal*. This goal is a concrete explanation of the wanted result. However, the *intent* is more of a set of abstract goals, some of which may be vague or impossible (e.g. drink something, survive forever, etc.).

Some approaches use trivial machine learning methods, along with a hand made plan library to match observations to their most likely plan using statistics. The issue with these common approaches is that they require an extensive amount of training data and need to be trained on each agent. This makes the practicality of such system quite limited. To address this issue, some works proposed hybrid approaches using logical constraints on the probabilistic method. These constraints are made to guide the resolution toward a more coherent solution. However, all probabilistic methods require an existing plan library that can be quite expensive to create.

A work from Ramirez and Geffner (2009), added an interesting method to solve this issue. Indeed, they noticed an interesting parallel between that problem with the field of automated planning. This analogy was made by using the Theory of mind Baker *et al.* (2011), which states that any agent will infer the intent of other agents using a projection of their own expectation on the observed behavior of the other agent.

This made the use of planning techniques possible to infer intent without the need for extensive and well-crafted plan libraries. Now only the domain of the possible actions, their effects and prerequisites are needed to infer the logical intent of an agent.

The main issue of planning for that particular use is computation time and search space size. This prevents most planners to make any decision before the intent is already realized and therefore being useless for assistance. This time constraint lead to the search of a real-time planner algorithm that is also expressive and flexible.

1.3 Contributions

In order to achieve such a planner, the first step was to formalize what is exactly needed to express a domain. Hierarchical and partially ordered plans gave the most expressivity and flexibility but at the cost of time and performance. This is why, a new formalism of knowledge representation was needed in order to increase the speed of the search space exploration while restricting it using semantic inference rules.

While searching for a knowledge representation model, some prototypes were done using standard ontology tools but all proved to be too slow and inexpressive for that application. This made the design of a lighter but more flexible knowledge representation model, a requirement of the project.

Then the planning formalism has to be adapted to our general knowledge representation tool. Since automated planning has a very diverse ecosystem of approaches and paradigms, its standard, the Planning Domain Description Language (PDDL) needs use of various extensions. However, no general formalism has been given for PDDL and some approaches often lack proper extensions (hierarchical planning, plan representation, etc). This is why a new formalism is proposed and compared to the one used as standard of the planning community.

Then finally, a couple of planners were designed to attempt answering the speed and flexibility requirements of human intent recognition. The first one is a prototype that aims to evaluate the advantages of repairing plans to use several heuristics. The second is a more complete prototype derived from the first (without plan repairs), which also implements a Breadth-First Search (BFS) approach to hierarchical decomposition of composite actions. This allows the algorithm to provide intermediaries plan that, while incomplete, are an abstraction of the result plans. This allows for anytime probability computation using existing techniques of invert planning.

1.4 Plan

In this document we will describe a few contributions from the new mathematical formalism to intent recognition. Each chapter builds on the previous one.

First we will present a new mathematical fondation, highlighting the flaws of the current ones. This fondation is used to create a formalism capable of describing all the classical mathematics.

In the third chapter, a new knowledge description system is presented as well as the associated grammar and inference framework.

The fourth chapter is an application of that knowledge description system to automated planning. Existing formalism and languages are compared to this new general one. This allows us to design a general planning framework that can express any existing type of domain.

Using this framework, two online planning algorithms are presented in chapter five: one that uses repairs on existing plans and one that uses hierarchical domains to create intermediary abstract plans.

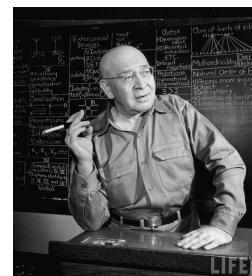
The final chapter is about intent recognition and its link to planning. Existing works are presented as well as a technique called *inverted planning*.

2 Foundation and Tools

Mathematics and logic are at the heart of all formal sciences, including computer science. Some of the most important problems in mathematics are the consistency and formalization issues. Research on these issues starts at the end of the 19th century, with Cantor inventing set theory. Then after a crisis in the beginning of the 20th century with Russel's paradox and Gödel's incompleteness theorem, revised versions of the set theory become one of the foundations of mathematics. The most accepted version is the Zermelo-Fraenkel axiomatic set theory with the axiom of Choice (ZFC). This effort leads to a formalization of mathematics itself, at least to a certain degree.

Cantor (1874)

Fraenkel et al.
(1973, vol. 67);
Ciesielski (1997)



Any knowledge must be expressed using a medium like language. Natural languages are quite expressive and allow for complex abstract ideas to be communicated between individuals. However, in science we encounter the first issues with such a language. It is culturally biased and improperly conveys formal notions and proof constructs. This is one of the main conclusions of the works of Korzybski (1958) on general semantics. The original goal of Korzybski was to pinpoint the errors that led humans to fight each other in World War I (WWI). He affirmed that the language is unadapted to convey information reliably about objective facts or scientific notions. There is a discrepancy between the natural language and the underlying structure of the reality. This issue is exacerbated in mathematics as the ambiguity in the definition of a term can be the cause for a contradiction and make the entire theory inconsistent.

Figure 2.1: Alfred
Russell Korzybski (1917)

"Mathematics may be defined as the subject in which we never know what we are talking about, nor whether what we are saying is true."

In this first chapter we analyze the issue of expression in mathematics and propose our solution. Then we use our foundation to build some important mathematical notions used later to formalize data structures and algorithms.

2.1 Properties

From this point we can apply some of these ideas to mathematics to analyze what properties a foundation of mathematics must hold.

2.1.1 Abstraction

abstraction (n.d.): *The process of formulating generalized ideas or concepts by extracting common qualities from specific examples*

Collins English
Dictionary (2014)

The idea behind abstraction is to simplify the representation of (potentially infinitely) complex instances. This mechanism is at the base of any knowledge representation system. Indeed, it is unnecessarily expensive to try to represent all properties of an object. An efficient way to reduce that knowledge representation is to prune away all irrelevant properties while also only keeping the one that will be used in the context. *This means that abstraction is a lossy process.* Information is lost when abstracting from an object.

Since this is done using a language as a medium, this language is a *host language*. Abstraction will refer to an instance using a *term* (also called symbol) of the host language. Since abstraction is a generalization process, if the host language is expressive enough, abstraction can be applied to already abstracted knowledge. The number of layers abstraction is needed for a term is called its *abstraction level*. Very general notions have a higher abstraction level and we usually represent reality using the null abstraction level. In practice abstraction uses terms of the host language to bind to a referenced instance in a lower abstraction level. This forms a structure that is strongly hierarchical with higher abstraction level terms on top.

Example 1. We can describe an individual organism with a name that is associated to this specific individual. If we name a dog "Rex" we abstract a lot of information about a complex, dynamic living being. We can also abstract from a set of qualities of the specimen to build higher abstraction. For example, its species would be *Canis lupus familiaris* from the *Canidae* family. Sometimes several terms can be used at the same abstraction level like the commonly used denomination "dog" in this case.



Terms are only a part of that structure. It is possible to combine several terms into a *formula* (also called proposition or statements).

2.1.2 Formalization

formal (adj.): *Relating to or involving outward form or structure, often in contrast to content or meaning.*

American
Heritage
Dictionary
(2011a)

A formalization is the act to make formal. The word "formal" comes from Latin *fōrmālis*, from *fōrma*, meaning form, shape or structure. This is the same base as for the word "formula". In mathematics and *formal sciences* the act of formalization is to reduce knowledge down to formulas. Like stated previously, a formula combines several terms. But a formula must follow rules at different levels:

- *Lexical* by using terms belonging in the host language.
- *Syntactic* as it must follow the grammar of the host language.
- *Semantic* as it must be internally consistent and meaningful.

The information conveyed from a formula can be reduced to one element: its semantic structure. Like its etymology suggest a formula is simply a structured statement about terms. This structure holds its meaning. Along with using abstraction, it becomes possible to abstract a formula and to therefore make a formula about other formulae should the host language allowing it.

Example 2. The formula using English “dog is man’s best friend” combines terms to hold a structure between words. It is lexically correct since it uses English words and grammatically correct since it can be grammatically decomposed as (n. v. n. p. adj. n.). In that the (n.) stands for nouns (v.) for verbs (adj.) for adjectives and (p.) for possessives. Since the verb “is” is the third person singular present indicative of “be”, and the adjective is the superlative of “good”, this form is correct in the English language. From there the semantic aspect is correct too but that is too subjective and extensive to formalize here. We can also build a formula about a formula like “this is a common phrase” using the referential pronoun “this” to refer to the previous formula.

However, there is a strong limitation of a formalization. Indeed, a complete formalization cannot occur about the host language. It is possible to express formulae about the host language but *it is impossible to completely describe the host language using itself*. This comes from two principal reasons. As abstraction is a loosy process one cannot completely describe a language that can evolve and be infinitely complex. And even when taking a simple language, there is a problem with describing the different levels of rules to process the language using itself. This supposes of the knowledge of the language *a priori*. This is contradictory and therefore impossible to achieve.

When abstracting a term, it may be useful to add information about the term to define it properly. That is why most formal system requires a *definition* of each term using a formula. This definition is the main piece of semantic information on a term and is used when needing to evaluate a term in a different abstraction level. However, this is causing yet another problem.

2.1.3 Circularity

circularity (n.d.): *Defining one word in terms of another that is itself defined in terms of the first word.*

American
Heritage
Dictionary
(2011b)

Defining a term requires using a formula in the host language to express the abstracted properties of the generalization. The problem is that most terms will have *circular* definitions. Like stated this means that every knowledge system will have some circularity in their definitions. This means that it is impossible to have a complete definition of a formal system without needing another formal system to describe its base.

Using definitions from the American Heritage Dictionary (2011b), we can find that the term “word” is defined using the word “meaning” that is in turn defined using the term “word”. Such circularity can happen to use an arbitrarily long chain of definition that will form a cycle in the dependencies.

This problem is very important as it is overlooked in most foundations of mathematics. This also makes Russel’s paradox unavoidable in such systems since there is always an infinite descent in the definition relation. Also, since a formalization cannot fully be self defined, another host language is generally used, sometimes without being acknowledged. This causes cycles in the dependencies of languages and theories in mathematics.

The only practical way to make some of this circularity disappear is to base a foundation of mathematics using natural language as host language for defining the most basic terms. This allows to acknowledge the problem in an instinctive way while being aware of it while building the theory.

2.2 Functional theory of mathematics

We aim to find the smallest possible set of axioms allowing to describe a foundation of mathematics. The following theory is a proposition for a possible foundation that takes into account the previously described constraints. It is inspired by category theory, and typed lambda calculus Awodey (2010).

In this part, as an introduction to the fundamentals of maths and logic, we propose another view of that foundation based on functions. The unique advantage of it lays in its explicit structure that have emergent reflexive properties. It also holds a coherent algebra that has a strong expressive power. This approach is loosely based on category theory. However, it differs in its priorities and formulation.

Our theory is axiomatic, meaning it is based on fundamental logical proposition called axioms. Those forms the base of the logical system and therefore are accepted without any need for proof. In a nutshell, axioms are true prior hypotheses.

The following axiom is the mandatory undefinable notion created from natural language. It is the explicit base of the formalism.

Axiom (Function). Let's *everything* be functions that associates a unique function (image) to each function (argument).

That axiom states that the formalism is based on *functions*. Any function can be used in any way as long as it has a single argument and returns only one value at a time (image which are also functions).

It is important to know that **everything** in our formalism *is* a function. Even notions such as literals, variables or set from classical mathematics are functions.

Next we need to lay off the base definitions of the formalism.

2.2.1 Formalism definition

This functional algebra at the base of our foundation is inspired by *operator algebra* Takesaki (2013, vol. 125). The problem with the operator algebra is that it supposes vectors and real numbers to work properly.

Since any formalism has circular definition, we will have two ways of defining each notion: using natural language as host language or by using notions that will be defined later. The main definitions we will express are using natural language but we'll add the formula using future notion as side notes. The goal is to explicitly separate fully defined notions from the circular base while giving the dependencies between notions to ease future improvements.

Here we define the basic notions of our functional algebra that dictates the rules of the formalism we are defining.

Definition 1 (Currying). Currying is the operation named after the mathematician Haskell Brooks Curry that allows multiple argument functions in a simpler monoidal formalism. A monome is a function that has only one argument and has only one value, as in our main axiom.

The operation of *currying* is a function $\langle \rangle$ that associates to each function f another function that recursively partially applies f with one argument at a time.

If we take a function h such that when applied to x gives the function g that takes an argument y , *unCurrying* is the function $\langle \rangle$ so that $f(x, y)$ behaves the same way as $(h)(x)(y)$.

Definition 2 (Application). We note the application of f with an argument x as $f(x)$. The application allows to recover the image y of x which is the value that f associates with x .

Along with Currying, function application can be used *partially* to make constant some arguments.

Definition 3 (Partial Application). We call *partial application* the application using an insufficient number of arguments to any function f . This results in a function that has fewer arguments with the first ones being locked by the partial application. It is interesting to note that currying is simply a recursion of partial applications.

From now on we will note $f(x, y, z, \dots)$ any function that has multiple arguments but will suppose that they are implicitly Curried. If a function only takes two arguments, we can also use the infix notation e.g. xfy for its application.

Example 3. Applying this to basic arithmetic for illustration, it is possible to create a function that will triple its argument by making a partial application of the multiplication function $\times(3)$ so we can write the operation to triple the number 2 as $\times(3)(2)$ or $\times(2, 3)$ or with the infix notation 2×3 .

Definition 4 (Association). Let's be the *association function* (\rightarrow) such that, for any two functions x and $f(x)$, the expression $x \rightarrow f(x)$ is equal to f .

Definition 5 (Specification). The *function of specification* (noted $:$) is a function that restricts the validity of an expression given a predicate. It can be intuitively be read as "such that".

The specification operator is extensively used in classical mathematics but informally, it is often seen as an extension of natural language and can be quite ambiguous. In the present document any usage of $(:)$ in any mathematical formula will follow the previously discussed definition.

Definition 6 (Null). The *null function* is the function between nothing and nothing. We note it \gg .

Definition 7 (Identity). The *identity function* is the function that associates any function to itself noted $(=)$. This function is therefore transparent as by definition $= (x)$ is the same as x . A more useful form of this function is as a binary predicate (see LATER).

The predicate can work in two ways: *equation* and *affectation*.

$$(\text{ }) = x \rightarrow x$$

$$(\text{ }) = (x, x \rightarrow T) \bowtie (x, y \rightarrow \perp)$$

- As an **equation**, the predicate tests whether two functions are the same.
- The **affectation** allows to assert that two functions have the same identity making them an alias to one another. An affectation is often used to define a new function.

In the rest of the document, classical equality and the identity function will refer to the same notion.

2.2.2 Literals and Variables

As everything is a function in our formalism, we use the null function to define notions of variables and literals.

Definition 8 (Literal). A literal is a function that associates null to its value. This consists of any function l written as $\text{ } \rightarrow \text{ } l$. This means that the function has only itself as an immutable value. We call *constants* functions that have no arguments and have as value either another constant or a literal.

$$l \Rightarrow \rightarrow l$$

Example 4. A good example of that would be the yet to be defined natural numbers. We can define the literal 3 as $3 \Rightarrow \rightarrow 3$. This is a fonction that takes no argument and is always valued to 3.

Definition 9 (Variable). A variable is a function that associates itself to null. This consists of any function x written as $x \rightarrow \text{ } \rightarrow \text{ }$. This means that the function requires an argument and has undefined value. Variables can be seen as a demand of value or expression and mean nothing without being defined properly.

$$x = x \rightarrow \rightarrow$$

An interesting property of this notation is that $\text{ } \rightarrow$ is both a variable and a constant. Indeed, by definition, $\text{ } \rightarrow$ is the function that associates $\text{ } \rightarrow \rightarrow \text{ }$ and fullfil both definitions.

When defining currying, we annotated with the formalism $(f) = f \rightarrow (x \rightarrow (f(x)))$. The obvious issue is the absence of stopping condition in that recursive expression. While the end of the recursion doesn't technically happen, in practice from the way variables and literals are defined, the recursion chain either ends up becoming a variable or a constant because it is undefined when currying a nullary function.

2.2.3 Functional algebra

Inspired by relational algebra and by category theory, we present a highly expressive and compact functional algebra. In the figure 2.2 we illustrate the different operators of this algebra and their properties.

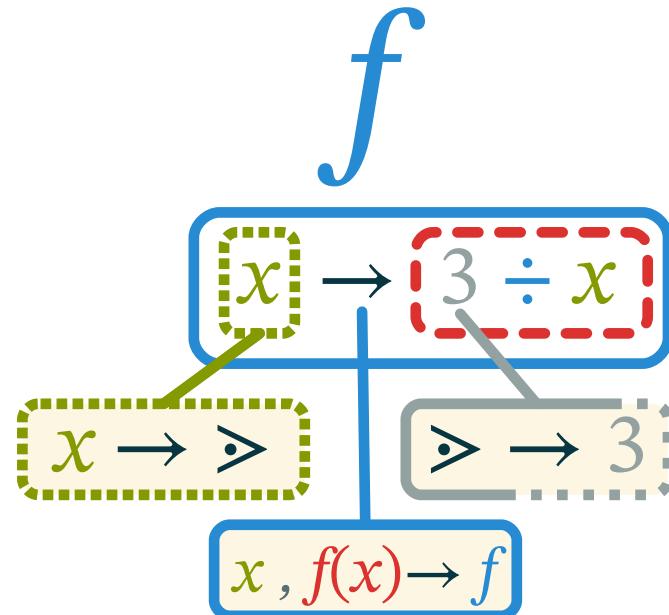


Figure 2.2: Illustration of basic functional operators and their properties.

FIXME: legend & explanation

The first operator of this algebra allows to combine several functions into one. This is very useful to add to the definition of a function by specifying special cases using a reduced notation.

Definition 10 (Combination). The *combination function* \bowtie is the function that associates any two functions to a new function that associates any functions associated by **either** functions.

$$f_1 \bowtie f_2 = x \rightarrow y \\ \text{with } f_1(x) = y \text{ or} \\ f_2(x) = y.$$

In other words, combination allows the union of two function by adding all their definitions into one function.

Example 5. For two functions f_1 and f_2 that are defined respectively by:

- $f_1 = (1 \rightarrow 2)$
- $f_2 = (2 \rightarrow 3)$

the combination $f_3 = f_1 \bowtie f_2$ will behave as follows:

- $f_3(1) = 2$
- $f_3(2) = 3$

Definition 11 (Superposition). The *superposition function* Δ is the function that associates any two functions to a new function that associates any function associated by **both** functions.

$$f_1 \Delta f_2 = x \rightarrow y \\ \text{with} \\ f_1(x) = f_2(x) = y.$$

We can say that the superposition is akin to an intersection where the resulting function is defined when both functions have the same behavior.

Example 6. Reusing the functions of the previous example, we can note that $f_3 \Delta f_1 = f_1$ because $1 \rightarrow 2$ is the only combination that both functions associates.

Definition 12 (Subposition). The *subposition function* is the function ∇ that associates any two functions f_1 and f_2 to a new function $f_1 \nabla f_2$ that associates anything associated by the first function f_1 but **not** by the second function f_2 .

The subposition is akin to a subtraction of function where we remove everything defined by the second function to the definition of the first.

We can also note a few properties of these functions:

- $f \Delta f = f$, a function superposed to itself is the same.
- $f \Delta >=>$, any function superposed by null is null.
- $f_1 \Delta f_2 = f_2 \Delta f_1$, the superposition order doesn't affect the result.
- $f \nabla f =>$, a function subposed by itself is always null.
- $f \nabla >= f$, subposing null to any function doesn't change it.

These functions are intuitively the functional equivalent of the union, intersection and difference from set theory. In our formalism we will define the set operations from these.

The following operators are the classical operations on functions.

Definition 13 (Composition). The *composition function* is the function that associates any two functions f_1 and f_2 to a new function such that: $f_1 \circ f_2 = x \rightarrow f_1(f_2(x))$.

Definition 14 (Inverse). The *inverse function* is the function that associates any function to its inverse such that if $y = f(x)$ then $x = \bullet(f)(y)$.

We can also use an infix version of it with the composition of functions: $f_1 \bullet f_2 = f_1 \circ \bullet(f_2)$.

These properties are akin to multiplication and division in arithmetic.

- $f \circ >=>$ ($>$ is the absorbing element)
- $f \circ == f$ ($=$ is the neutral element)
- $\bullet(>) =>$ and $\bullet(=) ==$
- $f_1 \circ f_2 \neq f_2 \circ f_1$

From now on, we will use numbers and classical arithmetic as we had defined them. However, we consider defining them from a foundation point of view, later using set theory and Peano's axioms.

In classical mathematics, the inverse of a function f is often written as f^{-1} . Therefore we can define the transitivity of the n^{th} degree as the power of a function such that:

- $f^{-1} = \bullet f$
- $f^0 = (=)$
- $f^1 = f$

$$\bullet f^n = f^{n-1} \circ f$$

By generalizing the formula, we can define the *transitive cover* of a function f and its inverse respectively as $f^+ = f^{+\infty}$ and $f^- = f^{-\infty}$. This cover is the application of the function to its result infinitely.

We also call *arity* the number of arguments (or the currying order) of a function noted $|f|$.

All this allows for a powerful notation system used throughout the rest of the document.

2.2.4 Properties

A modern approach of mathematics is called *reverse mathematics* as instead of building theorems from axioms, we search the minimal set of axioms required by a theorem. Inspired by this, we aim to minimize the formal basis of our system as well as identifying the circularity issues, we provide a dependency graph in figure 2.3.

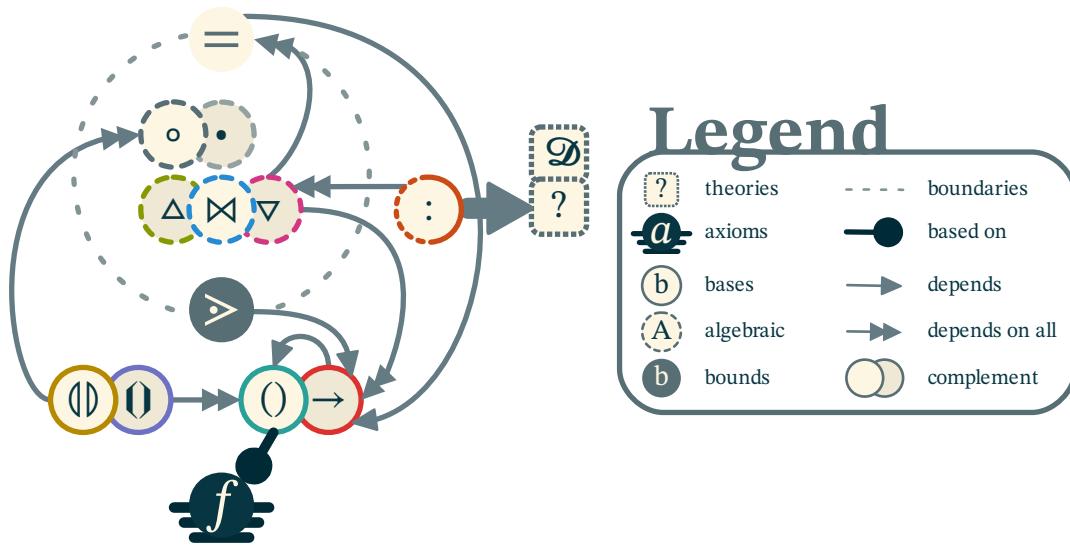


Figure 2.3: Dependency graph of notions in the functional theory

The algebra formed by the previously defined operations on functions is a semiring $(\mathbb{F}, \bowtie, \circ)$ with \mathbb{F} being the set of all functions.

Indeed, \bowtie is a commutative monoid having \gg as its identity element and \circ is a monoid with $(=)$ as its identity element.

Also the composition of the combination is the same as the combination of the composition. Therefore, \circ (composition) distributes over \bowtie (combination).

At last, using partial application composing with null gives null: $\circ(\gg) = ((\gg) \rightarrow \gg) \Rightarrow \gg$.

This foundation is now ready to define other fields of mathematics. We start with logic as it is a very basic formalism in mathematics. Then we redefine the ZFC set theory using our formalism as a base. And finally we will present derived mathematical tools to represent data structures and their properties.

2.3 First Order Logic

In this section, we present First Order Logic (FOL). FOL is based on boolean logic with the two literals T *true* and \perp *false*.

A function noted $(?)$ that have as only values either T or \perp is called a **predicate**.

$$\mathcal{D}(\bullet?) = \{\perp, T\}$$

We define the classical logic *entailment*, the predicate that holds true when a predicate (the conclusion) is true if and only if the first predicate (the premise) is true.

$$\vdash = (\perp, x \rightarrow T) \bowtie (T, x \rightarrow x)$$

Then we define the classical boolean operators \neg *not*, \wedge *and* and \vee *or* as:

- $\neg = (\perp \rightarrow T) \bowtie (T \rightarrow \perp)$, the negation associates true to false and false to true.
- $\wedge = x \rightarrow ((T \rightarrow x) \bowtie (\perp \rightarrow \perp))$, the conjunction is true when all its arguments are simultaneously true.
- $\vee = x \rightarrow ((T \rightarrow T) \bowtie (\perp \rightarrow x))$, the disjunction is true if all its arguments are not false.

The last two operators are curried function and can take any number of arguments as necessary and recursively apply their definition.

Functions that takes an expression as parameters are called *modifiers*. FOL introduces a useful kind of modifier used to modalize expressions: *quantifiers*. Quantifiers take an expression and a variable as arguments. Classical quantifiers are also predicates: they restrict the values that the variable can take.

The classical quantifiers are:

- The *universal quantifier* \forall meaning “for all”.
- The *existential quantifier* \exists meaning “it exists”.

$$\begin{aligned}\forall &= \S(\wedge) \\ \exists &= \S(\vee)\end{aligned}$$

They are sometimes extended with :

- The *uniqueness quantifier* $\exists!$ meaning “it exists a unique”.
- The *exclusive quantifier* \nexists meaning “it doesn’t exist”.

$$\begin{aligned}\exists! &= \S(= (1) \circ +) \\ \nexists &= \S(\neg \circ \wedge)\end{aligned}$$

Another exotic quantifier that isn’t a predicate can be proven useful (Hehner 2012):

- The *solution quantifier* \S meaning “those”.

$$\begin{aligned}\S &= f, x, ? \rightarrow \\ &\quad \{f(x) : ?\}\end{aligned}$$

The last three quantifiers are optional in FOL but will be conducive later on. It is interesting to note that most quantified expression can be expressed using the set builder notation discussed in the following section.



2.4 Set Theory

Since we need to represent knowledge, we will handle more complex data than simple booleans. The first theory we will define is the set theory. It is used as the classical foundation of mathematics. It is so important to mathematics that most other proposed foundation of mathematics invoke the concept of sets even before their first formula to describe the kind of notions they are introducing. The issue is then to define the sets themselves. At the beginning of his funding work on set theory, Cantor wrote:

Georg Cantor
(1895)

"A set is a gathering together into a whole of definite, distinct objects of our perception or of our thought—which are called elements of the set."

For Cantor, a set is a collection of concepts and percepts. In our case both notions are grouped in what we call *objects*, *entities* that are all ultimately *functions* in our formalism.

2.4.1 Base Definitions

This part is inspired by the work of Cantor (1895) and the set theory. The goal is to define the notions of set theory using our formalism.

Definition 15 (Set). A collection of *distinct* objects considered as an object in its own right. We define a set one of two ways (always using braces):

- In extension by listing all the elements in the set: $\{0, 1, 2, 3, 4\}$
- In intention by specifying the rule that all elements follow: $\{n : ?(n)\}$

Using our functional foundation, we can define any set as a predicate $\mathcal{S} = e \rightarrow T$ with e being a member of \mathcal{S} . This allows us to define the member function noted $e \in \mathcal{S}$ to indicate that e is an element of \mathcal{S} .

Another, useful definition using sets is the *domain* of a function f as the set of all arguments for which the function is defined. We call *co-domain* the domain of the inverse of a function. We can note them $f : \mathcal{D}(f) \mapsto \mathcal{D}(f)$. In the case of our functional version of sets, they are their own domain.

2.4.2 Set Operations

Along with defining the domains of functions using sets, we can use function on sets. This is very important in order to define ZFC and is extensively used in the rest of the document.

In this section, basic set operations are presented. The first one is the subset.

Definition 16 (Subset). A subset is a part of a set that is integrally contained within it. We note $\mathcal{S} \subset \mathcal{T} \vdash ((e \in \mathcal{S} \vdash e \in \mathcal{T}) \wedge \mathcal{S} \neq \mathcal{T})$, that a set \mathcal{S} is a proper subset of a more general set \mathcal{T} .

Definition 17 (Union). The union of two or more sets \mathcal{S} and \mathcal{T} is the set that contains all elements in *either* set. We can note it:

$$\mathcal{S} \cup \mathcal{T} = \{e : e \in \mathcal{S} \vee e \in \mathcal{T}\}$$

Definition 18 (Intersection). The intersection of two or more sets \mathcal{S} and \mathcal{T} is the set that contains only the elements member of *both* set. We can note it:

$$\mathcal{S} \cap \mathcal{T} = \{e : e \in \mathcal{S} \wedge e \in \mathcal{T}\}$$

Definition 19 (Difference). The difference of one set \mathcal{S} to another set \mathcal{T} is the set that contains only the elements contained in the first but not the last. We can note it:

$$\mathcal{S} \setminus \mathcal{T} = \{e : e \in \mathcal{S} \wedge e \notin \mathcal{T}\}$$

An interesting way to visualize relationships with sets is by using Venn diagrams. In figure 2.4 we present the classical set operations.

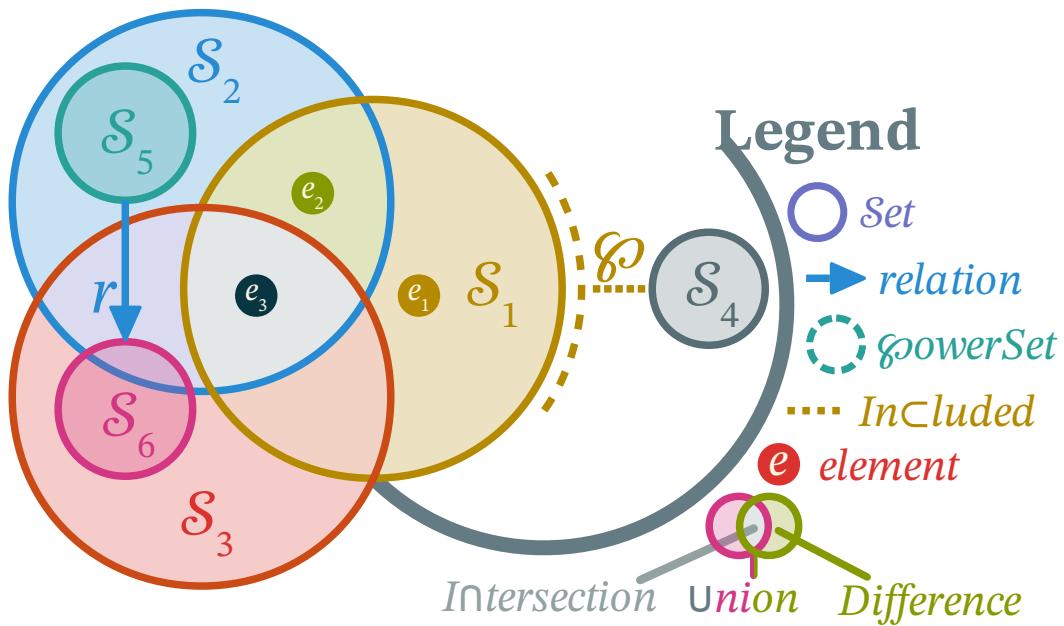


Figure 2.4: Example of an upgraded Venn diagram to illustrate operations on sets.

These diagrams have a lack of expressivity regarding complex operations on sets. Indeed, from their dimensionality it is complicated to express numerous sets having intersection and disjunctions. For example, it is difficult to represent the following notion.

Definition 20 (Cartesian product). The Cartesian product of two sets \mathcal{S} and \mathcal{T} is the set that contains all possible combination of an element of both sets. These combinations are a kind of ordered set called *tuples*. We note this product:

$$\mathcal{S} \times \mathcal{T} = \{\langle e_{\mathcal{S}}, e_{\mathcal{T}} \rangle : e_{\mathcal{S}} \in \mathcal{S} \wedge e_{\mathcal{T}} \in \mathcal{T}\}$$

From this we can also define the set power recursively by $\mathcal{S}^1 = \mathcal{S}$ and $\mathcal{S}^n = \mathcal{S} \times \mathcal{S}^{n-1}$.

The Cartesian product is the set equivalent of currying as:

$$\mathcal{S} \times \mathcal{T} = e_{\mathcal{S}}, e_{\mathcal{T}} \rightarrow \mathcal{S}(e_{\mathcal{S}}) \wedge \mathcal{T}(e_{\mathcal{T}})$$

The angles $\langle \rangle$ notation is used for tuples, those are another view on currying by replacing several arguments using a single one as an ordered list. A tuple of two elements is called a *pair*, of three elements a *triple*, etc. We can access elements in tuples using their index in the following way $e_2 = \langle e_1, e_2, e_3 \rangle_2$.

Definition 21 (Mapping). The mapping notation $\{\!\{ f(x) : x \in \mathcal{S} \}\!\}$ is a function such that $\{\!\{ f(x) : x \in \mathcal{S} \}\!\}$ will give the result of applying all elements in set \mathcal{S} as arguments of the function using the unCurrying operation recursively. If the function isn't specified, the mapping will select a member of the set non deterministically. The function isn't defined on empty sets or on sets with fewer members than arguments of the provided function.

Example 7. The classical sum operation on numbers can be noted:

$$\sum_{i=1}^3 2i = \{\!+(2 * i) : i \in [1, 3]\!\} = +(2 * 1)(+(2 * 2)(2 * 3))$$

2.4.3 The ZFC Theory

The most common axiomatic set theory is ZFC (Kunen 1980, vol. 102). In that definition of sets there are a few notions that come from its axioms. By being able to distinguish elements in the set from one another we assert that elements have an identity and we can derive equality from there:

Axiom (Extensionality). $\forall \mathcal{S} \forall \mathcal{T} : \forall e((e \in \mathcal{S}) = (e \in \mathcal{T})) \vdash \mathcal{S} = \mathcal{T}$

This means that two sets are equal if and only if they have all their members in common.

Another axiom of ZFC that is crucial in avoiding Russel's paradox ($\mathcal{S} \in \mathcal{S}$) is the following:

Axiom (Foundation). $\forall \mathcal{S} : (\mathcal{S} \neq \emptyset \vdash \exists \mathcal{T} \in \mathcal{S}, (\mathcal{T} \cap \mathcal{S} = \emptyset))$

This axiom uses the empty set \emptyset (also noted $\{\}$) as the set with no elements. Since two sets are equal if and only if they have precisely the same elements, the empty set is unique.

The definition by intention uses the set builder notation to define a set. It is composed of an expression and a predicate $? \cdot$ that will make any element e in a set \mathcal{T} satisfying it part of the resulting set \mathcal{S} , or as formulated in ZFC:

Axiom (Specification). $\forall? \forall \mathcal{T} \exists \mathcal{S} : (\forall e \in \mathcal{S} : (e \in \mathcal{T} \wedge ?(e)))$

The last axiom of ZFC we use is to define the power set $\wp(\mathcal{S})$ as the set containing all subsets of a set \mathcal{S} :

Axiom (Power set). $\wp(\mathcal{S}) = \{\mathcal{T} : \mathcal{T} \subseteq \mathcal{S}\}$

With the symbol $\mathcal{S} \subseteq \mathcal{T} \vdash (\mathcal{S} \subset \mathcal{T} \vee \mathcal{S} = \mathcal{T})$. These symbols have an interesting property as they are often used as a partial order over sets.

2.5 Graphs

With set theory, it is possible to introduce all of standard mathematics. A field of interest for this thesis is the study of the structure of data. Most of these structures use graphs and isomorphic derivatives.

Definition 22 (Graph). A graph is a mathematical structure g which is defined by its *connectivity function* χ that

2.5.1 Adjacency, Incidence and Connectivity

Definition 23 (Connectivity). The connectivity function is a combination of the classical adjacency and incidence functions of the graph. It is defined using a circular definition in the following way:

- *Adjacency*: $\chi_{\circ\circ} = v \rightarrow \{e : v \in \chi_{\rightarrow}(e)\}$
- *Incidence*: $\chi_{\leftarrow} = e \rightarrow \{v : e \in \chi_{\circ\circ}(v)\}$

Also: $\chi_{\circ\circ} = \bullet\chi_{\leftarrow}$

Defining either function defines the graph. For convenience, the connectivity function combine the adjacency and incidence:

$$\chi = \chi_{\circ\circ} \bowtie \chi_{\leftarrow}$$

Usually, graphs are noted $g = (V, E)$ with the set of vertices V (also called nodes) and edges E (arcs) that links two vertices together. Each edge is classically a pair of vertices ordered or not depending on if the graph is directed or not. It is possible to go from the set based definition to the functional relation using the following equation: $\mathcal{D}(\chi_{\leftarrow}) = E$

$$E \subseteq V^2$$

Example 8. A graph is often represented with lines or arrows linking points together like illustrated in figure 2.5. In that figure, the vertices v_1 and v_2 are connected through an undirected edge. Similarly v_3 connects to v_4 but not the opposite since they are bonded with a directed edge. The vertex v_8 is also connected to itself.

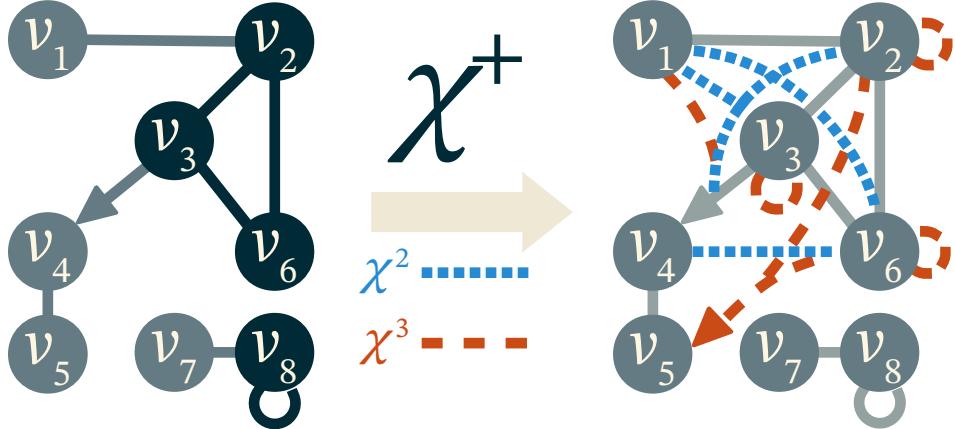


Figure 2.5: Example of the recursive application of the transitive cover to a graph.

2.5.2 Digraphs

In digraphs or *directional graphs* are a specific case of graphs where edges have a direction. This means that we can have two vertices v_1 and v_2 linked by an edge and while it is possible to go from v_1 to v_2 , the inverse is impossible. For such case the edges are ordered pairs and the incidence function can be decomposed into:

$$\chi_{\rightarrow} = \chi_{\rightarrow\rightarrow} \bowtie \chi_{\rightarrow\leftarrow}$$

We note $\chi_{\rightarrow\rightarrow}$ the **incoming relation** and $\chi_{\rightarrow\leftarrow}$ the **outgoing relation**.

In digraphs, classical edges can exist if allowed and will simply be bi-directional edges.

2.5.3 Path, cycles and transitivity

Most of the intrinsic information of a graph is contained within its structure. Exploring its properties require to study the “shape” of a graph and to find relationships between vertices. That is why graph properties are easier to explain using the transitive cover χ^+ of any graph $g = (V, E)$.

This transitive cover will create another graph in which two vertices are connected through an edge if and only if it exists a path between them in the original graph g . We illustrate this process in figure 2.5. Note how there is no edge in $\chi^2(g)$ between v_5 and v_6 and the one in $\chi^3(g)$ is directed towards v_5 because there is no path back to v_6 since the edge between v_3 and v_4 is directed.

Definition 24 (Path). We say that vertices v_1 and v_2 are *connected* if it exists a path from one to the other. Said otherwise, there is a path from v_1 to v_2 if and only if $\langle v_1, v_2 \rangle \in \mathcal{D}(\chi^+(g))$.

The notion of connection can be extended to entire graphs. An undirected graph g is said to be *connected* if and only if $\forall e \in V^2(e \in \mathcal{D}(\chi^+(g)))$.

Similarly we define *cycles* as the existence of a path from a given vertex to itself. For example, in figure 2.5, the cycles of the original graph are colored in blue. Some graphs can be strictly acyclical, enforcing the absence of cycles.

2.5.4 Trees

A **tree** is a special case of a graph. A tree is an acyclical connected graph. If a special vertex called a *root* is chosen, we call the tree a *rooted tree*. It can then be a directed graph with all edges pointing away from the root. When progressing away from the root, we call the current vertex *parent* of all exterior *children* vertices. Vertex with no children are called *leaves* of the tree and the rest are called *branches*.

An interesting application of trees to FOL is called *and/or trees* where each vertex has two sets of children: one for conjunction and the other for disjunction. Each vertex is a logic formula and the leaves are atomic logic propositions. This is often used for logic problem reduction. In figure 2.6 we illustrate how and/or trees are often depicted.

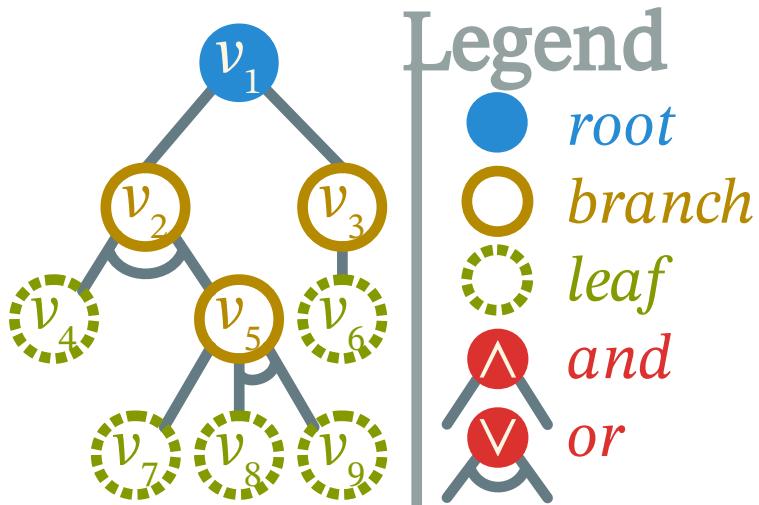


Figure 2.6: Example of and/or tree.

2.5.5 Quotient

Another notion often used for reducing big graphs is the quotienting as illustrated in figure 2.7.

Definition 25 (Graph Quotient). A quotient over a graph is the act of reducing a subgraph into a node while preserving the external connections. All internal structure becomes ignored and the subgraph now acts like a regular node. We note it $\div_f(g) =$

$(\{f(v) : v \in V\}, \{f(e) : e \in E\})$ with f being a function that maps any vertex either toward itself or toward its quotiented vertex.

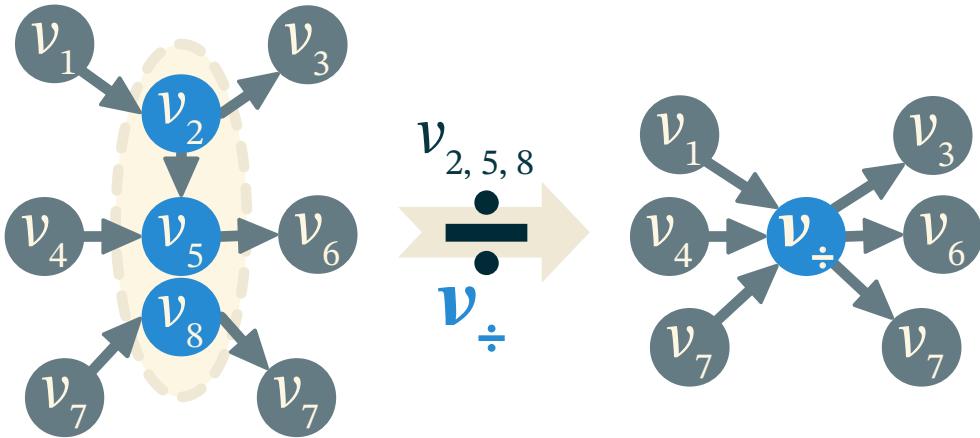


Figure 2.7: Example of graph quotient.

2.5.6 Hypergraphs

A generalization of graphs are **hypergraphs** where the edges are allowed to connect to more than two vertices. They are often represented using Venn-like representations but can also be represented with edges “gluing” several vertex like in figure 2.8.

An hypergraph is said to be *n-uniform* if the edges are restricted to connect to only *n* vertices together. In that regard, classical graphs are 2-uniform hypergraphs.

Hypergraphs have a special case where $E \subset V$. This means that edges are allowed to connect to other edges. In figure 2.8, this is illustrated by the edge e_3 connecting to three other edges. Information about these kinds of structures for knowledge representation is hard to come by and rely mostly on a form of “folk wisdom” within the mathematics community where knowledge is rarely published and mostly transmitted orally during lessons. One of the closest information available is this forum post (Kovitz 2018) that associated this type of graph to port graphs (Silberschatz 1981). Additional information was found in the form of a contribution of Vepstas (2008) on an encyclopedia article about hypergraphs. In that contribution, he says that a generalization of hypergraph allowing for edge-to-edge connections violate the axiom of **Foundation** of ZFC by allowing edge loops. Indeed, like in figure 2.8, an edge $e_9 = \{e_{10}\}$ can connect to another edge $e_{10} = \{e_9\}$ causing an infinite descent inside the \in relation in direct contradiction with ZFC.

This shows the limits of standard mathematics especially on the field of knowledge representation. Some structures needs higher dimensions than allowed by the structure of ZFC and FOL. However, it is important not to be mistaken: such non-standard set theories are more general than ZFC and therefore contains ZFC as a special case. All is a matter of restrictions.

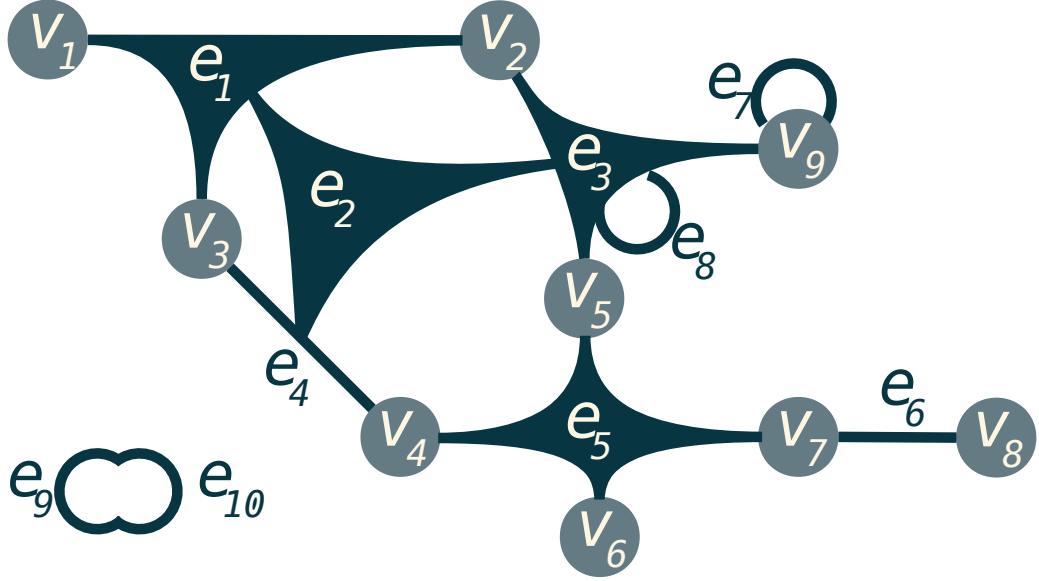


Figure 2.8: Example of hypergraph with total freedom on the edges specification.

2.6 Sheaf

In order to understand sheaves, we need to present a few auxiliary notions. Most of these definitions are adapted from (Vepštas 2008). The first of which is a seed.

Definition 26 (Seed). A seed corresponds to a vertex along with the set of adjacent edges. Formally we note a seed $\star = (v, \chi_g(v))$ that means that a seed build from the vertex v in the graph g contains a set of adjacent edges $\chi_g(v)$. We call the vertex v the *germ* of the seed. All edges in a seed do not connect to the other vertices but keep the information and are able to match the correct vertices through typing (often a type of a single individual). We call the edges in a seed *connectors*.

Seeds are extracts of graphs that contain all information about a vertex. Illustrated in the figure 2.9, seeds have a central germ (represented with discs) and connectors leading to a typed vertex (outlined circles). Those external vertices are not directly contained in the seed but the information about what vertex can fit in them is kept. It is useful to represent connectors like jigsaw puzzle pieces: they can match only a restricted number of other pieces that match their shape.

From there, it is useful to build a kind of partial graph from seeds called sections.

Definition 27 (Section). A section is a set of seeds that have their common edges connected. This means that if two seeds have an edge in common connecting both germs, then the seeds are connected in the section and the edges are merged. We note $g_\star = (V, \{\cup : E_{section}\})$ the graph formed by the section.

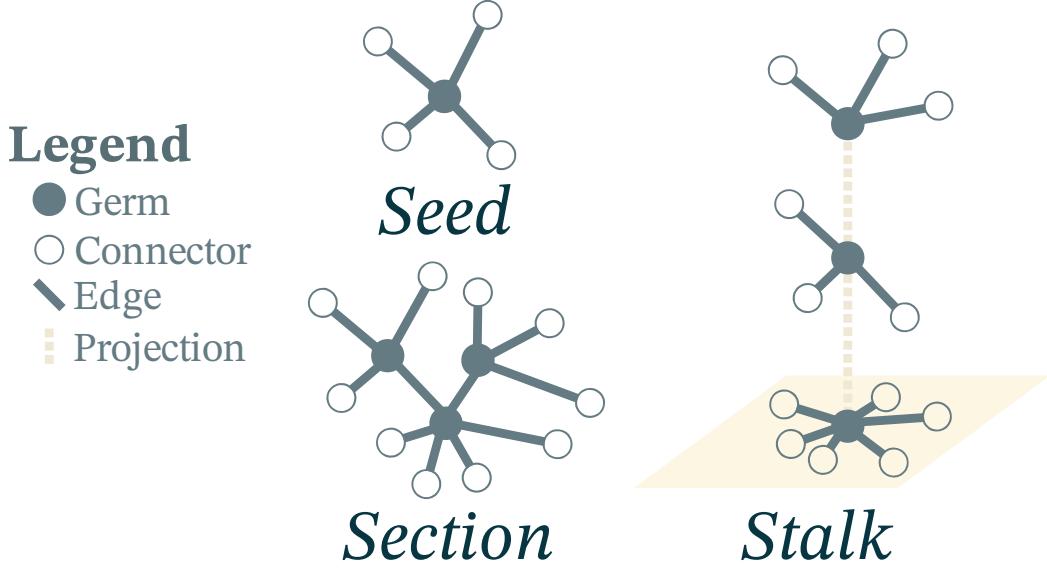


Figure 2.9: Example of a seed, a section and a stalk.

In figure 2.9, a section is represented. It is a connected section composed of seeds along with the additional seeds of any vertices they have in common. They are very similar to subgraph but with an additional border of typed connectors. This tool was originally mostly meant for big data and categorization over large graphs. As the graph quotient is often used in that domain, it was ported to sections instead of graphs allows us to define stalks.

Definition 28 (Stalk). Given a projection function $f : V \rightarrow V'$ over the germs of a section \star , the stalk above the vertex $v' \in V'$ is the quotient of all seeds that have their germ follow $f(v) = v'$.

The quotienning is used in stalks for their projection. Indeed, as shown in figure 2.9, the stalks are simply a collection of seeds with their germs quotiented into their common projection. The projection can be any process of transformation getting a set of seeds in one side and gives object in any base space called the image. Sheaves are a generalization of this concept to sections.

Definition 29 (Sheaf). A sheaf is a collection of sections, together with a projection. We note it $\mathcal{F} = \langle g_\star, glue \rangle$ with the function *glue* being the gluing axioms that the projection should respect depending on the application. The projected sheaf graph is noted as the fusion of all quotiented sections:

$$glue_{\mathcal{F}} = \{\div glue_\star : \{glue_\star \in g_\star\}\}$$

By merging common vertices into a section, we can build stack fields. These fields are simply a subcategory of sheaves. Illustrated in figure 2.10, a sheaf is a set of section with a projection relation.

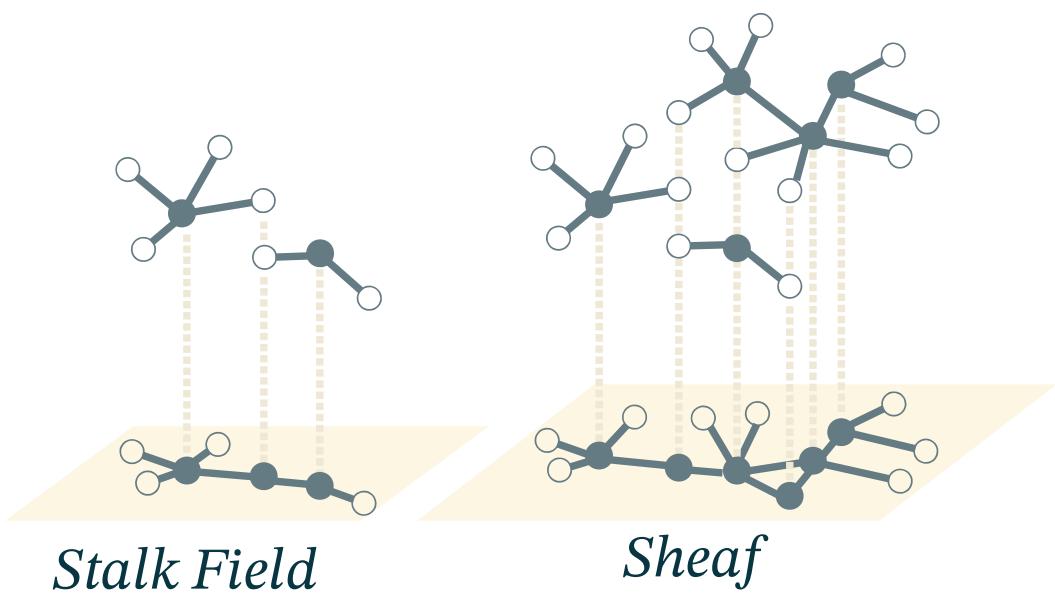


Figure 2.10: Example of sheaves.



Figure 3.1: Noam Chomsky 2017

3 Knowledge Representation

Knowledge representation is at the intersection of maths, logic, language and computer sciences. Knowledge description systems rely on syntax to interoperate systems and users to one another. The base of such languages comes from the formalization of automated grammars by Chomsky (1956). It mostly consists of a set of production rules aiming to describe all accepted input strings. Usually, the rules are hierarchical and deconstruct the input using simpler rules until it matches a terminal symbol. This deconstruction is called parsing and is a common operation in computer science. More tools for the characterization of computer language emerged soon after thanks to Backus (1959) while working on a programming language at IBM. This is how the Backus-Naur Form (BNF) metalanguage was created on top of Chomsky's formalization.

A similar process happened in the 1970s, when logic based knowledge representation gained popularity among computer scientists (Baader 2003). Systems at the time explored notions such as rules and networks to try and organize knowledge into a rigorous structure. At the same time other systems were built based on First Order Logic (FOL). Then, around the 1990s, the research began to merge in search of common semantics in what led to the development of Description Logics (DL). This domain is expressing knowledge as a hierarchy of classes containing individuals.

From there and with the advent of the world wide web, actors of the internet were on the lookout for standardization and interoperability of computer systems. One such standardization took the name of "semantic web" and aimed to create a widespread network of connected services sharing knowledge between one another in a common language. At the beginning of the 21st century, several languages were created, all based on the World Wide Web Consortium (W3C) specifications called Resource Description Framework (RDF) (Klyne and Carroll 2004). This language is based on the notion of statements as triples. Each can express a unit of knowledge. All the underlying theoretical work of DL continued with it and created more expressive derivatives. One such derivative is the family of languages called Web Ontology Language (OWL) (Horrocks *et al.* 2003). The ontologies and knowledge graphs are more recent names for the representation and definition of categories (DL classes), properties and relation between concepts, data and entities.

Nowadays, when designing a knowledge representation, one usually starts with existing framework. The most popular in practice is certainly the classical relational database, followed closely by more novel methods for either big data or more expressive solutions like ontologies.

In our case we need a tool that is more expressive than ontologies while remaining efficient. Of course, this will lead to compromises, but can also have some interesting properties.

3.1 Grammar and Parsing

Grammar is an old tool that used to be dedicated to linguists. With the funding works by Chomsky and his Context-Free Grammars (CFG), these tools became available to mathematicians and shortly after to computer scientists.

A CFG is a formal grammar that aims to generate a formal language given a set of hierarchical rules. Each rule is given a symbol as a name. From any finite input of text in a given alphabet, the grammar should be able to determine if the input is part of the language it generates.

3.1.1 BNF

In computer science, popular metalanguage called BNF was created shortly after Chomsky's work on CFG. The syntax is of the following form :

```
1 <rule> ::= <other_rule> | <terminal_symbol> | "literals"
```

A terminal symbol is a rule that does not depend on any other rule. It is possible to use recursion, meaning that a rule will use itself in its definition. This actually allows for infinite languages. Despite its expressive power, BNF is often used in one of its extended forms.

In this context, we present a widely used form of BNF syntax that is meant to be human readable despite not being very formal. We add the repetition operators `*` and `+` that respectively repeat 0 and 1 times or more the preceding expression. We also add the negation operator `~` that matches only if the following expression does not match. We also add parentheses for grouping expression and brackets to group literals.

Example 9. We can make a grammar for all sequence of `A` using the rule `<scream> ::= "A" +`. If we want to make a rule that prevent the use of the letter `z` we can write `<no-sleep> ::= ~"z"`.

3.1.2 Dynamic Grammar

A regular grammar is static, it is set once and for all and will always produce the same language. In order to be more flexible we need to talk about dynamic grammars and their associated tools and explain our choice of grammatical framework.

One of the main tools for both static and dynamic grammar is a parser. It is the program that will interpret the input into whatever usage it is meant for. Most of the time, a parser will transform the input into another similarly structured language. It can be a storage inside objects or memory, or compiled into another format, or even just for syntax coloration. Since a lot of usage requires the same kind of function, a new kind of tool emerged to make the creation of a parser simpler. We call those tools parser or compiler generators (Paulson 1982). They take a grammar description as input and gives the program of a parser of the generated language as an output.

For dynamic grammar, these tools can get more complicated. There are a few ways a grammar can become dynamic. The most straightforward way to make a parser dynamic is to introduce code in the rule handling that will tweak variables affecting the parser itself (Souto *et al.* 1998). This allows for handling context in CFG without needing to rewrite the grammar.

Another kind of dynamic grammar is grammar that can modify themselves. In order to do this a grammar is evaluated with reified objects representing parts of itself (Hutton and Meijer 1996). These parts can be modified dynamically by rules as the input gets parsed (Renggli *et al.* 2010; Alessandro and Piumarta 2007). This approach uses Parsing Expression Grammars (PEG)(Ford 2004) with Packrat parsing that Packrat parsing backtracks by ensuring that each production rule in the grammar is not tested more than once against each position in the input stream (Ford 2002). While PEG is easier to implement and more efficient in practice than their classical counterparts (Loff *et al.* 2018; Henglein and Rasmussen 2017), it offsets the computation load in memory making it actually less efficient in general (Becket and Somogyi 2008).

Some tools actually just infer entire grammars from inputs and software (Höschele and Zeller 2017; Grünwald 1996). However, these kinds of approaches require a lot of input data to perform well. They also simply provide the grammar after expensive computations.

My system uses a grammar, composed of classical rules and is extended using meta-rules that activates once the classical grammar fails.

3.2 Description Logics

One of the most standard and flexible way of representing knowledge is by using ontologies. They are based mostly on the formalism of Description Logics (DL). It is based on the notion of classes (or types) as a way to make the knowledge hierarchically structured. A class is a set of individuals that are called instances of the classes. Classes got the same basic properties as sets but can also be constrained with logic formula. Constraints can be on anything about the class or its individuals. Knowledge is also encoded in relations that are predicates over attributes of individuals.

It is common when using DLs to store statements into three boxes (Baader 2003):

- The TBox for terminology (statements about types)
- The RBox for rules (statements about properties) (Bürckert 1994)
- The ABox for assertions (statements about individual entities)

These are used mostly to separate knowledge about general facts (intentional knowledge) from specific knowledge of individual instances (extensional knowledge). The extra RBox is for “knowhow” or knowledge about entity behavior. It restricts usages of roles (properties) in the ABox. The terminology is often hierarchically ordered using a subsumption relation noted \subseteq . If we represent classes or type as a set of individuals then this relation is akin to the subset relation of set theory.

There are several versions and extensions of DL. They all vary in expressivity. Improving the expressivity of DL system often comes at the cost of less efficient inference engines that can even become undecidable for some extensions of DL.

3.3 Ontologies and their Languages

Most AI problem needs a way to represent knowledge. The classical way to do so has been more and more specialized for each AI community. Each their Domain Specific Language (DSL) that neatly fit the specific use it is intended to do. There was a time when the branch of AI wanted to unify knowledge description under the banner of the "semantic web". From numerous works, a repeated limitation of the "semantic web" seems to come from the languages used. In order to guarantee performance of generalist inference engines, these languages have been restricted so much that they became quite complicated to use and quickly cause huge amounts of recurrent data to be stored because of some forbidden representation that will push any generalist inference engine into undecidability.

The most basic of these languages is perhaps RDF Turtle (Beckett and Berners-Lee 2011). It is based on triples with an XML syntax and has a graph as its knowledge structure (Klyne and Carroll 2004). A RDF graph is a set of RDF triples $\langle \text{sub}, \text{pro}, \text{obj} \rangle$ which fields are respectively called subject, property and object. It can also be seen as a partially labeled directed graph (V, E) with V being the set of RDF nodes and E being the set of edges. This graph also comes with an incomplete label relation that associates a unique string called a Uniform Resource Identifier (URI) to most nodes. Nodes without an URI are called blank nodes. It is important that, while not named, blank nodes have a distinct internal identifier from one another that allows to differentiate them.

Built on top of RDF, the W3C recommended another standard called OWL. It adds the ability to have hierarchical classes and properties along with more advanced description of their arity and constraints. OWL is, in a way, more expressive than RDF (Van Harmelen et al. 2008, 1,p825). It adds most formalism used in knowledge representation and is widely used and interconnected. OWL comes in three versions: OWL Lite, OWL DL and OWL Full. The lite version is less advanced but its inference is decidable, OWL DL contains all notions of DL and the full version contains all features of OWL but is strongly undecidable.

```
1 ex:ontology rdf:type owl:Ontology .
2 ex:name rdf:type owl:DatatypeProperty .
3 ex:author rdf:type owl:ObjectProperty .
4 ex:Book rdf:type owl:Class .
5 ex:Person rdf:type owl:Class .
6
7 _:x rdf:type ex:Book .
8 _:x ex:author _:x1 .
9 _:x1 rdf:type ex:Person .
10 _:x1 ex:name "Fred"^^xsd:string .
```

W3C (2004a)

The expressivity can also come from a lack of restriction. If we allow some freedom of expression in RDF statements, its inference can quickly become undecidable (Motik

2007). This kind of extremely permissive language is better suited for specific usage for other branches of AI. Even with this expressivity, several works still deem existing ontology system as not expressive enough, mostly due to the lack of classical constructs like lists, parameters and quantifiers that don't fit the triple representation of RDF.

One of the ways which have been explored to overcome these limitations is by adding a 4th field in RDF. This field is meant for context and annotations. This field is used for information about any statement represented as a triple, such as access rights, probabilities, or most of the time the source of the data (Tolksdorf *et al.* 2004). One of the other uses of the fourth field of RDF is to reify statements (Hernández *et al.* 2015). Indeed by identifying each statement, it becomes possible to efficiently form statements about statements.

A completely different approach is done by Hart and Goertzel (2008) in his framework for Artificial General Intelligence (AGI) called OpenCog. The structure of the knowledge is based on a rhizome, a collection of trees, linked to each other. This structure is called Atomspace. Each vertex in the tree is an atom, leaf vertexes are nodes, the others are links. Atoms are immutable, indexed objects. They can be given values that can be dynamic and, since they are not part of the rhizome, are an order of magnitude faster to access. Atoms and values alike are typed.

The goal of such a structure is to be able to merge concepts from widely different domains of AI. The major drawback being that the whole system is very slow compared to pretty much any domain specific software.

In my system, a similar structure is used but along with ontology oriented notions.

3.4 Self

As we have seen, the most used knowledge description systems (e.g. RDF, Ontologies and relational databases) have a common drawback: they are static. This means that they are created to be optimized for a specific use case, or gets general at the cost of efficiency. The main issue is that such system are unable to adapt to the use case by themselves. To fix this issue, a new knowledge representation model must be presented. The goal is to make a minimal language framework that can adapt to its use to become as specific as needed. If it becomes specific it must start from a generic base. Since that base language must be able to evolve to fit the most cases possible, it must be neutral and simple.

To summarize, that framework must maximize the following criteria:

1. **Neutral:** Must be independent from preferences and be localization.
2. **Permissive:** Must allow as many data representation as possible.
3. **Minimalist:** Must have the minimum number of base axioms and as little native notions as possible.
4. **Adaptive:** Must be able to react to user input and be as flexible as possible.

In order to respect these requirements, we developed a framework for knowledge description. This Structurally Expressive Language Framework (SELF) is our answer to these criteria. SELF is inspired by RDF Turtle and Description Logic.

3.4.1 Knowledge Structure

SELF extends the RDF graphs by adding another label to the edges of the graph to uniquely identify each statement. This basically turns the system into a quadruple storage even if this forth field is transparent to the user.

Axiom (Structure). A SELF graph is a set of statements that transparently include their own identity. The closest representation of the underlying structure of SELF is as follows:

$$g_{\mathbb{U}} = (\mathbb{U}, S) : S = \{s = \langle sub, pro, obj \rangle : s \in \mathcal{D} \vdash s \wedge \mathcal{D}\}$$

with:

- $sub, obj \in \mathbb{U}$ being entities representing the *subject* and *object* of the *statement* s ,
- $pro \in P$ being the *property* of the statement s ,
- $\mathcal{D} \subset S$ is the *domain* of the *world* $g_{\mathbb{U}}$,
- $S, P \subset \mathbb{U}$ with S the set of statements and P the set of properties,

This means that the world $g_{\mathbb{U}}$ is a graph with the set of entities \mathbb{U} as vertices and the set of statements S as edges. This model also suppose that every statement s must be true if they belong to the domain \mathcal{D} . This graph is a directed 3-uniform hypergraph.

Since sheaves are a representation of hypergraphs, we can encode the structure of SELF into a sheaf-like form. Each seed is a statement, the germ being the statement vertex. It is always accompanied of an incoming connector (its subject), an outgoing connector (its object) and a non-directed connector (its property). The sections are domains and must be coherent. Each statement, along with its property, makes a stalk as illustrated in figure 3.2.

The difference with a sheaf is that the projection function is able to map the pair statement-property into a labeled edge in its projection space. We map this pair into a classical labeled edge that connects the subject to the object of the statement in a directed fashion. This results in the projected structure being a correct RDF graph.

3.4.1.1 Consequences

The base knowledge structure is more than simply convenience. The fact that statements have their own identity, changes the degrees of freedom of the representation. RDF has a way to represent reified statements that are basically blank nodes with properties that are related to information about the subject, property and object of a designated statement. The problem is that such statements are very differently represented and need 3 regular statements just to define. Using the fourth field, it becomes possible to make statements about *any* statements. It also becomes possible to

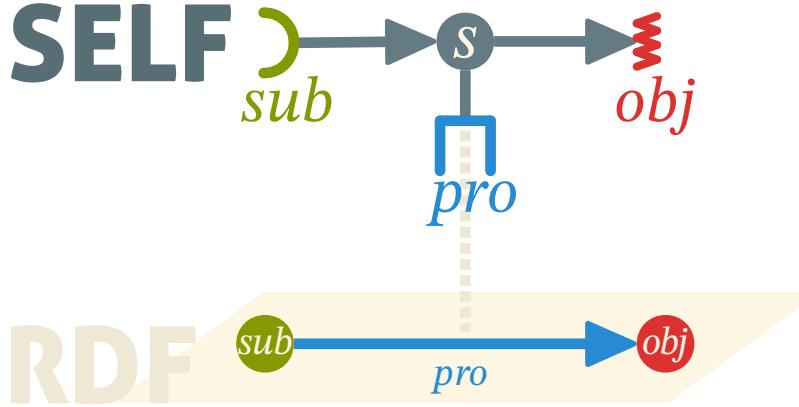


Figure 3.2: Projection of a statement from the SELF to RDF space.

express modal logic about statements or to express, various traits like the probability or the access rights of a statement.

The knowledge structure holds several restrictions on the way to express knowledge. As a direct consequence, we can add several theorems to the logic system underlying SELF. The axiom of [Structure](#) is the only axiom of the system.

Theorem 1 (Identity). *Any entity is uniquely distinct from any other entity.*

This theorem comes from the axiom of [Extensionality](#) of ZFC. Indeed it is stated that a set is an unordered collection of distinct objects. Distinction is possible if and only if intrinsic identity is assumed. This notion of identity entails that a given entity cannot change in a way that would alter its identifier.

Theorem 2 (Consistency). *Any statement in a given domain is consistent with any other statements of this domain.*

Consistency comes from the need for a coherent knowledge system and is often a requirement of such constructs. This theorem also is a consequence of the axiom of [Structure](#): $s \in \mathcal{D} \vdash s \wedge \mathcal{D}$.

Theorem 3 (Uniformity). *Any object in SELF is an entity. Any relations in SELF are restricted to \mathbb{U} .*

This also means that all native relations are closed under \mathbb{U} . This allows for a uniform knowledge database.

3.4.1.2 Native Properties

In the following, we suppose all notions from previous chapter. The difference is that we define and use only a subset of the functions defined in the SELF formalism. In

relation to the theory of SELF, we use the functional theory previously defined as the underlying formalism.

FIXME: Case variation for crossref

Theorem 1 lead to the need for two native properties in the system : *equality* and *name*.

The **equality relation** $=: \mathbb{U} \rightarrow \mathbb{U}$, behaves like the classical operator. Since the knowledge database will be expressed through text, we also need to add an explicit way to identify entities. This identification is done through the **name relation** $\nu : \mathbb{U} \rightarrow L_{String}$ that affects a string literal to some entities. This lead us to introduce literals into SELF that is also entities that have a native value.

The axiom of **Structure** puts a type restriction on property. Since it compartments \mathbb{U} using various named subsets, we must adequately introduce an explicit type system into SELF. That type system requires a **type relation** (named using the colon) $:: : \mathbb{U} \rightarrow T$. That relation is complete as all entities have a type. Theorem 3 causes the set of entities to be universal. Type theory, along with Description Logic (DL), introduces a **subsumption relation** $\subseteq : T \rightarrow T$ as a partial ordering relation to the types. Since types can be seen as sets of instances, we simply use the subset relation from set theory. In our case, the entity type is the greatest element of the lattice formed by the set of types with the subsumption relation (T, \subseteq) .

The theorem 3 also allows for some very interesting meta-constructs. That is why we also introduce a signed **Meta relation** $\mu : \mathbb{U} \rightarrow D$ with $\mu^* = \bullet\mu$. This allows to create domain from certain entities and to encapsulate domains into entities. μ^* is for reification and μ is for abstraction. This Meta relation also allows to express value of entities, like lists or various containers.

To fulfill the principle of adaptability and in order to make the type system more useful, we introduce the **parameter relation** $\rho : \mathbb{U} \rightarrow \mathbb{U}$. This relation affects a list of parameters, using the Meta relation, to some parameterized entities. This also allows for variables in statements.

Since axiom of **Structure** gives the structure of SELF a hypergraph shape, we must port some notions of graph theory into our framework. Introducing the **statement relation** $\chi : S \rightarrow \mathbb{U}$ reusing the same symbol as for the adjacency and incidence relation of graphs. This isn't a coincidence as this relation has the same properties.

Example 10. Since statements are triplets and edges, s_0 gives the subject of a statement s . Respectively, s_1 and s_2 give the property and object of any statement. For adjacencies, χ can give the set of statements any entity is the object or subject of. For any property pro , the notation $\chi(pro)$ gives the set of statements using this property.

This allows us to port all the other notions of graphs using this relation as a base.

In figure 3.3, we present all the native relations along with their domains and most subsets of \mathbb{U} .

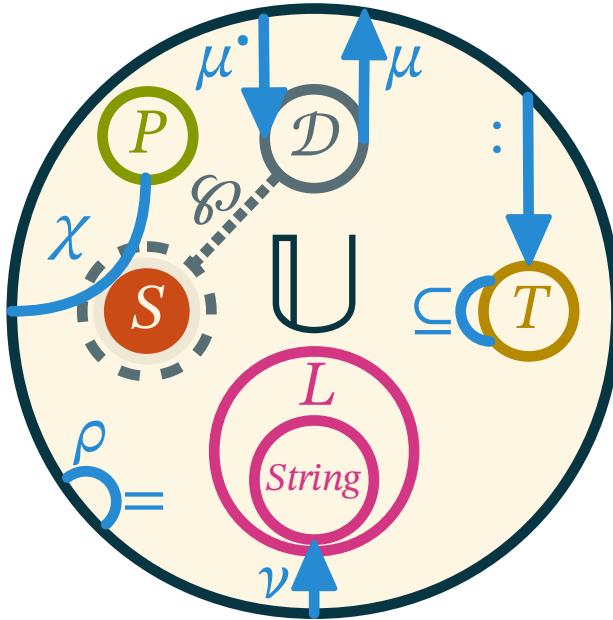


Figure 3.3: Venn diagram of subsets of \mathbb{U} along with their relations. Dotted lines mean that the sets are defined as subset of the wider set.

3.4.2 Syntax

Since we need to respect the requirements of the problem, the RDF syntax cannot be used to express the knowledge. Indeed, RDF states native properties as English nodes with a specific URI that isn't neutral. It also isn't minimalist since it uses an XML syntax so verbose that it is not used for most examples in the documents that defines RDF because it is too confusing and complex (W3C 2004b; W3C 2004c). The XML syntax is also quite restrictive and cannot evolve dynamically to adapt to the usage.

We need to define a new language that has contradictive qualities. It must be general, yet specific and minimalistic while expressive.

So the solution to the problem is to actually define two languages that fit the criteria: one minimalist and one adaptive. The issue is that we don't want the user to learn two languages and the second kind of language must be very specific and that violates the principle of neutrality we try to respect.

The only solution is to make a mechanism to adapt the language as it is used. We start off with a simple framework that uses a grammar.

The description of g_0 is pretty straightforward: it mostly is just a triple representation separated by whitespaces. The goal is to add a minimal syntax consistent with the axiom of [Structure](#). In listing 3.1, we give a simplified version of g_0 . It is written in a pseudo-BNF fashion, which is extended with the classical repetition operators * and + along with the negation operator ~. All tokens have names in uppercase. We also add the following rule modifiers:

- `<~name>` are ignored for the parsing. However, the tokens are consumed and therefore acts like separators for the other rules.
- `<?name>` are inferred rules and tokens. They play a key role for the process of derivation explained in section 3.4.3.

```

1 <~COMMENT: <INLINE: "://" (~["\n", "\r"])*>
2 | <BLOCK: "/*" (~[*/*])* > //Ignored
3 <~WHITE_SPACE: " " | "\t" | "\n" | "\r" | "\f">
4 <LITERAL: <INT> | <FLOAT> | <CHAR> | <STRING>> //Java definition
5 <ID: <TYPE: <UPPERCASE>(<LETTERS>|<DIGITS>)* >
6 | <ENTITY: <LOWERCASE>(<LETTERS>|<DIGITS>)*>
7 | <SYMBOL: (~[<LITERALS>, <LETTERS>, <DIGITS>])*>>
8
9 <worselfld> ::= <first> <statement>* <EOF>
10 <first> ::= <subject> <?EQUAL> <?SOLVE> <?EOS>
11 <statement> ::= <subject> <property> <object> <EOS>
12 <subject> ::= <entity>
13 <property> ::= <ID> | <?meta_property>
14 <object> ::= <entity>
15 <entity> ::= <ID> | <LITERAL> | <?meta_entity>

```

Listing 3.1: Simplified pseudo-BNF description for basic SELF.

In order to respect the principle of neutrality, the language must not suppose of any regional predisposition of the user. There are few exceptions for the sake of convenience and performance. The first exception is that the language is meant to be read from left to right and have an occidental biased `subject verb object` triple description. Another exception is for literals that use the same grammar as in classical Java. This means that the decimal separator is the dot (.). This concession is made for reasons of simplicity and efficiency, but it is possible to define literals dynamically in theory (see section 7.2.1).

Even if sticking to the ASCII subset of characters is a good idea for efficiency, SELF can work with UTF-8 and exploits the Unicode Character Database (UCD) for its token definitions (Unicode Consortium 2018b). This means that SELF comes keywords free and that the definition of each symbol is left to the user. Each notion and symbol is inferred (with the exception of the first statement which is closer to an imposed configuration file).

In g_0 , the first two token definitions are ignored. This means that comments and white-spaces will act as separation and won't be interpreted. Comments are there only for convenience since they do not serve any real purpose in the language. It was arbitrarily decided to use Java-style comments. White-spaces are defined against UCD's definition of the separator category `Zs` (see Unicode Consortium 2018a, chap. 4).

FIXME: Lines ref are borken

Line 4 uses the basic Java definition for liberals. In order to keep the independence from any natural language, boolean laterals are not natively defined (since they are English words).

Another aspect of that language independence is found starting at line 5 where the definitions of `<UPPERCASE>`, `<LOWERCASE>`, `<LETTERS>` and `<DIGITS>` are defined from the UCD (respectively categories `Lu`, `Ll`, `L`, `Nd`). This means that any language's upper case can

be used in that context. For performance and simplicity reasons we will only use ASCII in our examples and application.

The rule at line 1 is used for the definition of three tokens that are important for the rest of the input. `<EQUAL>` is the symbol for equality and `<SOLVE>` is the symbol for the *solution quantifier* (and also the language pendant of μ^*). The most useful token `<EOS>` is used as a statement delimiter. This rule also permits the inclusion of other files if a string literal is used as a subject. The underlying logic of this first statement will be presented in section 3.4.5.1. In the following examples we will consider that `<EQUAL> ::= "="`, `<SOLVE> ::= "?"` and `<EOS> ::= ";"`.

At line 11, we can see one of the most defining features of g_0 : statements. The input is nothing but a set of statements. Each component of the statements are entities. We defined two specific rules for the subject and object to allow for eventual runtime modifications. The property rule is more restricted in order to guarantee the non-ambiguity of the grammar.

3.4.3 Dynamic Grammar

The syntax we described is only valid for g_0 . As long as the input is conforming to these rules, the framework keeps the minimal behavior. In order to access more features, one needs to break a rule. We add a second outcome to handling with violations : **derivation**. There are several kinds of possible violations that will interrupt the normal parsing of the input :

- Violations of the `<first>` statement rule : This will cause a fatal error.
- Violations of the `<statement>` rule : This will cause a derivation if an unexpected additional token is found instead of `<EOS>`. If not enough tokens are present, a fatal error is triggered.
- Violations of the secondary rules (`<subject>`, `<entity>`, ...) : This will cause a fatal error except if there is also an excess of token in the current statement which will cause derivation to happen.

Derivation will cause the current input to be analyzed by a set of meta-rules. The main restriction of these rules is given in g_0 : each statement must be expressible using a triple notation. This means that the goal of the meta-rules is to find an interpretation of the input that is reducible to a triple and to augment g_0 by adding an expression to any `<meta_*>` rules. If the input has fewer than 3 entities for a statement then the parsing fails. When there is extra input in a statement, there is a few ways the infringing input can be reduced back to a triple.

3.4.3.1 Containers

The first meta-rule is to infer a container. A container is delimited by, at least, a left and right delimiter (they can be the same symbol) and an optional middle delimiter. We infer the delimiters using the algorithm 1.

Algorithm 1 Container meta-rule

```
1: function CONTAINER(Token current)
2:   LOOKAHEAD(current, EOS)                                > Populate all tokens of the statement
3:   for all token in horizon do
4:     if token is a new symbol then delimiters.APPEND(token)
5:   if LENGTH(delimiters) < 2 then
6:     if COHERENTDELIMITERS(horizon, delimiters[0]) then
7:       INFERMIDDLE(delimiters[0])                         > New middle delimiter in existing containers
8:       return Success
9:     return Failure
10:    while LENGTH(delimiters) > 0 do
11:      for all (left, middle, right) in SORTEDDELIMITERS(delimiters) do
12:        if COHERENTDELIMITERS(horizon, left, middle, right) then
13:          INFERDELIMITER(left, right)
14:          INFERMIDDLE(middle)                            > Ignored if null
15:          delimiters.REMOVE(left, middle, right)
16:          break
17:        if LENGTH(delimiters) stayed the same then return Success
18:    return Success
```

The function sortedDelimiters at line 11 is used to generate every ordered possibility and sort them using a few criteria. The default order is possibilities grouped from left to right. All coupled delimiters that are mirrors of each other following the UCD are preferred to other possibilities.

Checking the result of the choice is very important. At line 12 a function checks if the delimiters allow for triple reduction and enforce restrictions.

Example 11. For example, a property cannot be wrapped in a container (except if part of parameters). This is done in order to avoid a type mismatch later in the interpretation.

Once the inference is done, the resulting calls to inferDelimiter will add the rules listed in listing 3.2 to \mathbb{G}_0 . This function will create a `<container>` rule and add it to the definition of `<meta_entity>`. Then it will create a rule for the container named after the UCD name of the left delimiter (searching in the `NamesList.txt` file for an entry starting with "left" and the rest of the name or defaulting to the first entry). Those rules are added as a conjunction list to the rule `<container>`. It is worthy to note that the call to inferMiddle will add rules to the token `<MIDDLE>` independently from any container and therefore, all containers share the same pool of middle delimiters.

```
1 <meta_entity> ::= <container>
2 <container> ::= <parenthesis> | ...
3 <parenthesis> ::= "(" [<naked_entity>] (<?MIDDLE> <naked_entity>)* ")"
4 <naked_entity> ::= <statement> | <entity>
```

Listing 3.2: Rules added to the current grammar for handling the new container for parenthesis

The rule at line 4 is added once and enables the use of meta-statements inside containers. It is the language pendant of the μ relation, allowing to wrap abstraction in a safe way.

Example 12. If we parse the expression `a = (b, c);`, we start by tokenizing it as `<ENTITY> <EQUAL> <SYMBOL><ENTITY><SYMBOL><ENTITY><SYMBOL> <EOS>` (ignoring whitespaces and comments). This means that the statement is 4 tokens too long to form a triple. This triggers a parsing error and then an evaluation using meta-rules. All the `<ID>` tokens are new symbols, but they don't have the same subtype. This means that candidate delimiters are `(`, `,` and `)`. To Infer the inference there's only one combination and the left delimiter and right delimiters are found via their Unicode description. The comma is left to be inferred as the middle delimiter. The grammar is rewritten and the statement becomes `<ENTITY> <EQUAL> <container> <EOS>` which is a valid triple statement.

3.4.3.2 Parameters

If the previous rule didn't fix the parsing of the statement, we continue with the following meta-rule. Parameters are extra containers that are used after an entity. Every container can be used as parameters. We detail the analysis in algorithm 2.

Algorithm 2 Parameter meta-rule

```

1: function PARAMETER(Entity[] statement)
2:   reduced = statement
3:   while LENGTH(reduced) > 3 do
4:     for i from 0 to LENGTH(reduced) - 1 do
5:       if NAME(reduced[i]) not null and
6:         TYPE(reduced[i+1]) = Container and
7:           COHERENTPARAMETERS(reduced, i) then
8:             param = INFERPARAMETER(reduced[i], reduced[i+1])
9:             reduced.REMOVE(reduced[i], reduced[i+1])
10:            reduced.INSERT(param, i)                                > Replace parameterized entity
11:            break
12:            if LENGTH(statement) stayed the same then return Success
13:   return Failure

```

The goal is to match extra containers with the preceding named entity. The container is then combined with the preceding entity into a parameterized entity.

The call to inferParameter will add the rule in listing 3.3, replacing `<?container>` with the name of the container used.

Example 13. In this case we have to parse `f(x) = x;;`. If we already have the parenthesis delimiter defined, it becomes `<ENTITY> <container> <EQUAL> <ENTITY> <EOS>`. Since the statement isn't a triple we execute the meta-rules. The container rule finds no new symbols and fails. Then the parameter meta-rule will reduce the statement by bounding the first entity to the following container and mark it as a parameterized entity. This gives `<meta_entity> <EQUAL> <ENTITY> <EOS>`.

```

1 <meta_entity> ::= <ID> <?container>
2 <meta_property> ::= <ID> <?container>

```

Listing 3.3: Rules added to the current grammar for handling parameters

3.4.3.3 Operators

A shorthand for parameters is the operator notation. It allows to affect a single parameter to an entity without using a container. This is essentially syntactic sugar: a feature that makes the language easier to write. It is most used for special entities like quantifiers or modifications. This is why, once used, the parent entity takes a polymorphic type, meaning that type inference will not issue errors for any usage of them. Details of the way the operators are reduced is exposed in algorithm 3.

Algorithm 3 Operator meta-rule

```
1: function OPERATOR(Entity[] statement)
2:   reduced = statement
3:   while LENGTH(reduced) > 3 do
4:     for i from 0 to LENGTH(reduced) - 1 do
5:       if ν(reduced[i]) not null and
6:         ν(reduced[i+1]) not null and
7:           (ν(reduced[i]) is a new symbol or
8:             reduced[i] has been parameterized before) and
9:             COHERENTOPERATOR(reduced, i) then
10:               op = INFEROPERATOR(reduced[i], reduced[i+1])
11:               reduced.REMOVE(reduced[i], reduced[i+1])
12:               reduced.INSERT(op, i)                                > Replace parameterized entity
13:               break
14:             if LENGTH(statement) stayed the same then return Success
15:   return Failure
```

TODO: More explanation

From the call of inferOperator, comes new rules explicated in listing 3.4. The call also adds the operator entity to an inferred token <OP>.

```
1 <meta_entity> ::= <?OP> <ID>
2 <meta_property> ::= <?OP> <ID>
```

Listing 3.4: Rules added to the current grammar for handling operators

Example 14. With the input !x = 0; we parse <SYMBOL> <ENTITY> <EQUAL> <LITERAL> <EOS>. This cannot be a container since the new symbol is at the beginning without any mirroring possible. It cannot be a parameter since no container is present. But this will conclude with the operator meta-rule as the new symbol precedes an entity. This becomes <meta_entity> <EQUAL> <LITERAL> <EOS> and becomes a valid statement.

If all meta-rules fail, then the parsing fails and returns an error to the user like in classical occurrences.

3.4.4 Contextual Interpretation

While parsing another important part of the processing is done after the success of a grammar rule. The grammar in SELF is valued, meaning that each rule has to return an entity. A set of functions are used to then populate the knowledge description system

with the right entities or retrieve an existing one that corresponds to what is being parsed.

When parsing, the rules `<entity>` and `<property>` will trigger the creation or retrieval of an entity. This mechanism will use the name of the entity to retrieve an entity with the same name in a given scope. If no such entity exists it is created and added to the current scope.

3.4.4.1 Naming and Scope

When parsing an entity by name, the system will first request for an existing entity with the same name. If such an entity is retrieved, it is returned instead of creating a new one. The validity of a name is limited by the notion of scope.

A scope is the reach of an entity's direct influence. It affects the naming relation by removing variable names. Scopes are delimited by containers and statements. This local context is useful when wanting to restrict the scope of the declaration of an entity. The main goal of such restriction is to allow for a similar mechanism as the RDF namespaces. This also makes the use of variables possible, akin to RDF blank nodes.

The scope of an entity has three special values :

- Variable: This scope restricts the scope of the entity to only the other entities in its scope.
- Local: This scope is temporarily bound to a given entity during the parsing. This scope is limited to the statement being interpreted.
- Global: This scope means that the name has no scope limitation.

The scope of an entity also contains all its parent entities, meaning all containers or statement the entity is part of. This is used when choosing between the special values of the scope. The process is detailed in algorithm 4.

The process happens for each entity created or requested by the parser. If a given entity is part of any other entity, the enclosing entity is added to its scope. When an entity is enclosed in any entity while already being a parameter of another entity, it becomes a variable since it is referenced twice in the same statement.

3.4.4.2 Instanciation identification

When a parameterized entity is parsed, another process starts to identify if a compatible instance already exists. From theorem 1, it is impossible for two entities to share the same identifier. This makes mandatory to avoid creating an entity that is equal to an existing one. Given the order of which parsing is done, it is not always possible to determine the parameter of an entity before its creation. In that case a later examination will merge the new entity onto the older one and discard the new identifier.

Algorithm 4 Determination of the scope of an entity

```
1: function INFERSCOPE(Entity  $e$ )
2:   Entity[] reach = []
3:   if :  $(e) = S$  then
4:     for all  $i \in \chi(e)$  do reach.APPEND(INFERVARIABLE( $i$ ))     $\triangleright$  Adding scopes nested in statement  $e$ 
5:   for all  $i \in \mu^*(e)$  do reach.APPEND(INFERVARIABLE( $i$ ))     $\triangleright$  Adding scopes nested in container  $e$ 
6:   if  $\exists \rho(e)$  then
7:     Entity[] param = INFERSCOPE( $\rho(e)$ )
8:     for all  $i \in \text{param}$  do param.REMOVE(INFERSCOPE( $i$ ))     $\triangleright$  Remove duplicate scopes from
parameters
9:     for all  $i \in \text{param}$  do reach.APPEND(INFERVARIABLE( $i$ ))     $\triangleright$  Adding scopes from parameters of  $e$ 
10:    SCOPE( $e$ )  $\leftarrow$  reach
11:    if GLOBAL  $\notin$  SCOPE( $e$ ) then SCOPE( $e$ )  $\leftarrow$  SCOPE( $e$ )  $\cup$  {LOCAL}
return reach
12: function INFERVARIABLE(Entity  $e$ )
13:   Entity[] reach = []
14:   if LOCAL  $\in$  SCOPE( $e$ ) then
15:     for all  $i \in \text{SCOPE}(e)$  do
16:       if  $\exists e_p \in \cup : \rho(p) = i$  then     $\triangleright e$  is already a parameter of another entity  $e_p$ 
17:         SCOPE( $e$ )  $\leftarrow$  SCOPE( $e$ )  $\setminus$  {LOCAL}
18:         SCOPE( $e_p$ )  $\leftarrow$  SCOPE( $e_p$ )  $\cup$  SCOPE( $e$ )
19:         SCOPE( $e$ )  $\leftarrow$  SCOPE( $e$ )  $\cup$  {VARIABLE,  $p$ }
20:       reach.APPEND( $e$ )
21:   reach.APPEND(SCOPE( $e$ ))
return reach
```

3.4.5 Structure as a Definition

The derivation feature on its own does not allow to define most of the native properties. For that, one needs a light inference mechanism. This mechanism is part of the default inference engine. This engine only works on the principle of structure as a definition. Since all names must be neutral from any language, that engine cannot rely on classical mechanisms like configuration files with keys and values or predefined keywords.

To use SELF correctly, one must be familiar with the native properties and their structure or implement their own inference engine to override the default one.

3.4.5.1 Quantifiers

In SELF quantifiers differ from their mathematical counterparts. The quantifiers are special entities that are meant to be of a generic type that matches any entities including quantifiers. There are infinitely many quantifiers in SELF but they are all derived from a special one called the *solution quantifier*. We mentioned it briefly during the definition of the grammar g_0 . It is the language equivalent of μ^* and is used to extract and evaluate reified knowledge (see section 3.4.1.2).

Example 15. The statement `bob is <SOLVE>(x)` will give either a default container filled with every value that the variable `x` can take or if the value is unique, it will take that value. If there is no value it will default to `<NULL>`, the exclusion quantifier.

How are other quantifiers defined? We use a definition akin to Lindström quantifiers (1966) which is a generalization of counting quantifiers (Gradel et al. 1997). Meaning that a quantifier is defined as a constrained range over the quantified variable. We suppose five quantifiers as existing in SELF as native entities.

- The **solution quantifier** `<SOLVE>` noted \S in classical mathematics, turns the expression into the possible value range of its variable. It is like replacing it by the natural expression “those x that”.
- The **universal quantifier** `<ALL>` behaves like \forall and forces the expression to take every possible value of its variable.
- The **existential quantifier** `<SOME>` behaves like \exists and forces the expression to match *at least one* arbitrary value for its variable.
- The **uniqueness quantifier** `<ONE>` behaves like $\exists!$ and forces the expression to match *exactly one* arbitrary value for its variable.
- The **exclusion quantifier** `<NULL>` behaves like $\neg\exists$ and forces the expression to not match the value of its variable.

The last four quantifiers are inspired from Aristotle's square of opposition (D'Alfonso 2011).

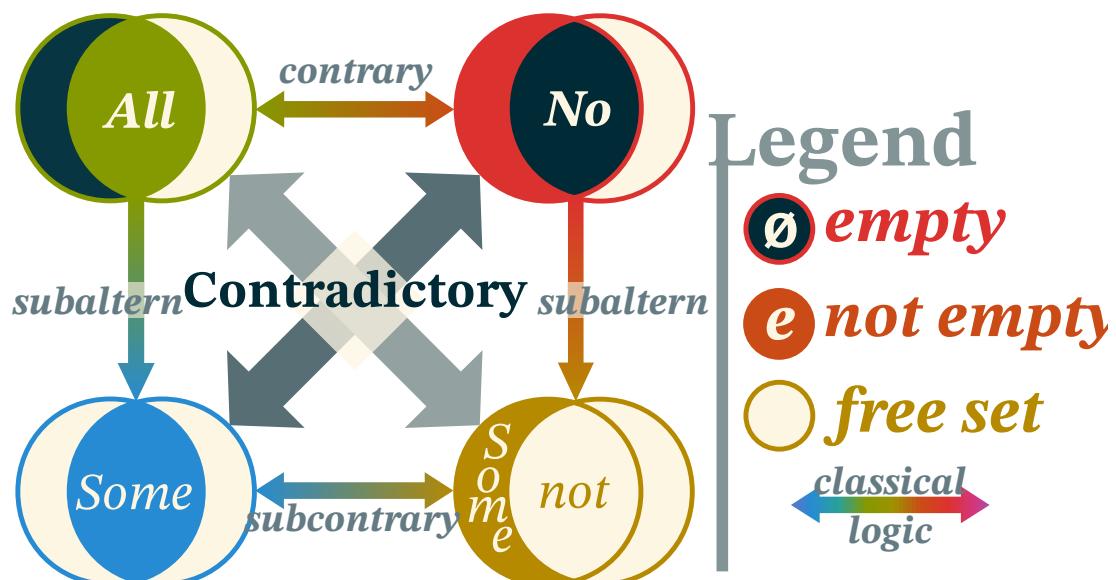


Figure 3.4: Aristotle's square of opposition

In SELF, quantifiers are not always followed by a quantified variable and can be used as a value. In that case the variable is simply anonymous. We use the exclusion quantifier as a value to indicate that there is no value, sort of like `null` or `nil` in programming languages.

Example 16. If we want to express the fact that a glass of water is not empty we can write either `glass contains ~(~)`; or `glass ~(contains) ~` with `<NULL> = ~`. This shows that `<NULL>` is used for negation and to indicate the absence of value.

This property is quite handy as it require only one symbol and allows for complex constructs that are difficult to explain using available paradigms.



In listing 3.5, we present an example file that is meant to define most of the useful native properties along with default quantifiers.

```

1 * =? ;
2 ?(x) = x; //Optional definition
3 ?~ = { };
4 ?_ ~(=) ~;
5 ?!_ = {_};
6
7 (*e, !T) : (e :: T); *T : (T :: Type);
8 *T : (Entity / T);
9
10 :: :: Property(Entity, Type);
11 (__) :: Statement;
12 (~, !, _, *) :: Quantifier;
13 ()::Group;
14 {}::Set;
15 []::List;
16 < >::Tuple;
17 Collection/(Set,List,Tuple);
18 0 :: Integer; 0.0::Float;
19 '\0'::Character; ""::String;
20 Literal/(Boolean, Integer, Float, Character, String);
21
22 (*e, !(s::String)) : (e named s);
23 (*e(p), !p) : (e param p);
24 *(s p o):(((s p o) subject s),((s p o) property p),((s p o) object o));

```

Listing 3.5: The default lang.w file.

At line 1, we give the first statement that defines the solution quantifier's symbol. The reason this first statement is shaped like this is that global statements are always evaluated to be a true statement. Since domains are sets of statements, this means that anything equaling the solution quantifier at this level will be evaluated as a domain. This is because the entity is a domain **by structure**. If it is a single entity then it becomes synonymous to the entire SELF domain and therefore contains everything. We can infer that it becomes the universal quantifier.

If it is a string literal, then it must be either a file path or URL or a valid SELF expression.

Example 17. Using the first statement, we can include external domains akin to the `import` directive in Java. Writing `"path/lang.w" = ? ;` as a first statement will make the process parse the file located at `path/lang.w` and insert it at this spot.

All statements up to line 5 are quantifiers definitions. On the left side we got the quantifier symbol used as a parameter to the solution quantifier using the operator notation. On the right we got the domain of the quantifier. The exclusive quantifier

has as a range the empty set. For the existential quantifier we have only a restriction of it not having an empty range. At last, the uniqueness quantifier got a set with only one element matching its variable (noting that anonymous variables doesn't match necessarily other anonymous variables in the same statement).

In listing 3.5 the type hierarchy can be illustrated by the figure 3.5. It consists of entities that are either parameterized or not and that has a value or not.

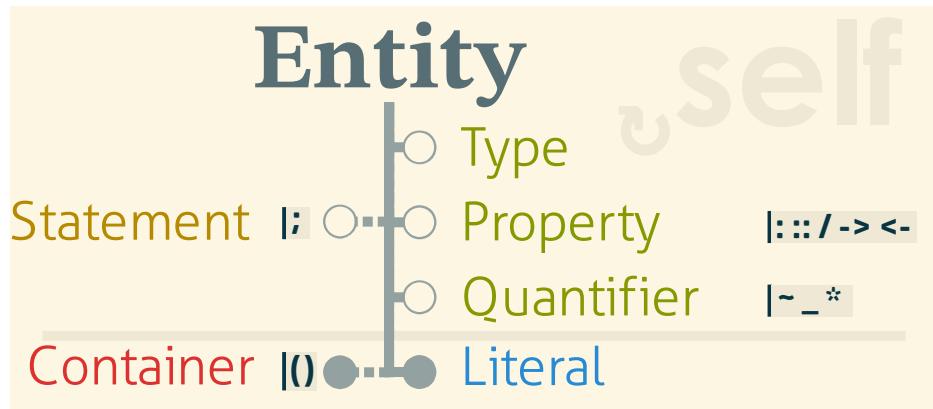


Figure 3.5: Hierarchy of types in SELF

TODO: Explain figure

3.4.5.2 Inferring Native Properties

All native properties can be inferred by structure using quantified statements. Here is the structural definition for each of them:

- $=$ (at line 1) is the equality relation given in the first statement.
- \subseteq (at line 8) is the first property to relate a particular type to all types. That type becomes the entity type.
- μ^* (at line 1) is the solution quantifier discussed above given in the first statement.
- μ is represented using containers.
- ν (at line 22) is the first property affecting a string literal uniquely to each entity.
- ρ (at line 23) is the first property to effect to all entities a possible parameter list.
- $:$ (at line 7) is the first property that matches every entity to a type.
- χ (at line 24) is the first property to match for all statements.

We limit the inference to one symbol to eliminate ambiguities and prevent accidental re-definition of native properties. This also improves performance as the inference is stopped after finding a first matching entity that can be used programmatically using a single constant.

3.4.6 Extended Inference Mechanisms

In this section we present the default inference engine. It is quite limited since it is meant to be universal and the goal of SELF is to provide a framework that can be used by specialists to define and code exactly what tools they need.

Inference engines need to create new knowledge but this knowledge shouldn't be simply merged with the explicit user provided domain. Since this knowledge is inferred, it is not exactly part of the domain but must remain consistent with it. This knowledge is stored in a special scope dedicated to each inference engine. This way, inference engines can use defeasible logic or have dynamic inference from any knowledge insertion in real time.

3.4.6.1 Type Inference

Type inference works on matching types in statements. The main mechanism consists in inferring the type of properties in a restrictive way. Properties have a parameterized type with the type of their subject and object. The goal is to make that type match the input subject and object.

For that we start by trying to match the types. If the types differ, the process tries to reduce the more general type against the lesser one (subsumption-wise). If they are incompatible, the inference uses some light defeasible logic to undo previous inferences. In that case the types are changed to the last common type in the subsumption tree.

However, this may not always be possible. Indeed, types can be explicitly specified as a safeguard against mistakes. If that's the case, an error is raised and the parsing or knowledge insertion is interrupted.

3.4.6.2 Instantiation

Another inference mechanism is instantiation. Since entities can be parameterized, they can also be defined against their parameters. When those parameters are variables, they allow entities to be instantiated later.

Since entities are immutable, updating their instance can be quite tricky. Indeed, parsing happens from left to right and therefore an entity is often created before all the instantiation information are available. Even harder are completion of definition in several separate statements. In all cases, a new entity is created and then the inference realize that it is either matching a previous definition and will need to be merged with the older entity or it is a new instance and needs all properties duplicated and instantiated.

This gives us two mechanisms to take into account: merging and instantiating.

Merging is pretty straightforward: the new entity is replaced with the old one in all of the knowledge graph. containers, parameterized entities, quantifiers and statements must be duplicated with the correct value and the original destroyed. This is a heavy

and complicated process but seemingly the only way to implement such a case with immutable entities.

Instanciating is similar to merging but even more complicated. It starts with computing a relation that maps each variable that needs replacing with their grounded value. Then it duplicates all knowledge about the parent entity while applying the replacement map.

3.5 Example

In the following section, a use case of the framework will be presented. First we have to explain a few notions.

3.5.1 Modality of Statements

In the field of logic there exists one special flavor of it called *modal logic*. It lays the emphasis upon the qualifications of statements, and especially the way they are interpreted. This is a very appropriate example for SELF. The modality of a statement acts like a modifier, it specifies a property regarding its plausibility, origin or validity.

Example 18. In the figure 3.6, we present a case of three persons gossiping, Alice, Becky and Carol. The presentation is inspired by the work of Schwarzentruber (2018). Here is a list of the statements in this example:

- Alice said to Becky that Carol should *probably* change her style from C_1 to C_2 .
- Becky said to Alice that she finds the Carol's style *usually* good.
- Alice told Carol that Becky told her that she should *sometimes* change her style to C_2 .

The following statement can be inferred:

- Carol *possibly* thinks that Becky thinks that the style C_2 is *often* good.

In the example, all modalities are *emphasized*. One can notice an interesting property of these statements in that they are about other statements. This kind of description is called *higher order knowledge*.

3.5.2 Higher order knowledge

SELF is based on the ability to easily process higher order knowledge. In that case the term *order* refers to the level of abstraction of a statement (Schwarzentruber 2018). For such usages, a hypergraph structures such as SELF is using is a clear advantage in terms of expressivity and ease of manipulation of those statements. This is due to the higher dimensionality of sheaves (and by extension hypergraphs) that makes meta-statement as simple to express as any other statement. This chain of abstractions using meta-statements is where the higher order knowledge is encoded.

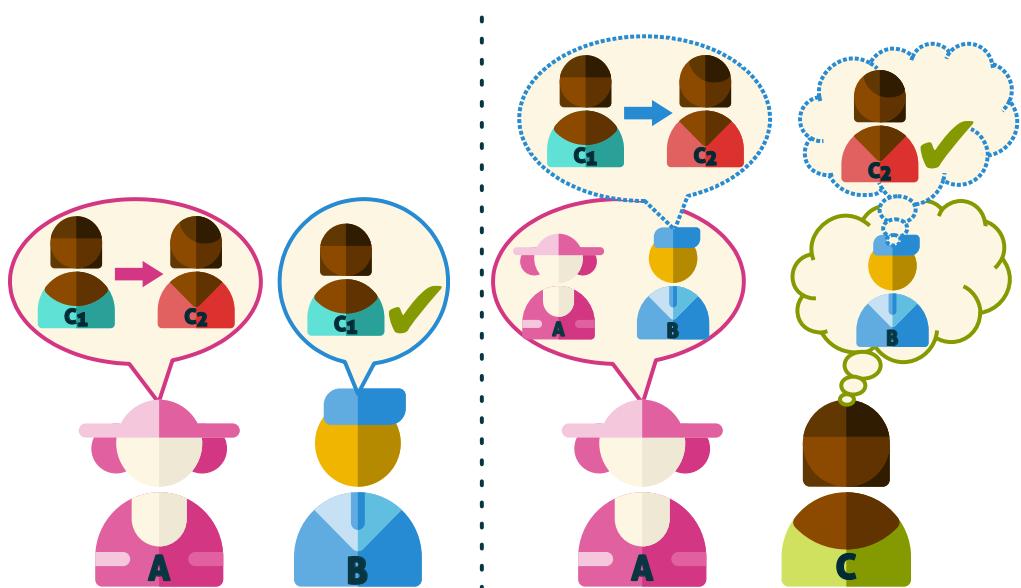


Figure 3.6: Example of modal logic propositions: Alice gossips about what Beatrice said about Claire

In the following listings, we present the previous example using RDF and SELF to describe knowledge of the gossip.

```
1 @prefix : <http://genn.io/self/gossip#> .
2 @prefix owl: <http://www.w3.org/2002/07/owl#> .
3 @prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
4 @prefix xml: <http://www.w3.org/XML/1998/namespace> .
5 @prefix xsd: <http://www.w3.org/2001/XMLSchema#> .
6 @prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
7 @base <http://genn.io/self/gossip> .

8
9 <http://genn.io/self/gossip> rdf:type owl:Ontology ;
10           owl:imports rdf: .

11
12 :modality rdf:type owl:AnnotationProperty ;
13           rdfs:range :Modality .
14
15 :told_a rdf:type owl:ObjectProperty .
16 :told_b rdf:type owl:ObjectProperty .
17 :told_c rdf:type owl:ObjectProperty .
18 :Modality rdf:type owl:Class .
19 :good rdf:type owl:NamedIndividual .
20 :is rdf:type owl:NamedIndividual ,
21           rdf:Property .
22 :probably rdf:type owl:NamedIndividual ,
23           :Modality .
24 :s1 rdf:type owl:NamedIndividual ,
25           rdf:Statement ;
26     rdf:object :c2 ;
27     rdf:predicate :worsethan ;
28     rdf:subject :c ;
29     :modality :probably .
30 :s2 rdf:type owl:NamedIndividual ;
31     rdf:object :good ;
32     rdf:predicate :is ;
33     rdf:subject :c ;
34     :modality :usually .
35 :s3 rdf:type owl:NamedIndividual ,
36           rdf:Statement ;
37     rdf:object :s4 ;
38     rdf:predicate :told_a ;
39     rdf:subject :b .
40 :s4 rdf:type owl:NamedIndividual ,
41           rdf:Statement ;
42     rdf:object :c2 ;
43     rdf:predicate :should ;
44     rdf:subject :c ;
45     :modality :sometimes .
46 :should rdf:type owl:NamedIndividual ,
47           rdf:Property .
48 :sometimes rdf:type owl:NamedIndividual ,
49           :Modality .
50 :told_a rdf:type owl:NamedIndividual ,
51           rdf:Property .
52 :usually rdf:type owl:NamedIndividual ,
53           :Modality .
54 :worsethan rdf:type owl:NamedIndividual ,
55           rdf:Property .
56 :a rdf:type owl:NamedIndividual ;
57   :told_b :s1 ;
```

```
58 :told_c :s3 .  
59 :b rdf:type owl:NamedIndividual ;  
60 :told_a :s2 .  
61 :c rdf:type owl:NamedIndividual .  
62 :c2 rdf:type owl:NamedIndividual .
```

```
1 "lang.s" = ? ;  
2 a told(b) probably(c worsethan ctwo);  
3 b told(a) usually(c is good);  
4 a told(c) (b told(a) sometimes(c should ctwo));
```

It is obvious that the SELF version is an order of magnitude more concise than RDF to express modal logic. The 4 lines of SELF are **equivalent** to the 62 lines of RDF. In the RDF version we use the reified statements `:s1`, `:s2`, `:s3` and `:s4` along with a `:modality` annotation to express high order knowledge and modalities. In SELF, everything is inferred by structure and one can start exploiting their database right away.

4 General Planning Framework

When designing intelligent systems, an important feature is the ability to make decisions and act accordingly. To act, one should plan ahead. This is why the field of automated planning is being actively researched in order to find efficient algorithms to find the best course of action in any given situation. The previous chapter allowed to lay the basis of knowledge representation. How knowledge about the planning domains are represented is a main factor to take into account in order to conceive most planning algorithm.

Automated planning really started being formally investigated after the creation of the Stanford Research Institute Problem Solver (STRIPS) by Fikes and Nilsson (1971). This is one of the most influential planners, not because of its algorithm but because of its input language. Any planning system needs a way to express the information related to the input problem. Any language done for this purpose is called an *action language*. STRIPS will be mostly remembered for its eponymous action language that is at the base of any modern derivatives.

All action languages are based on mainly two notions: *actions* and *states*. A state is a set of *fluents* that describe aspects of the world modeled by the domain. Each action has a logic formula over states that allows its correct execution. This requirement is called *precondition*. The mirror image of this notion is called possible *effects* which are logic formulas that are enforced on the current state after the action is executed. The domain is completed with a problem, most of the time specified in a separate file. The problem basically contains two states: the *initial* and *goal* states.

4.1 Illustration

To illustrate how automated planners works, we introduce a typical planning problem called **block world**.

Example 19. In this example, a robotic grabbing arm tries to stack blocks on a table in a specific order. The arm is only capable of handling one block at a time. We suppose that the table is large enough so that all the blocks can be put on it without any stacks. Figure 4.1 illustrates the setup of this domain.

The possible actions are `pickup`, `putdown`, `stack` and `unstack`. There are at least three fluents needed:

- one to state if a given block is `down` on the table,
- one to specify which block is `held` at any moment and
- one to describe which block is stacked `on` which block.

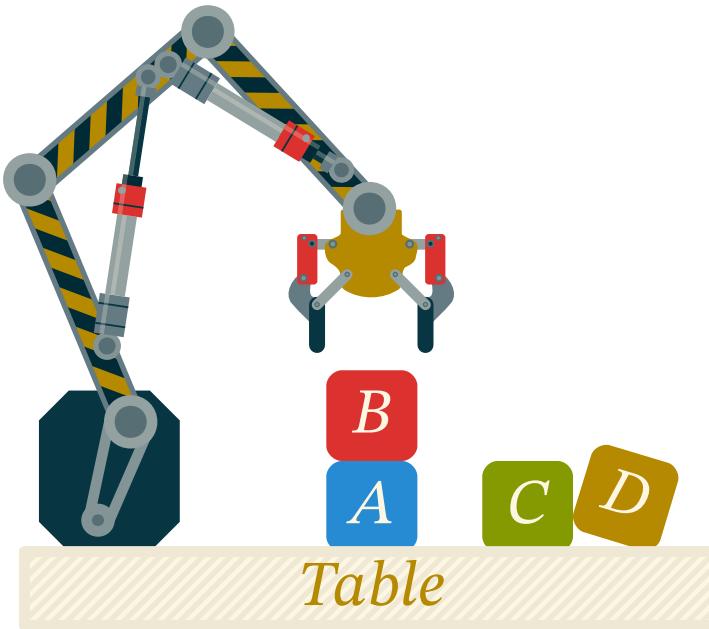


Figure 4.1: The block world domain setup.

We also need a special block to state when `noblock` is held or on top of another block. This block is a constant.

The knowledge we just described is called *planning domain*.

In that example, the initial state is described as stacks and a set of blocks directly on the table. The goal state is usually the specification of one or many stacks that must be present on the table. This part of the description is called *planning problem*.

In order to solve it we must find a valid sequence of actions called a *plan*. If this plan can be executed in the initial state and result in the goal state, it is called a *solution* of the planning problem. To be executed, each action must be done in a state satisfying its precondition and will alter that state according to its effects. A plan can be executed if all its actions can be executed in the sequence of the plan.

Example 20. For example, in the block world domain we can have an initial state with the *blockB* on top of *blockA* and the *blockC* being on the table. In figure 4.2, we give a plan solution to the problem consisting of having the stack $\langle \text{block}A, \text{block}B, \text{block}C \rangle$ from that initial state.

Every automated planner aims to find at least one such solution in any way shape or form in the least amount of time with the best plan quality. The quality of a plan is often measured by how hard it is to execute, whether by its execution time or by the resources needed to accomplish it. This metric is often called *cost* of a plan and is often simply the sum of the costs of its actions.

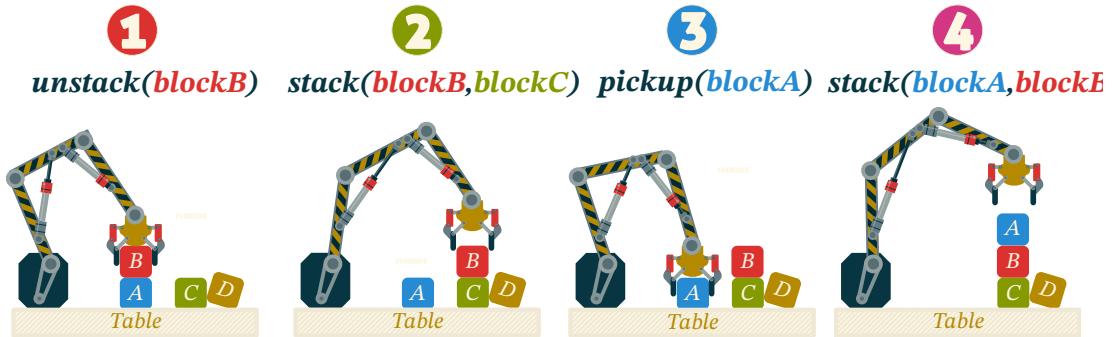


Figure 4.2: An example of a solution to a planning problem with a goal that requires three blocks stacked in alphabetical order.

Automated planning is very diverse. A lot of paradigms shift the definition of the domain, actions and even plan to widely varying extents. This is the reason why making a general planning formalism was deemed so hard or even impossible:

"It would be unreasonable to assume there is one single compact and correct syntax for specifying all useful planning problems." Sanner (2010)

Indeed, the block world example domain we give is mostly theoretical since there is infinitely more subtlety into this problem such as mechatronic engineering, balancing issues and partial ability to observe the environment and predict its evolution as well as failure in the execution. In our example, we didn't mention the misplaced *blockD* that could very well interfere with any execution in unpredictable ways. This is why so many planning paradigms exist and why they are all so diverse: they try to address an infinitely complex problem, one sub-problem at a time. In doing so we lose the general view of the problem and by simply stating that this is the only way to resolve it we close ourselves to other approaches that can become successful. Like once said:

"The easiest way to solve a problem is to deny it exists." Asimov (1973)

However, in the next section we aim to create such a general planning formalism. The main goal is to design the automated planning community with a general unifying framework it so badly needs.

4.2 Formalism

In this section, a general formalism of automated planning is proposed. The goal is to explain what is planning and how it works. First we must express the knowledge domain formalism, then we describe how problems are represented and lastly how a general planning algorithm can be envisioned.

4.2.1 Planning domain

In order to conceive a general formalism for planning domains, we base its definition on the formalism of SELF. This means that all parts of the domain must be a member of the universe of discourse \mathbb{U} .

4.2.1.1 Fluents

First, we need to define the smallest unit of knowledge in planning, the fluents.

Definition 30 (Fluent). A planning fluent is a predicate $f \in F$.

Fluents are signed. Negative fluents are noted $\neg f$ and behave as a logical complement. We do not use the closed world hypothesis: fluents are only satisfied when another compatible fluent is provided.

The name “fluent” comes from their fluctuating value. Indeed the truth value of a fluent is meant to vary with time and specifically by acting on it. In this formalism we represent fluents using either parameterized entities or using statements for binary fluents.

Example 21. In our example we have four predicates. They can form countless fluents like $held(no-block)$, $on(blockA, blockB)$ or $\neg down(blockA)$. Their when expressing a fluent we suppose its truth value is T and denote falsehood using the negation \neg .

4.2.1.2 States

When expressing states, we need a formalism to express sets of fluents as formulas.

Definition 31 (State). A state is a logical formula of fluents. Since all logical formulas can be reduced to a simple form using only \wedge , \vee , and \neg , we can represent states as *and/or trees*. This means that the leaves are fluents and the other nodes are states. We note states using small squares \square as it is often the symbol used in the representation of automates and grafsets.

Example 22. In the domain block world, we can express a couple of states as :

- $\square_1 = held(noblock) \wedge on(blockA, blockB) \wedge down(blockC)$
- $\square_2 = held(blockC) \wedge down(blockA) \wedge down(blockB)$

In such a case, both state \square_1 and \square_2 have their truth value being the conjunction of all their fluents. We can express a disjunction in the following way: $\square_3 = \square_1 \vee \square_2$. In that case \square_3 is the root of the and/or tree and all its direct children are or vertices. The states \square_1 and \square_2 have their children as *and vertices*. All the leaves are fluents.

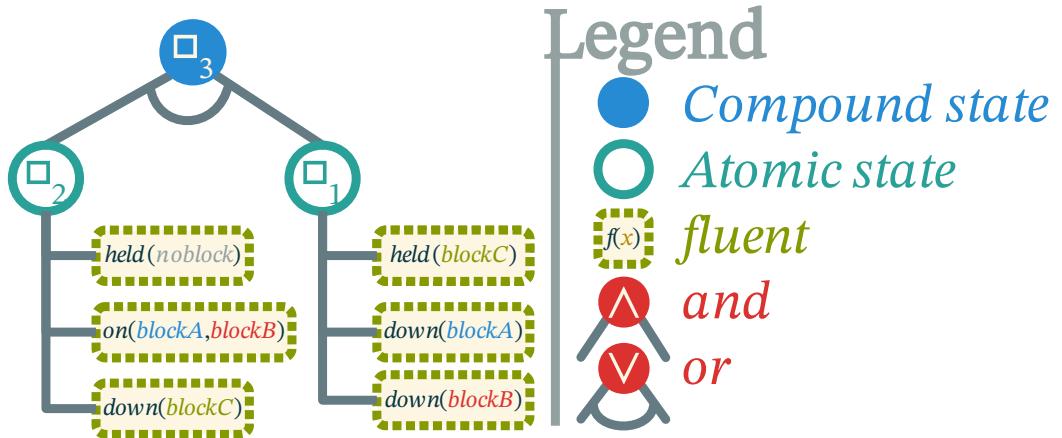


Figure 4.3: Example of a state encoded as an and/or tree.

4.2.1.3 Verification and binding constraints

When planning, there are two operations that are usually done on states: verify if a precondition fits a given state and then apply the effects of an action. In our model we consider preconditions and effects as states.

The verification is the operation $\square_{pre} \models \square$ that has either no value when the verification fails or a binding map for variables and fluents with their respective values.

The algorithm is a regular and/or tree exploration and evaluation applied on the state $\square = \square_{pre} \wedge \square$. During the valuation, if an inconsistency is found then the algorithm returns nothing. Otherwise, at each node of the tree, the algorithm will populate the binding map and verify if the truth value of the node holds under those constraints. All quantified variables are also registered in the binding map to enforce coherence in the root state. If the node is a state, the algorithm recursively applies until it reaches fluents. Once \square is valued as true, the binding map is returned.

Example 23. Using previously defined example states $\square_{1,2,3}$, and adding the following:

- $\square_4(x) = \{held(noblock), down(x)\}$ and
- $\square_5(y) = \{held(y), \neg down(y)\}$,

We can express a few examples of fluent verification:

- $held(noblock) \models held(x) = \{x = noblock\}$
- $\neg held(x) \models held(x) = \emptyset$

4.2.2 Application

Once the verification is done, the binding map is kept until the planner needs to apply the action to the state. The application of an effect state is noted $\square_{eff}(\square) = \square'$ and

is very similar to the verification. The algorithm will traverse the state \square and use the binding map to force the values inside it. The binding map is previously completed using \square_{eff} to enforce the application of its new value in the current state. This leads to changing the state \square progressively into \square' and the application algorithm will return this state.

4.2.2.1 Actions

Actions are the main mechanism behind automated planning, they describe what can be done and how it can be done.

Definition 32 (Action). An action is a parameterized tuple $a(args) = \langle pre, eff, \gamma, \mathbb{C}, d, \mathbb{P}, \Pi \rangle$ where:

- pre and eff are states that are respectively the **preconditions and the effects** of the action.
- γ is the state representing the **constraints**.
- \mathbb{C} is the intrinsic **cost** of the action.
- d is the intrinsic **duration** of the action.
- \mathbb{P} is the prior **probability** of the action succeeding.
- Π is a set of **methods** that decompose the action into smaller simpler ones.

Operators take many names in difference planning paradigm: actions, steps, tasks, etc. In our case we call operators, all fully lifted actions and actions are all the possible instances (including operators).

In order to be more generalist, we allow in the constraints description, any time constraints, equalities or inequalities, as well as probabilistic distributions. These constraints can also express derived predicates. It is even possible to place arbitrary constraints on order and selection of actions.

Actions are often represented as state operators that can be applied in a given state to alter it. The application of actions is done by using the action as a relation on the set of states $a : \square \rightarrow \square$ defined as follows:

$$a(\square) = \begin{cases} \emptyset, & \text{if } pre \models \square = \emptyset \\ eff(\square), & \text{using the binding map otherwise} \end{cases}$$

Example 24. A useful action we can define from previously defined states is the following:

$$pickup(x) = \langle \square_4(x), \square_5(x), (x : Block), 1.0\mathbb{C}, 3.5s, 75\%, \emptyset \rangle$$

That action can pick up a block x in 3.5 seconds using a cost of 1.0 with a prior success probability of 75%.

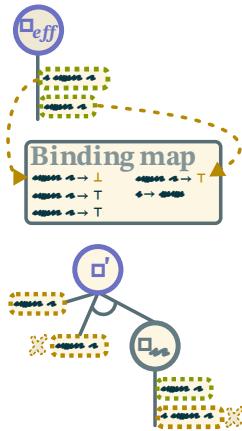


Figure 4.5: Example of application of a state with the updated binding constraints

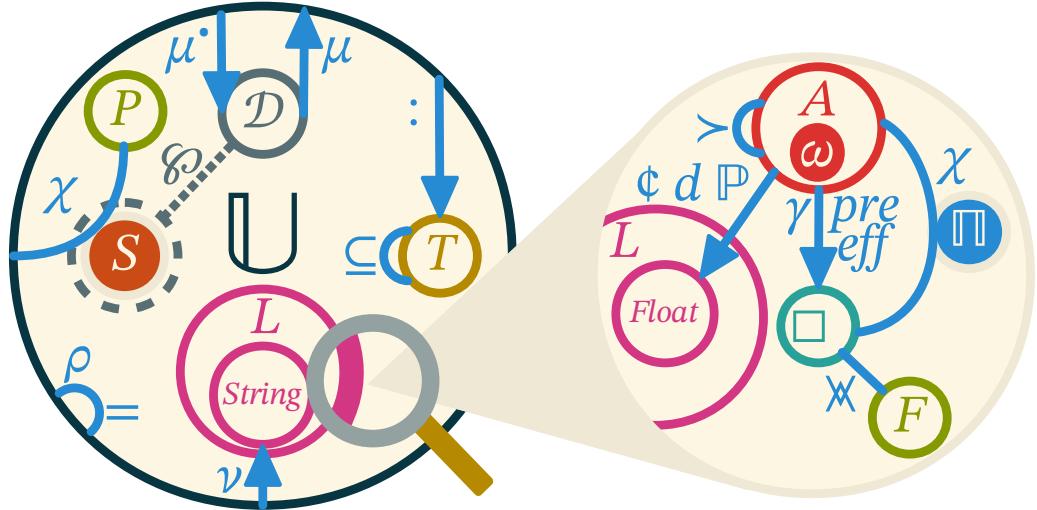


Figure 4.6: Venn diagram extended from the one from SELF to add all planning knowledge representation.

4.2.2.2 Domain

The planning domain specifies the allowed operators that can be used to plan and all the fluents they use as preconditions and effects.

Definition 33 (Domain). A planning domain \mathcal{D} is a set of **operators** which are fully lifted *actions*, along with all the relations and entities needed to describe their pre-conditions and effects.

Example 25. In the previous examples the domain was named block world. It consists in four actions: *pickup*, *putdown*, *stack* and *unstack*. Usually the domain is self contained, meaning that all fluents, types, constants and operators are contained in it.

4.2.3 Planning problem

The aim of an automated planner is to find a plan to satisfy the goal. This plan can be of multiple forms, and there can even be multiple plans that meet the demand of the problem.

4.2.3.1 Solution to Planning Problems

Definition 34 (Partial Plan / Method). A partially ordered plan is an *acyclic* directed graph $\Pi = (A_\Pi, E)$, with:

- A_{π} the set of **steps** of the plan as vertices. A step is an action belonging in the plan. A_{π} must contain an initial step a_{π}^0 and goal step a_{π}^* as convenience for certain planning paradigms.
- E the set of **causal links** of the plan as edges. We note $l = a_s \xrightarrow{\square} a_t$ the link between its source a_s and its target a_t caused by the set of fluents \square . If $\square = \emptyset$ then the link is used as an ordering constraint.

This definition can express any kind of plans, either temporal, fully or partially ordered or even hierarchical plans (using the methods of the actions \sqcap). It can even express diverse planning results.

The notation can be reminiscent of functional affectation and it is on purpose. Indeed, those links can be seen as relations that only affect their source to their target and the plan is a graph with its adjacency function being the combination of all links.

In our framework, *ordering constraints* are defined as the transitive cover of causal links over the set of steps. We note ordering constraints: $a_a > a_s$, with a_a being *anterior* to its successor a_s . Ordering constraints cannot form cycles, meaning that the steps must be different and that the successor cannot also be anterior to its anterior steps: $a_a \neq a_s \wedge a_s > a_a$. If we need to enforce order, we simply add a link without specifying a cause. The use of graphs and implicit order constraints help to simplify the model while maintaining its properties. Totally ordered plans are made by specifying links between all successive actions of the sequence.

Example 26. In the section 4.1, we described a classical fully ordered plan, illustrated in figure 4.2. A partially ordered plan has a tree-like structure except that it also meets in a "sink" vertex (goal step). We explicit this structure in figure 4.7.

4.2.3.2 Planning Problem

With this formalism, the problem is very simplified but still general.

Definition 35 (Problem). The planning problem is defined as the **root operator** ω which methods are potential solutions of the problem. Its preconditions and effects are respectively used as initial state and goal description.

As actions are very general, it is interesting to make the problem and domain space homogenous by using an action to describe any problem. Most of the specific notions of this framework are optional. Any planner using it will probably define what features it supports when compiling input domains and problems.

All notions explained so far are represented in the figure 4.6 adding to the SELF Venn diagram.

4.3 Planning search

A general planning algorithm can be described as a guided exploration of a search space. The detailed structure of the search space as well as search iterators is dependent on the planning paradigm used and therefore are parameters of the algorithm.

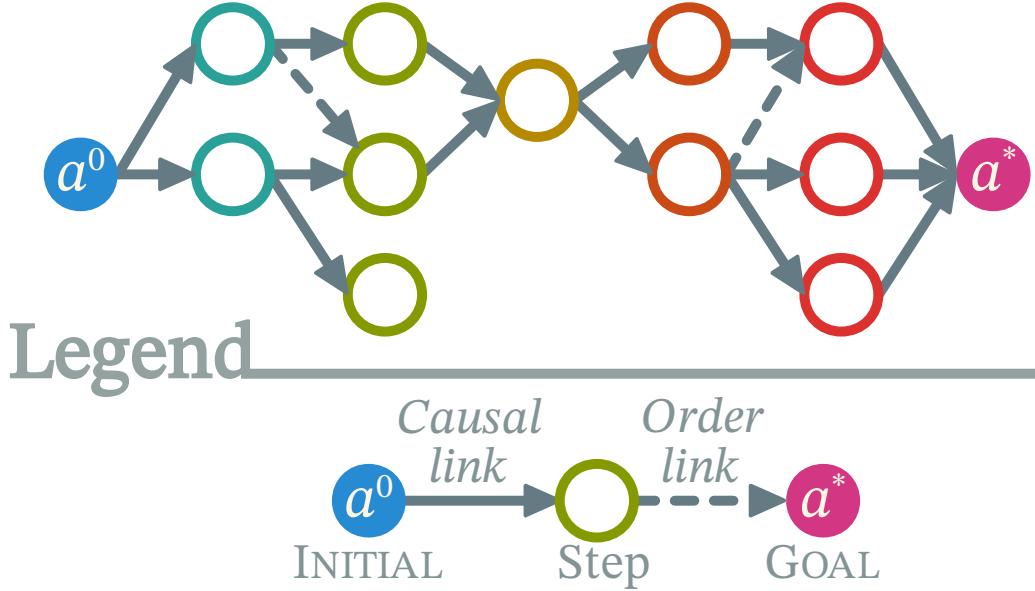


Figure 4.7: Structure of a partially ordered plan.

4.3.1 Search space

Definition 36 (Planner). A planning algorithm, often called planner, is an exploration of a search space \mathbb{S} partially ordered by an iterator $\chi_{\mathbb{S}}$ guided by a heuristic h . From any problem p every planner can derive two pieces of information immediately:

- the starting point $s_0 \in \mathbb{S}$ and
- the solution predicate $?_{s^*}$ that gives the validity of any potential solution in the search space.

Formally the problem can be viewed as a path-finding problem in the directed graph $g_{\mathbb{S}}$ formed by the vertex set \mathbb{S} and the adjacency function $\chi_{\mathbb{S}}$. The set of solutions is therefore expressed as:

$$\mathbb{S}^* = \{s^* : \langle s_0, s^* \rangle \in \chi_{\mathbb{S}}^+(s_0) \wedge ?_{s^*}\}$$

We note a provided heuristic $h(s)$. It gives off the shortest predicted distance to any point of the solution space. The exploration is guided by it by minimizing its value.

The search in automated planning can have a lot of requirements. Often time is limited and results may need to meet a certain set of specifications.

4.3.2 Solution constraints

As finding a plan is computationally expensive, it is sometimes better to try to find either a more generally applicable plan or a set of alternatives. This is especially important in the case of execution monitoring or human interactions as proposing several relevant solutions to pick from is a very interesting feature.

For this, one will prefer either probabilistic or diverse planning. The main problem is the additional parameters required and the change of behavior of the algorithm. In order to handle such formalism, the planning algorithm needs additional specifications.

We note $\gamma_{\mathbb{S}}$ the set of constraints on the nature of the solution. This contains notably the following optional elements:

- Δ the plan deviation metric to compare how much two plans are different.
- k is the number of expected different solutions. This simply makes the process return when either it found k solutions or when it determined that $k > |\mathbb{S}^*|$.

For probabilistic planning, all elements of the probability distributions used are typically included in the domain but the solution constraints parameter can also contain all required information from any planning paradigm, present or future as long as it is expressible in SELF.

4.3.3 Temporal constraints

Another aspect of planning lies in its timing. Indeed sometimes acting needs to be done before a deadline and planning are useful only during a finite timeframe. This is done even in optimal planning as researchers evaluating algorithms often need to set a timeout in order to be able to complete a study in a reasonable amount of time. Indeed, often in efficiency graphs, planning instances are stopped after a defined amount of time.

This time component is quite important as it often determines the planning paradigm used. It is expressed as two parameters:

- $t_{\mathbb{S}}$ the allotted time for the algorithm to find at least a fitting solution.
- t^* additional time for plan optimization.

This means that if the planner cannot find a fitting solution in time it will either return a timeout error or a partial or abstract solution that needs to be refined. Anytime planners will also use the extra time parameter to optimize the solution some more. If the amount of time is either unknown or unrestricted the parameters can be omitted and their value will be set to infinity.

4.4 General planner

A general planner $\Pi^*(g_S, s_0, ?_{s^*}, h, \gamma_S, \mathcal{D}, t_S, t^*)$ is an algorithm that can find solutions using any valid instance of the planning formalism.

For the planning algorithm itself, we simply use a parameterized instance of any search algorithms. In our case we chose the K* algorithm (Aljazzar and Leue 2011, alg. 1). This algorithm uses the classical algorithm A* to explore the graph while using Dijkstra on some sections to find the k shortest paths. The parameters are as follows: $K^*(g_S, s_0, ?_{s^*}, h)$. In this case, the solution predicate contains the solution and time constraints.

Of course this algorithm is merely an example of a general planner algorithm. The algorithm has been chosen to be general and its efficiency hasn't been tested.

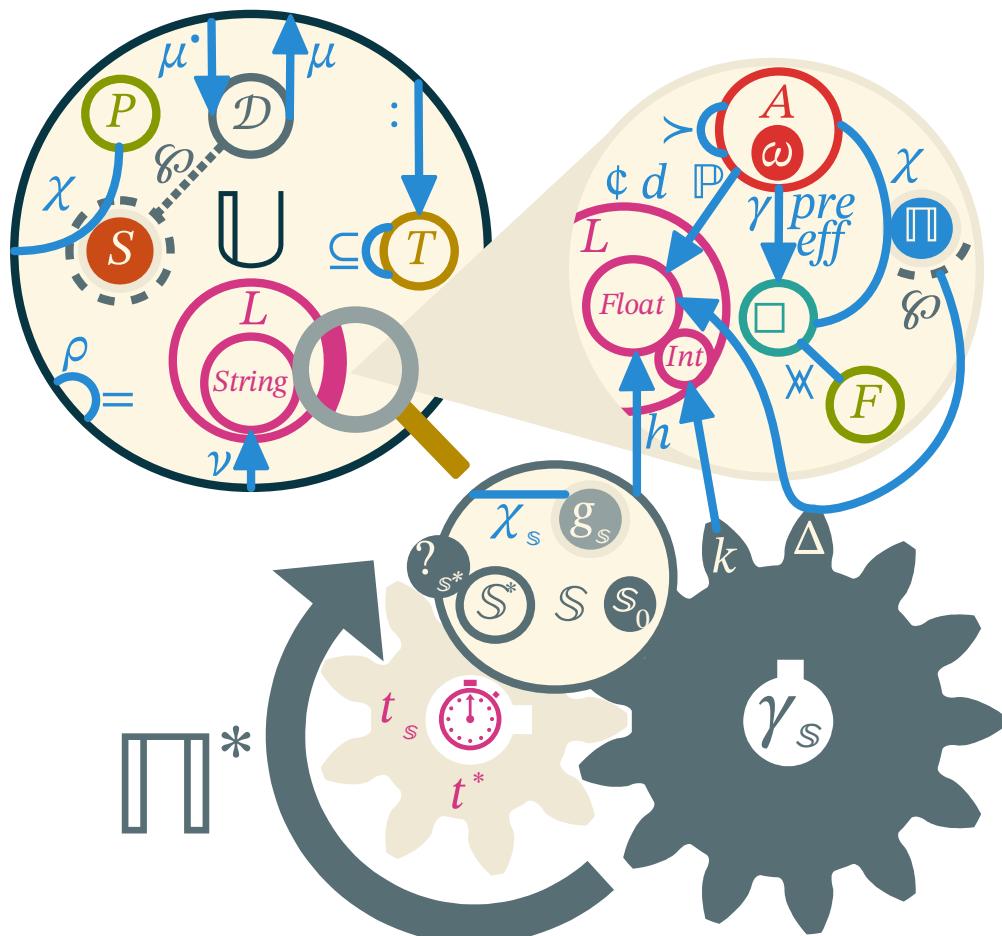


Figure 4.8: Venn diagram extended with general planning formalism.

4.5 Classical Formalism

One of the most comprehensive work on summarizing the automated planning domain was done by Ghallab *et al.* (2004). This book explains the different planning paradigm of its time and gives formal description of some of them. This work has been updated later (Ghallab *et al.* 2016) to reflect the changes occurring in the planning community.

4.5.1 State-transition planning

The most classical representation of automated planning is using the state transition approach: actions are operators on the set of states and a plan is a finite-state automaton. We can also see any planning problem as either a graph exploration problem or even a constraint satisfaction problem. In any way that problem is isomorph to its original formulation and most efficient algorithms use a derivative of A* exploration techniques on the state space.

The parameters for state space planning are trivial:

$$\Pi_{\square}^* = \Pi^*((\square, A), \text{pre}(\omega), \text{eff}(\omega))$$

This formulation takes advantage of several tools previously described. It uses the partial application of a function to omit the last parameters. It also defines the search graph using the set of all states \square as the vertices set and the set of all available actions A as the set of edges while considering actions as relations that can be applied to states to make the search progress toward an eventual solution. We also use the binary nature of states to use the effects of the root operator as the solution predicate.

Usually, we would set both t_S and t^* to an infinite amount as it is often the case for such planners. These parameters are left to the user of the planner.

State based planning usually suppose total knowledge of the state space and action behavior. No concurrence or time constraints are expressed and the state and action space must be finite as well as the resulting state graph. This process is also deterministic and doesn't allow uncertainty. The result of such planning is a totally ordered sequence of actions called a plan. The total order needs to be enforced even if it is unnecessary.

All those features are important in practice and lead to other planning paradigms that are more complex than classical state-based planning.

4.5.2 Plan space planning

Plan Space Planning (PSP) is a form of planning that uses plan space as its search space. It starts with an empty plan and tries to iteratively refine that plan into a solution.

The transformation into a general planner is more complicated than for state-based planning as the progress is made through refinements. We note the set of possible

refinement of a given plan $r = \pi \rightarrow \{\odot : \otimes(\pi)\}$. Each refinement is a new plan in which we fixed a *flaw* using one of the possible *resolvers* (see **LATER**).

$$\Pi_{\Pi}^* = \Pi^* ((\Pi, \{r(\pi) : \pi \in \Pi\}), (\{a^0, a^*\}, \{a^0 \rightarrow a^*\}), \otimes(s) = \emptyset)$$

with a^0 and a^* being the initial and goal steps of the plan corresponding s_0 such that $\text{eff}(a^0) = \text{pre}(\omega)$ and $\text{pre}(a^*) = \text{eff}(\omega)$. The iterator is all the possible resolutions of all flaws on any plan in the search space and the solution predicate is true when the plan has no more flaws.

Details about flaws, resolvers and the overall Partial Order Causal Links (POCL) algorithm will be presented **LATER**.

This approach can usually give a partial plan if we set t_s too low for the algorithm to complete. This plan is not a solution but can eventually be useful as an approximation for certain use cases (like intent recognition, see **LATER**).

4.5.3 Case based planning

Another plan oriented planning is called Case-Based Planning (CBP). This kind of planning relies on a library \mathcal{C} of already complete plans and try to find the most appropriate one to repair.

$$\Pi_{\mathcal{C}} = \Pi^* ((\mathcal{C}, \odot), (\pi : \pi \in \mathcal{C} \wedge \Delta(\pi, \omega) = \{\min : \mathcal{C} \times \{\omega\}\}, s(\text{pre}(\omega)) \neq \emptyset))$$

The planner selects a plan that fits the best efficiently with the initial and goal state of the problem. This plan is then repaired and validated iteratively. The problem with this approach is that it may be unable to find a valid plan or might need to populate and maintain a good plan library. For such case an auxiliary planner is used (preferably a diverse planner; that gives several solutions).

4.5.4 Probabilistic planning

Probabilistic planning tries to deal with uncertainty by generating a policy instead of a plan. The initial problem holds probability laws that govern the execution of any actions. It is sometimes accompanied with a reward function instead of a deterministic goal. We use the set of a pair of states with applicable actions as the search space: $\square + A = \{\langle \square, a \rangle : a \in A \wedge \square \in \square \wedge a(\square) \neq \emptyset\}$. We can also note the policy iteration $pol = \langle \square, a \rangle \rightarrow \langle a(\square), \{a' \in A \wedge a'(a(\square)) \neq \emptyset\} \rangle$

$$\Pi_p = \Pi^* ((\square + A, pol), \langle \text{pre}(\omega), \{a \in A \wedge a(\text{pre}(\omega)) \neq \emptyset\} \rangle, \forall s_1 \models \text{eff}(\omega))$$

At each iteration a state is chosen from the frontier. The frontier is updated with the application of a non-deterministically chosen pair of the last policy insertion. The search stops when all elements in the frontier are goal states.

4.5.5 Hierarchical planning

Hierarchical Task Networks (HTN) are a totally different kind of planning paradigm. Instead of a goal description, HTN uses a root task that needs to be decomposed. The task decomposition is an operation that replaces a task (action) by one of its methods Π .

$$\Pi_\omega = \Pi^* = ((\Pi, s \rightarrow \{\pi \in \Pi(\{a \in A_s \wedge \Pi(a) \neq \emptyset\})\}), \omega, \forall a \in A_s : \Pi(a) = \emptyset)$$

TODO: Explain more

4.6 Existing Languages and Frameworks

4.6.1 Classical

After STRIPS, one of the first languages to be introduced to express planning domains like ADL (Pednault 1989). That formalism adds negation and conjunctions into literals to STRIPS. It also drops the closed world hypothesis for an open world one: anything not stated in conditions (initials or action effects) is unknown.

The current standard was strongly inspired by Penberthy *et al.* (1992) and his UCPOP planner. Like STRIPS, UCPOP had a planning domain language that was probably the most expressive of its time. It differs from ADL by merging the add and delete lists in effects and to change both preconditions and effects of actions into logic formula instead of simple states.

The PDDL language was created for the first major automated planning competition hosted by AIPS in 1998 (Ghallab *et al.* 1998). It came along with syntax and solution checker written in Lisp. It was introduced as a way to standardize the notation of planning domains and problems so that libraries of standard problems can be used for benchmarks. The main goal of the language was to be able to express most of the planning problems of the time.

With time, the planning competitions became known under the name of International Planning Competitions (IPC) regularly hosted by the ICAPS conference. With each installment, the language evolved to address issues encountered the previous years. The current version of PDDL is 3.1 (Kovacs 2011). Its syntax goes similarly as described in listing 4.1.

```
1 (define (domain <domain-name>
2   (:requirements :<requirement-name>)
3   (:types <type-name>)
4   (:constants <constant-name> - <constant-type>)
5   (:predicates (<predicate-name> ?<var> - <var-type>))
6   (:functions (<function-name> ?<var> - <var-type>) - <function-type>)
7
8   (:action <action-name>
9     :parameters (?<var> - <var-type>))
```

```

10      :precondition (and (= (<function-name> ?<var>) <value>)
11          (<predicate-name> ?<var>))
12      :effect
13      (and (not (<predicate-name> ?<var>))
14          (assign (<function-name> ?<var>) ?<var>)))

```

Listing 4.1: Simplified explanation of the syntax of PDDL.

PDDL uses the functional notation style of LISP. It defines usually two files: one for the domain and one for the problem instance. The domain describes constants, fluents and all possible actions. The problem lays the initial and goal states description.

Example 27. For example, consider the classic block world domain expressed in listing 4.2. It uses a predicate to express whether a block is on the table because several blocks can be on the table at once. However it uses a 0-ary function to describe the one block allowed to be held at a time. The description of the stack of blocks is done with an unary function to give the block that is on top of another one. To be able to express the absence of blocks it uses a constant named `no-block`. All the actions described are pretty straightforward: `stack` and `unstack` make sure it is possible to add or remove a block before doing it and `pick-up` and `put-down` manages the handling operations.

```

1 (define (domain BLOCKS-object-fluents)
2   (:requirements :typing :equality :object-fluents)
3   (:types block)
4   (:constants no-block - block)
5   (:predicates (on-table ?x - block))
6   (:functions (in-hand) - block
7   (on-block ?x - block) ;;what is in top of block ?x
8
9   (:action pick-up
10    :parameters (?x - block)
11    :precondition (and (= (on-block ?x) no-block) (on-table ?x) (= (in-hand) no-block))
12    :effect
13    (and (not (on-table ?x))
14        (assign (in-hand) ?x)))
15
16   (:action put-down
17    :parameters (?x - block)
18    :precondition (= (in-hand) ?x)
19    :effect
20    (and (assign (in-hand) no-block)
21        (on-table ?x)))
22
23   (:action stack
24    :parameters (?x - block ?y - block)
25    :precondition (and (= (in-hand) ?x) (= (on-block ?y) no-block))
26    :effect
27    (and (assign (in-hand) no-block)
28        (assign (on-block ?y) ?x)))
29
30   (:action unstack
31    :parameters (?x - block ?y - block)
32    :precondition (and (= (on-block ?y) ?x) (= (on-block ?x) no-block)
33    (= (in-hand) no-block))
34    :effect
35    (and (assign (in-hand) ?x)

```

```
35      (assign (on-block ?y) no-block)))
```

Listing 4.2: Classical PDDL 3.0 definition of the domain Block world

However, PDDL is far from a universal standard. Some efforts have been made to try and standardize the domain of automated planning in the form of optional requirements. The latest of the PDDL standard is the version 3.1 (Kovacs 2011). It has 18 atomic requirements as represented in figure 4.9. Most requirements are parts of PDDL that either increase the complexity of planning significantly or that require extra implementation effort to meet.

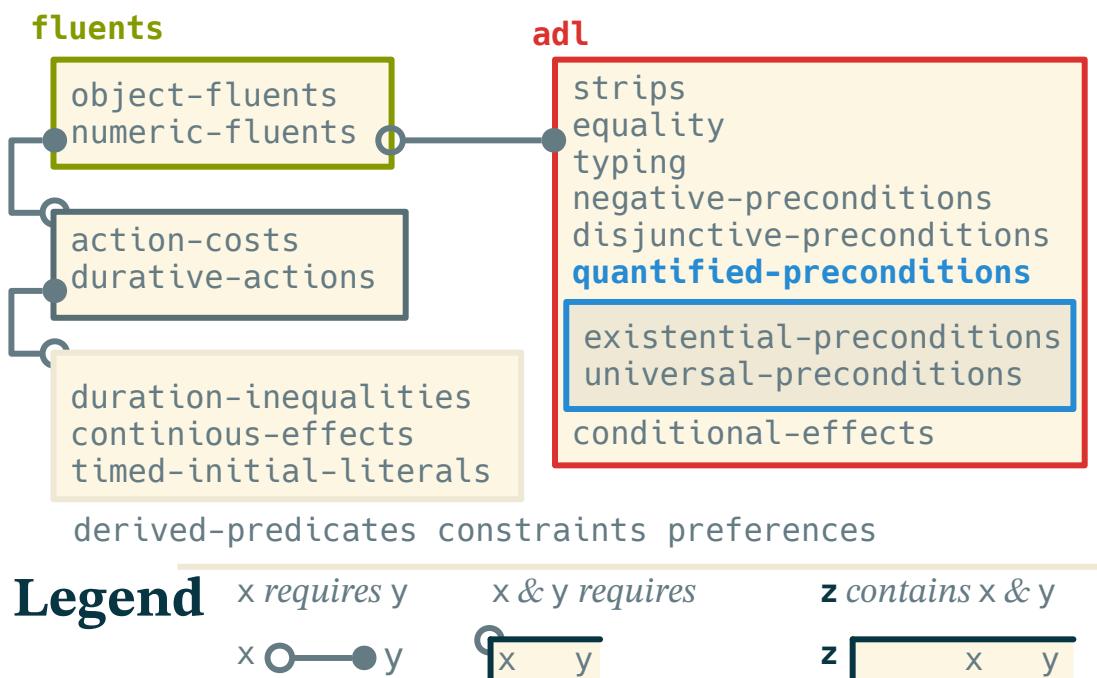


Figure 4.9: Dependencies and grouping of PDDL requirements.

Even with that flexibility, PDDL is unable to cover all of automated planning paradigms. This caused most subdomains of automated planning to be left in a state similar to before PDDL: a collection of languages and derivatives that aren't interoperable. The reason for this is the fact that PDDL isn't expressive enough to encode more than a limited variation in action and fluent description.

Another problem is that PDDL isn't made to be used by planners to help with their planning process. Most planners will totally separate the compilation of PDDL before doing any planning, so much so that most planners of the latest IPC used a framework that translates PDDL into a useful form before planning, adding computation time to the planning process. The list of participating planners and their use of language is presented in table ??.

Table 4.1: Planners participating in the Classic track of the 2018 International Planning Competition (IPC). The table states whether the planner used a translation and a pre-processing system to handle PDDL. Most of the planners are based on FastDownward directly. {#tbl:ipc}

Name	Trans	Pre	Lang	Base	Rank
Delfi	Yes	Yes	C++	FD	1
Complementary	Yes	Yes	C++	FD	2
Planning-PDBs	Yes	Yes	C++	FD	3
Scorpion	Yes	Yes	C++	FD	4
FDMS	Yes	Yes	C++	FD	5
DecStar	Yes	Yes	C++	LAMA	6
Metis	Yes	Yes	C++	FD	7
MSP	Yes	Yes	Lisp	FD	8
Symple	Yes	Yes	C++	FD	9
Ma-plan	No	Yes	C	None	10

The domain is so diverse that attempts to unify it haven't succeeded so far. The main reason behind this is that some paradigms are vastly different from the classical planning description. Sometimes just adding a seemingly small feature like probabilities or plan reuse can make for a totally different planning problem. In the next section we describe planning paradigms and how they differ from classical planning along with their associated languages.

4.6.2 Temporality oriented

When planning, time can become a sensitive constraint. Some critical tasks may require to be completed within a certain time. Actions with duration are already a feature of PDDL 3.1. However, PDDL might not provide support for external events (i.e. events occurring independent from the agent). To do this one must use another language.

4.6.2.1 PDDL+

PDDL+ is an extension of PDDL 2.1 that handles process and events (Fox and Long 2002). It can be viewed as similar to PDDL 3.1 continuous effects but it differs on the expressivity. A process can have an effect on fluents at any time. They can happen either from the agent's own doing or being purely environmental. It might be possible in certain cases to model this using the durative actions, continuous effects and timed initial literals of PDDL 3.1.

In listing 4.3, we reproduce an example from Fox and Long (2002). It shows the syntax of durative actions in PDDL+. The timed preconditions are also available in PDDL 3.1, but the `increase` and `decrease` rate of fluents is an exclusive feature of PDDL+.

```

1 (:durative-action downlink
2   :parameters (?r - recorder ?g - groundStation)
3   :duration (> ?duration 0)
4   :condition (and (at start (inView ?g))
5                 (over all (inView ?g))
6                 (over all (> (data ?r) 0)))
7   :effect (and (increase (downlinked)
8                  (* #t (transmissionRate ?g)))
9                  (decrease (data ?r)
10                     (* #t (transmissionRate ?g))))))

```

Listing 4.3: Example of PDDL+ durative action from Fox's paper.

The main issue with durative actions is that time becomes a continuous resource that may change the values of fluents. The search for a plan in that context has a higher complexity than regular planning.

4.6.3 Probabilistic

Sometimes, acting can become unpredictable. An action can fail for many reasons, from logical errors down to physical constraints. This call for a way to plan using probabilities with the ability to recover from any predicted failures. PDDL doesn't support using probabilities. That is why all IPC's tracks dealing with it always used another language than PDDL.

4.6.3.1 PPDDL

PPDDL is such a language. It was used during the 4th and 5th IPC for its probabilistic track (Younes and Littman 2004). It allows for probabilistic effects as demonstrated in listing 4.4. The planner must take into account the probability when choosing an action. The plan must be the most likely to succeed. But even with the best plan, failure can occur. This is why probabilistic planning often gives policies instead of a plan. A policy dictates the best choice in any given state, failure or not. While this allows for much more resilient execution, computation of policies are exponentially harder than classical planning. Indeed the planner needs to take into account every outcome of every action in the plan and react accordingly.

```

1 (define (domain bomb-and-toilet)
2   (:requirements :conditional-effects :probabilistic-effects)
3   (:predicates (bomb-in-package ?pkg) (toilet-clogged)
4                (bomb-defused)))
5   (:action dunk-package
6    :parameters (?pkg)
7    :effect (and (when (bomb-in-package ?pkg)
8                  (bomb-defused))
9                  (probabilistic 0.05 (toilet-clogged))))))

```

Listing 4.4: Example of PPDDL use of probabilistic effects from Younes's paper.

4.6.3.2 RDDL

Another language used by the 7th IPC's uncertainty track is RDDL (Sanner 2010). This language has been chosen because of its ability to express problems that are hard to encode in PDDL or PPDDL. Indeed, RDDL is capable of expressing Partially Observable Markovian Decision Process (POMDP) and Dynamic Bayesian Networks (DBN) in planning domains. This along with complex probability laws allows for easy implementation of most probabilistic planning problems. Its syntax differs greatly from PDDL, and seems closer to Scala or C++. An example is provided in listing 4.5 from Sanner (2010). In it, we can see that actions in RDDL don't need preconditions or effects. In that case the reward is the closest information to the classical goal and the action is simply a parameter that will influence the probability distribution of the events that conditioned the reward.

```
1 //////////////////////////////////////////////////////////////////
2 // A simple propositional 2-slice DBN (variables are not parameterized).
3 //
4 // Author: Scott Sanner (ssanner [at] gmail.com)
5 //////////////////////////////////////////////////////////////////
6 domain prop_dbn {
7
8   requirements = { reward-deterministic };
9
10  pvariables {
11    p : { state-fluent, bool, default = false };
12    q : { state-fluent, bool, default = false };
13    r : { state-fluent, bool, default = false };
14    a : { action-fluent, bool, default = false };
15  };
16
17  cpfs {
18    // Some standard Bernoulli conditional probability tables
19    p' = if (p ^ r) then Bernoulli(.9) else Bernoulli(.3);
20
21    q' = if (q ^ r) then Bernoulli(.9)
22      else if (a) then Bernoulli(.3) else Bernoulli(.8);
23
24    // KronDelta is like a DiracDelta, but for discrete data (boolean or
25    // int)
26    r' = if (~q) then KronDelta(r) else KronDelta(r <= q);
27  };
28
29  // A boolean functions as a 0/1 integer when a numerical value is needed
30  reward = p + q - r; // a boolean functions as a 0/1 integer when a
31  // numerical value is needed
32}
33
34 instance inst_dbn {
35
36  domain = prop_dbn;
37  init-state {
38    p = true; // could also just say 'p' by itself
39    q = false; // default so unnecessary, could also say '~q' by itself
40    r; // same as r = true
41  };
42
43  max-nondef-actions = 1;
```

```

42 horizon = 20;
43 discount = 0.9;
44 }

```

Listing 4.5: Example of RDDL syntax by Sanner.

4.6.4 Multi-agent

Planning can also be a collective effort. In some cases, a system must account for other agents trying to either cooperate or compete in achieving similar goals. The problem that arise is coordination. How to make a plan meant to be executed with several agents concurrently ? Several multi-agent action languages have been proposed to answer that question.

4.6.4.1 MAPL

Another extension of PDDL 2.1, MAPL was introduced to handle synchronization of actions (Brenner 2003). This is done using modal operators over fluents. In that regard, MAPL is closer to the PDDL+ extension proposed earlier. It encodes durative actions that will later be integrated into the PDDL 3.0 standard. MAPL also introduce a synchronization mechanism using speech as a communication vector. This seems very specific as explicit communication isn't a requirement of collaborative work. Listing 4.6 is an example of the syntax of MAPL domains. PDDL 3.0 seems to share a similar syntax.

```

1 (:state-variables
2   (pos ?a - agent) - location
3   (connection ?p1 ?p2 - place) - road
4   (clear ?r - road) - boolean)
5 (:durative-action Move
6   :parameters (?a - agent ?dst - place)
7   :duration (:= ?duration (interval 2 4))
8   :condition
9   (at start (clear (connection (pos ?a) ?dst)))
10  :effect (and
11    (at start (:= (pos ?a) (connection (pos ?a) ?dst)))
12    (at end (:= (pos ?a) ?dst))))

```

Listing 4.6: Example of MAPL syntax by Brenner.

4.6.4.2 MA-PDDL

Another aspect of multi-agent planning is the ability to affect tasks and to manage interactions between agents efficiently. For this MA-PDDL seems more adapted than MAPL. It is an extension of PDDL 3.1, that makes easier to plan for a team of heterogeneous agents (Kovács 2012). In the example in listing 4.7, we can see how action can be affected to agents. While it makes the representation easier, it is possible to obtain similar effect by passing an agent object as parameters of an action in PDDL 3.1. More complex expressions are possible in MA-PDDL, like referencing the action of other

agents in the preconditions of actions or the ability to affect different goals to different agents. Later on, MA-PDDL was extended with probabilistic capabilities inspired by PPDDL (Kovács and Dobrowiecki 2013).

```

1 (define (domain ma-lift-table)
2   (:requirements :equality :negative-preconditions
3                 :existential-preconditions :typing :multi-agent)
4   (:types agent) (:constants table)
5   (:predicates (lifted (?x - object) (at ?a - agent ?o - object)))
6   (:action lift :agent ?a - agent :parameters ()
7    :precondition (and (not (lifted table)) (at ?a table)
8                      (exists (?b - agent)
9                           (and (not (= ?a ?b)) (at ?b table) (lift ?b))))
10   :effect (lifted table)))

```

Listing 4.7: Example of MA-PDDL syntax by Kovacs.

4.6.5 Hierarchical

Another approach to planning is using Hierarchical Tasks Networks (HTN) to resolve some planning problems. Instead of searching to satisfy a goal, HTNs try to find a decomposition to a root task that fits the initial state requirements and that generate an executable plan.

4.6.5.1 UMCP

One of the first planner to support HTN domains was UCMP by Erol *et al.* (1994). It uses Lisp like most of the early planning systems. Apparently PDDL was in part inspired by UCMP's syntax. Like for PDDL, the domain file describes action (called operators here) and their preconditions and effects (called postconditions). The syntax is exposed in listing 4.8. The interesting part of that language is the way decomposition is handled. Each task is expressed as a set of methods. Each method has an expansion expression that specifies how the plan should be constructed. It also has a pseudo precondition with modal operators on the temporality of the validity of the literals.

```

1 (constants a b c table) ; declare constant symbols
2 (predicates on clear) ; declare predicate symbols
3 (compound-tasks move) ; declare compound task symbols
4 (primitive-tasks unstack dostack restack) ; declare primitive task symbols
5 (variables x y z) ; declare variable symbols
6
7 (operator unstack(x y)
8   :pre ((clear x)(on x y))
9   :post ((~on x y)(on x table)(clear y)))
10 (operator dostack (x y)
11   :pre ((clear x)(on x table)(clear y))
12   :post ((~on x table)(on x y)(~clear y)))
13 (operator restack (x y z)
14   :pre ((clear x)(on x y)(clear z))
15   :post ((~on x y)(~clear z)(clear y)(on x z)))
16
17 (declare-method move(x y z)
18   :expansion ((n restack x y z)))

```

```

19           :formula (and (not (veq y table))
20                     (not (veq x table))
21                     (not (veq z table)))
22                     (before (clear x) n)
23                     (before (clear z) n)
24                     (before (on x y) n)))
25
26 (declare-method move(x y z)
27   :expansion ((n dostack x z))
28   :formula (and (veq y table)
29                 (before (clear x) n)
30                 (before (on x y) n)))

```

Listing 4.8: Example of the syntax used by UCMP.

4.6.5.2 SHOP2

The next HTN planner is SHOP2 by Nau *et al.* (2003). It remains to this day, one of the reference implementation of an HTN planner. The SHOP2 formalism is quite similar to UCMP's: each method has a signature, a precondition formula and eventually a decomposition description. This decomposition is a set of methods like in UCMP. The methods can also be partially ordered allowing more expressive plans. An example of the syntax of a method is given in listing 4.9.

```

1 (:method
2   ; head
3   (transport-person ?p ?c2)
4   ; precondition
5   (and
6     (at ?p ?c1)
7     (aircraft ?a)
8     (at ?a ?c3)
9     (different ?c1 ?c3))
10  ; subtasks
11  (:ordered
12    (move-aircraft ?a ?c1)
13    (board ?p ?a ?c1)
14    (move-aircraft ?a ?c2)
15    (debark ?p ?a ?c2)))

```

Listing 4.9: Example of method in the SHOP2 language.

4.6.5.3 HDDL

A more recent example of HTN formalism comes from the PANDA framework by Bercher *et al.* (2014). This framework is considered the current standard of HTN planning and allows for great flexibility in domain description. PANDA takes previous formalism and generalize them into a new language exposed in listing 4.10. That language was called HDDL.

```

1 (define (domain transport)
2   (:requirements :typing :action-costs)
3   (:types

```

```

4      location target locatable - object
5      vehicle package - locatable
6      capacity-number - object
7  )
8  (:predicates
9      (road ?l1 ?l2 - location)
10     (at ?x - locatable ?v - location)
11     (in ?x - package ?v - vehicle)
12     (capacity ?v - vehicle ?s1 - capacity-number)
13     (capacity-predecessor ?s1 ?s2 - capacity-number)
14  )
15
16  (:task deliver :parameters (?p - package ?l - location))
17  (:task unload :parameters (?v - vehicle ?l - location ?p - package))
18
19  (:method m-deliver
20    :parameters (?p - package ?l1 ?l2 - location ?v - vehicle)
21    :task (deliver ?p ?l2)
22    :ordered-subtasks (and
23      (get-to ?v ?l1)
24      (load ?v ?l1 ?p)
25      (get-to ?v ?l2)
26      (unload ?v ?l2 ?p))
27  )
28  (:method m-unload
29    :parameters (?v - vehicle ?l - location ?p - package ?s1 ?s2 - capacity-number)
30    :task (unload ?v ?l ?p)
31    :subtasks (drop ?v ?l ?p ?s1 ?s2)
32  )
33
34  (:action drop
35    :parameters (?v - vehicle ?l - location ?p - package ?s1 ?s2 - capacity-number)
36    :precondition (and
37      (at ?v ?l)
38      (in ?p ?v)
39      (capacity-predecessor ?s1 ?s2)
40      (capacity ?v ?s1)
41    )
42    :effect (and
43      (not (in ?p ?v))
44      (at ?p ?l)
45      (capacity ?v ?s2)
46      (not (capacity ?v ?s1)))
47    )
48  )
49 )

```

Listing 4.10: Example of HDDL syntax as used in the PANDA framework.

4.6.5.4 HPDDL

A very recent language proposition was done by RAMOUL (2018). He proposes HPDDL with a simple syntax similar to the one of UCMP. In listing 4.11 we give an example of HPDDL method. Its expressive power seems similar to that of UCMP and SHOP.

```

1 (:method do_navigate
2   :parameters(?x - rover ?from ?to - waypoint)
3   :expansion((tag t1 (navigate ?x ?from ?mid))
4             (tag t2 (visit ?mid))
5             (tag t3 (do_navigate ?x ?mid ?to)))
6             (tag t4 (unvisited ?mid)))
7   :constraints((before (and (not (can_traverse ?x ?from ?to)) (not
8     (visited ?mid))
9           (can_traverse ?x ?from ?mid)) t1)))

```

Listing 4.11: Example of HPDDL syntax as described by Ramoul.

4.6.6 Ontological

Another idea is to merge automated planning and other artificial intelligence fields with knowledge representation and more specifically ontologies. Indeed, since the “semantic web” is already widespread for service description, why not make planning compatible with it to ease service composition ?

TODO: cite Eva

4.6.6.1 WebPDDL

This question finds its first answer in 2002 with WebPDDL. This language, explicated in listing 4.12, is meant to be compatible with RDF by using URI identifiers for domains (McDermott and Dou 2002). The syntax is inspired by PDDL, but axioms are added as constraints on the knowledge domain. Actions also have a return value and can have variables that aren’t dependant on their parameters. This allows for greater expressivity than regular PDDL, but can be partially emulated using PDDL 3.1 constraints and object fluents.

```

1 (define (domain www-agents)
2   (:extends (uri "http://www.yale.edu/domains/knowing")
3             (uri "http://www.yale.edu/domains/regression-planning")
4             (uri "http://www.yale.edu/domains/commerce"))
5   (:requirements :existential-preconditions :conditional-effects)
6   (:types Message - Obj Message-id - String)
7   (:functions (price-quote ?m - Money)
8               (query-in-stock ?pid - Product-id)
9               (reply-in-stock ?b - Boolean) - Message)
10  (:predicates (web-agent ?x - Agent)
11    (reply-pending a - Agent id - Message-id msg - Message)
12    (message-exchange ?interlocutor - Agent
13      ?sent ?received - Message
14      ?eff - Prop)
15    (expected-reply a - Agent sent expect-back - Message))
16  (:axiom
17    :vars (?agt - Agent ?msg-id - Message-id ?sent ?reply - Message)
18    :implies (normal-step-value (receive ?agt ?msg-id) ?reply)
19    :context (and (web-agent ?agt)
20                  (reply-pending ?agt ?msg-id ?sent)
21                  (expected-reply ?agt ?sent ?reply)))
22  (:action send

```

```

23      :parameters (?agt - Agent ?sent - Message)
24      :value (?sid - Message-id)
25      :precondition (web-agent ?agt)
26      :effect (reply-pending ?agt ?sid ?sent))
27 (:action receive
28      :parameters (?agt - Agent ?sid - Message-id)
29      :vars (?sent - Message ?eff - Prop)
30      :precondition (and (web-agent ?agt) (reply-pending ?agt ?sid ?sent))
31      :value (?received - Message)
32      :effect (when (message-exchange ?agt ?sent ?received ?eff) ?eff)))

```

Listing 4.12: Example of WebPDDL syntax by Mc Dermott.

4.6.6.2 OPT

This previous work was updated by McDermott (2005). The new version is called OPT and allows for some further expressivity. It can express hierarchical domains with links between actions and even advanced data structure. The syntax is mostly an update of WebPDDL. In listing 4.13, we can see that the URI was replaced by simpler names, the action notation was simplified to make the parameter and return value more natural. Axioms were replaced by facts with a different notation.

```

1 (define (domain www-agents)
2   (:extends knowing regression-planning commerce)
3   (:requirements :existential-preconditions :conditional-effects)
4   (:types Message - Obj Message-id - String )
5   (:type-fun (Key t) (Feature-type (keytype t)))
6   (:type-fun (Key-pair t) (Tup (Key t) t))
7   (:functions (price-quote ?m - Money)
8             (query-in-stock ?pid - Product-id)
9             (reply-in-stock ?b - Boolean) - Message)
10  (:predicates (web-agent ?x - Agent)
11    (reply-pending a - Agent id - Message-id msg - Message)
12    (message-exchange ?interlocutor - Agent
13      ?sent ?received - Message
14      ?eff - Prop)
15    (expected-reply a - Agent sent expect-back - Message))
16  (:facts
17    (freevars (?agt - Agent ?msg-id - Message-id
18              ?sent ?reply - Message)
19    (<- (and (web-agent ?agt)<!-- -->
20          (reply-pending ?agt ?msg-id ?sent)
21          (expected-reply ?agt ?sent ?reply))
22          (normal-value (receive ?agt ?msg-id) ?reply))))
23  (:action (send ?agt - Agent ?sent - Message) - (?sid - Message-id)
24    :precondition (web-agent ?agt)
25    :effect (reply-pending ?agt ?sid ?sent))
26  (:action (receive ?agt - Agent ?sid - Message-id) - (?received -
27    Message)
28    :vars (?sent - Message ?eff - Prop)
29    :precondition (and (web-agent ?agt)
30                      (reply-pending ?agt ?sid ?sent))
31    :effect (when (message-exchange ?agt ?sent ?received ?eff) ?eff)))

```

Listing 4.13: Example of the updated OPT syntax as described by Mc Dermott.

4.7 Color and general planning representation

From the general formalism of planning proposed earlier, it is possible to create an instantiation of the SELF language for expressing planning domains. This extension was the primary goal of creating SELF and uses almost all features of the language.

4.7.1 Framework

In order to describe this planning framework into SELF, we simply put all fields of the actions into properties. Entities are used as fluents, and the entire knowledge domain as constraints. We use parameterized types as specified **BEFORE**.

```
1 "lang.w" = ? ; //include default language file.
2 Fluent = Entity;
3 State = (Group(Fluent), Statement);
4 BooleanOperator = (&,|);
5 (pre,eff, constr)::Property(Action,State);
6 (costs,lasts,probability) ::Property(Action,Float);
7 Plan = Group(Statement);
8 -> ::Property(Action,Action); //Causal links
9 methods ::Property(Action,Plan);
```

Listing 4.14: Content of the file "planning.w"

The file presented in listing 4.14, gives the definition of the syntax of fluents and actions in SELF. The first line includes the default syntax file using the first statement syntax. The fluents are simply typed as entities. This allows them to be either parameterized entities or statements. States are either a set of fluents or a logical statement between states or fluents. When a state is represented as a set, it represents the conjunction of all fluents in the set.

Then at line 5, we define the preconditions, effects and constraint formalism. They are represented as simple properties between actions and states. This allows for the simple expression of commonly expressed formalism like the ones found in PDDL. Line 6 expresses the other attributes of actions like its cost, duration and prior probability of success.

Plans are needed to be represented in the files, especially for case based and hierarchical paradigms. They are expressed using statements for causal link representation. The property `->` is used in these statements and the causes are either given explicitly as parameters of the property or they can be inferred by the planner. We add a last property to express methods relative to their actions.

4.7.2 Example domain

Using the classical example domain used earlier, we can write the following file in listing 4.15.

```

1 "planning.w" = ? ; //include base terminology
2
3 (! on !, held(!), down(_)) :: Fluent;
4
5 pickUp(x) pre (~ on x, down(x), held(~));
6 pickUp(x) eff (~ (down(x)), held(~));
7
8 putDown(x) pre (held(x));
9 putDown(x) eff (held(~), down(x));
10
11 stack(x, y) pre (held(x), ~ on y);
12 stack(x, y) eff (held(~), x on y);
13
14 unstack(x, y) pre (held(~), x on y);
15 unstack(x, y) eff (held(x), ~ on y);

```

Listing 4.15: Blockworld written in SELF to work with Color

At line line 1, We need to include the file defined in listing 4.14. After that line 3 defines the allowed arrity of each relation/function used by fluents. This restricts eventually the cardinality between parameters (one to many, many to one, etc).

Line 5 encodes the action *pickup* defined earlier. It is interesting to note that instead of using a constant to denote the absence of block, we can use an anonymous exclusive quantifier to make sure no block is held. This is quite useful to make concise domains that stay expressive and intuitive.

4.7.3 Differences with PDDL

SELF+Color is more concise than PDDL. It will infere most types and declaration. Variables are also inferred if they are used more than once in a statement and also part of parameters.

While PDDL uses a fixed set of extensions to specify the capabilities of the domain, SELF uses inclusion of other files to allow for greater flexibility. In PDDL, everything must be declared while in SELF, type inference allows for usage without definition. It is interesting to note that the use of variables names *x* and *y* are arbitrary and can be changed for each statement and the domain will still be functionally the same. The line 3 in listing 4.2 is a specific feature of SELF that is absent in PDDL. It is possible to specify constraints on the cardinality of properties. This limits the number of different combinations of values that can be true at once. This is typically done in PDDL using several predicate or constraints.

Most of the differences can be summarized saying that 'SELF do it once, PDDL needs it twice'. This doesn't only mean that SELF is more compact but also that the expressivity allows for a drastic reduction of the search space if taken into account. Thiébaux *et al.* (2005) advocate for the recognition of the fact that expressivity isn't just a convenience but is crucial for some problems and that treating it like an obstacle by trying to compile it away only makes the problem worse. If a planner is agnostic to the domain and problem, it cannot take advantages of clues that the instantiation of an action or even its name can hold (Babli *et al.* 2015).

Whatever the time and work that an expert spends on a planning domain it will always be incomplete and fixed. SELF allows for dynamical extension and even addresses the use of reified actions as parameters. Such a framework can be useful in multi-agent systems where agents can communicate composite actions to instruct another agent. It can also be useful for macro-action learning that allows to improve hierarchical domains from repeating observations. It can also be used in online planning to repair a plan that failed. And at last this framework can be used for explanation or inference by making easy to map two similar domains together. (lots of **CITATION**).

Also another difference between SELF and PDDL is the underlying planning framework. We presented the one of SELF (listing 4.14) but PDDL seems to suppose a more classical state based formalism. For example the fluents are of two kinds depending if they are used as preconditions or effects. In the first case, the fluent is a formula that is evaluated like a predicate to know if the action can be executed in any given state. Effects are formula enforcing the values of existing fluent in the state. SELF just suppose that the new knowledge is enforcing and that the fluents are of the same kind since verification about the coherence of the actions are made prior to its application in planning.

5 Online and Flexible Planning Algorithms

In this chapter, we present planners and approaches to inverted planning and intent recognition. To do that we must first have an efficient online planning algorithm that can take into account observed plans or fluents and find the most likely plan to be pursued by an external agent. The planning process must be done in real time and take into account new observations to make new predictions. This requires the use of online planners. In such cases, the planning process works using distinct phases. In figure 5.1, we illustrate the components of the process. Only the planning part is meant to have real-time constraints on its execution the rest is usually a linear process and can be negligible in terms of execution times.



Figure 5.1: Planning phases for online planning

Classical planning can be used for such a work but lacks flexibility when needing to replan at high frequency. The planner must be either able to reuse previously found plans or be able to compute quickly plans that are good approximation of the intended goal. We could use probabilistic planning, especially Partially Observable Markovian Decision Process (POMDP) to directly encode the intent recognition problem but that approach has been explored in great detail already, including numerous Bayesian network approaches. Further discussions of inverted planning and intent recognition can be found **LATER**.

It was decided to explore more expressive and flexible approaches to use the semantics of the planning domain to attempt to guide the search to a more logical plan. This approach uses either repair heuristics or explanation to provide fast predictions of intended goals.

5.1 Existing Algorithms

In order to make a planner capable of repairing plans, the most fitting paradigm is PSP as described **BEFORE**. Using the plan space for search allows to modify the refinement process into repairing existing plans.

The second approach using explanations is hierarchical. The planner will use a HTN planning domain that contains composite actions (or tasks) that have several methods (as plans) to realize them.

First, PSP will be presented in more details regarding its classical formulation and definition.

5.1.1 Plan Space Planning

All PSP algorithms work in a similar way: their search space is the set of all plans and its iteration operation is plan refinement. This means that every PSP planner searches for *flaws* in the current plan and then computes a set of *resolvers* that potentially fix each of them. The algorithm usually starts with an empty plan only having the initial and goal step and recursively refine the plan until all flaws have been solved.

In general, PSP is faster than naive classical planning. However, with the advent of efficient state based heuristics used in Fast Forward (FF) (Hoffmann 2001) and LAMA (Richter and Westphal 2010), plan space planning has been left behind regarding raw performance. While PSP delays commitment and therefore can make very efficient choices that can be faster than classical planning, most formulation of PSP problems leads to significant increase in complexity (Tan and Gruninger 2014). The backtracking in PSP algorithms along with heavy data structures such as plans to modify at each iteration makes the approach slower by design without an excellent heuristic.

Works on PSP didn't stop at that point (Nguyen and Kambhampati 2001) since it has unique advantages over classical planning. Indeed, by using backward chaining, PSP algorithms have are sound and complete and therefore guarantee to find a solution if it exists (Sjöberg and Nissar 2015).

As PSP finds partially ordered plans, it is also by nature more flexible. Indeed, multiple totally ordered plans are contained within a partially ordered one, they are called linearizations. So, when wanting to have several plans with low diversity, PSP is the way to go.

5.1.1.1 Definitions

In **BEFORE** we have formalized how PSP works in the general planning formalism. However, since this formalism is being introduced in the present document, it isn't used by the rest of the community. This means that we still need to define the classical Partially Ordered with Causal Link (POCL) algorithm. In order to define this algorithm, we need to explain the notions of flaws and resolvers.

Definition 37 (Flaws). Flaws are constraints violations within a plan. The set of flaws in a plan π is noted \otimes_π . There are different kinds of flaws in classical PSP and additional ones can be defined depending on the application.

Classical flaws often have a few common features. They are often **constructive** since they require an *addition* of causal links and steps in a plan to be fixed. They have a *proper fluent f* that is the cause of the violation in the plan the flaw is representing

and a *need*er a_n that is the action requiring the proper fluent be fulfilled. In classical PSP flaws are either:

- **Subgoals**, also called *open condition* that are yet to be supported by a *provider* a_p . We note subgoals $\otimes_{a_n}^{\dagger}(f)$.
- **Threats** are caused by steps that can break a causal link with their effects. They are called *breakers* of the threatened link. A step a_b threatens a causal link $l_t = a_p \xrightarrow{f} a_n$ if and only if $(\text{eff}(a_b) \not\models f) \wedge (a_b \succ a_p \wedge a_n \succ a_b)$. Said otherwise, the breaker can potentially cancel an effect of a providing step a_p , before it gets used by its needer a_n . We note threats $\otimes_{a_n}^{\dagger}(f, a_b)$.

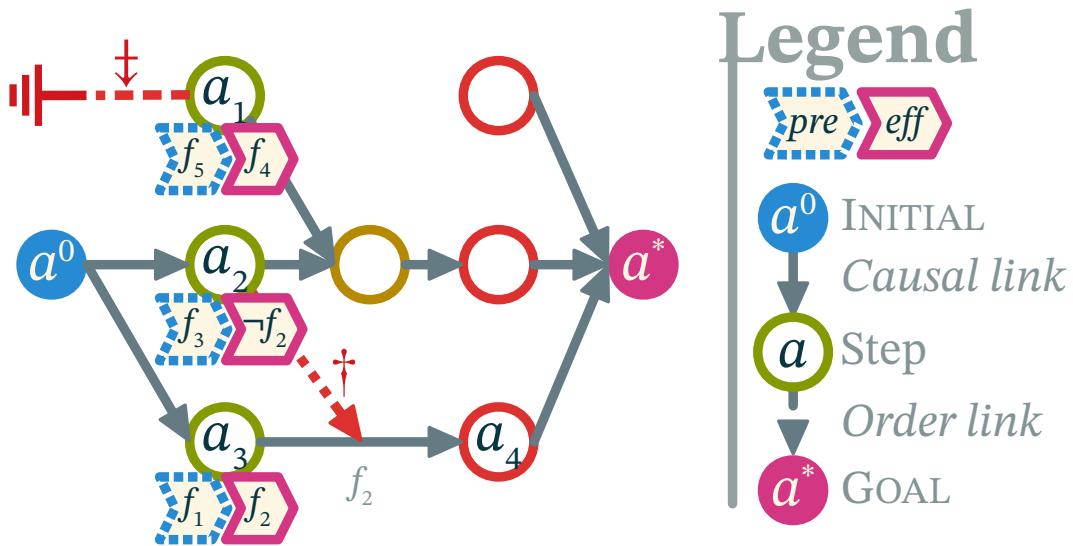


Figure 5.2: Example of partial plan having flaws

Example 28. In figure 5.2 we present a partially ordered plan with two typical flaws. The first is a subgoal missing from the plan to fulfill the f_5 precondition of a_1 .

The second flaw in the figure 5.2 is the threat between a_2 and a causal link outgoing from a_3 . This happens because nothing prevents a_2 to be executed after a_3 and negate the fluent f_2 needed by the next step.

These flaws need to be fixed in order for the plan to be valid. In POCL it is done by finding their resolvers.

Definition 38 (Resolvers). A resolver is a plan refinement that attempts to solve a flaw \otimes_{a_n} . Since classical flaws are constructive, the classical resolvers are called *positive*. They are defined as follows:

- For subgoals, the resolvers are a potential causal link containing the proper fluent f of a given subgoal in their causes while taking the needer step a_n as their target and a **provider** step a_p as their source. They are noted $\odot_{a_p}^+(f) = a_p \xrightarrow{f} a_n$.

- For threats, we usually consider only two resolvers: **demotion** ($a_b > a_p$) and **promotion** ($a_n > a_b$) of the breaker relative to the threatened link. We call the added causeless causal link a **guarding link**. The resolvers for threats are noted $\odot_>^+(\otimes_{a_n}^\dagger(f, a_b)) = a_p \rightarrow a_b$ for promotion and $\odot_<^+(\otimes_{a_n}^\dagger(f, a_b)) = a_b \rightarrow a_n$ for demotion.

It is possible to introduce extra resolvers to fix custom flaws. In such a case we call positive resolvers, those which add causal links and steps to the plan and negative those that removes causal links and steps. It is preferable to engineer flaws and resolver not to mix positive and negative aspect at once because of the complicated side effects that might result from it.

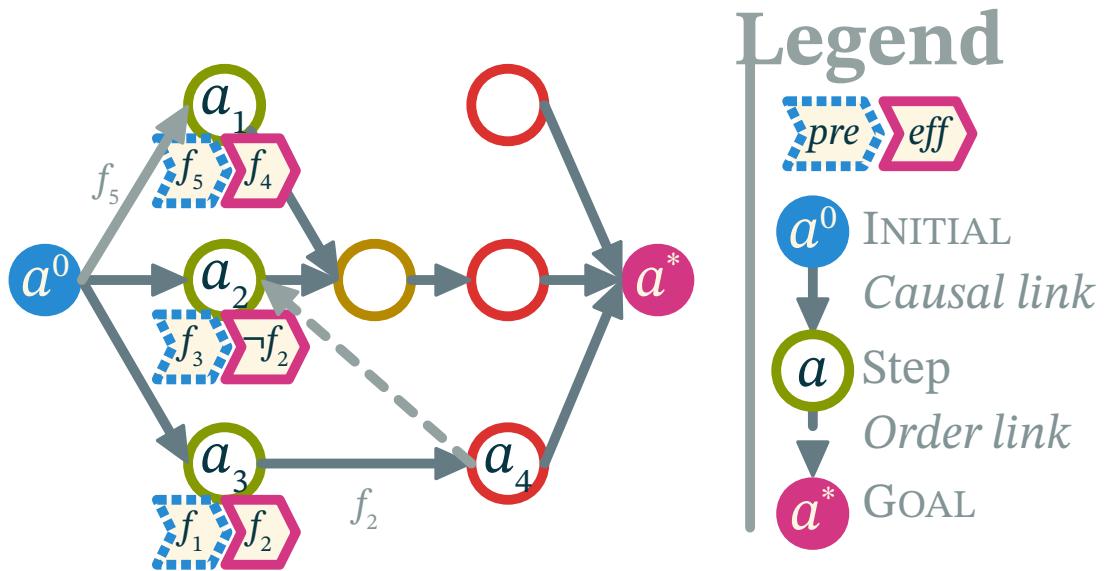


Figure 5.3: Example of resolvers that fixes the previously illustrated flaws

Example 29. From our previous example, we present the complete plan in figure 5.3. The subgoal needs to be fixed by inserting another causal link to provide the missing fluent and inserting any necessary steps to do so. In that case the initial state happens to provide the necessary fluent so we simply add a causal link for it.

For the threat, the solution is to either promote or demote a_2 so that it doesn't interfere with the causal link between a_3 and a_4 . We chose here to demote a_2 so it requires a_4 to be executed before it.

The application of a resolver does not necessarily mean progress. It can have consequences that may require reverting its application in order to respect the backtracking of the POCL algorithm.

Definition 39 (Side effects). Flaws that are caused by the application of a resolver are called *related flaws*. They are inserted into the *agenda* with each application of a resolver:

- Related subgoals* are all the new open conditions inserted by new steps.

An agenda is a flaw container used for the flaw selection of POCL.

- *Related threats* are the causal links threatened by the insertion of a new step or the deletion of a guarding link.

Flaws can also become irrelevant when a resolver is applied. It is always the case for the targeted flaw, but this can also affect other flaws. Those *invalidated flaws* are removed from the agenda upon detection:

- *Invalidated subgoals* are subgoals satisfied by the new causal links or the removal of their needer.
- *Invalidated threats* happen when the breaker no longer threatens the causal link because the order guards the threatened causal link or either of them have been removed.

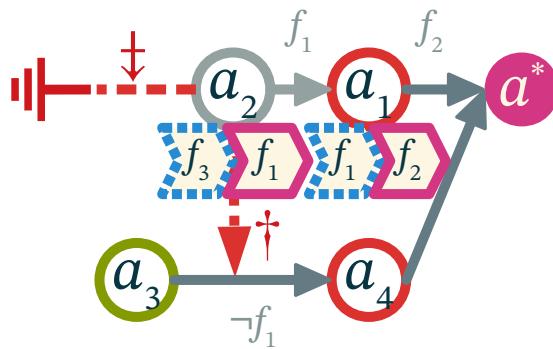


Figure 5.4: Example of the side effects of the application of a resolver

Example 30. In our example, by adding the step a_2 to fix an unsupported subgoal needed by a_1 , we introduced another subgoal to support the new step that also threatens the causal link between a_3 and a_4 .

5.1.1.2 Classical POCL Algorithm

In algorithm 5 we present a generic version of POCL inspired by Ghallab et al. (2004, sec. 5.4.2).

For our version of POCL we follow a refinement procedure that works in several generic steps. In figure ?? we detail the resolution of a subgoal as done in the algorithm 5.

The first is the search for resolvers. It is often done in two separate steps: first, select the candidates and then check each of them for validity. This is done using the polymorphic function `solve` at line 5.

In the case of subgoals, variable unification is performed to ensure the compatibility of the resolvers. Since this step is time-consuming, the operator is instantiated accordingly at this step to factories the computational effort. Composite operators have also all their methods instantiated at this step if they are selected as a candidate.

Then a resolver is picked non-deterministically for applications (this can be heuristically driven). At line 7 the resolver is effectively applied to the current plan. All side effects and invalidations are handled during the update of the agenda at line 8. If a problem occurs, line 11 backtracks and tries other resolvers. If no resolver fits the flaw,

Algorithm 5 POCL Algorithm

```
1: function POCL(Agenda  $\mathcal{A}$ , Action  $\omega$ )
2:   if  $\mathcal{A} = \emptyset$  then                                 $\triangleright$  Populated agenda needs to be provided
3:     return Success                                 $\triangleright$  Stops all recursion
4:   Flaw  $\otimes \leftarrow \{\mathcal{A}\}$                        $\triangleright$  Heuristically chosen flaw
5:   Resolvers  $\odot \leftarrow \text{SOLVE}(\otimes, \{\Pi(\omega)\})$      $\triangleright$  The root operator has only one method for PSP
6:   for all  $\odot \in \odot$  do                          $\triangleright$  Non-deterministic choice operator
7:     APPLY( $\odot, \pi$ )                                 $\triangleright$  Apply resolver to partial plan
8:     Agenda  $\mathcal{A}' \leftarrow \text{UPDATE}(\mathcal{A})$ 
9:     if POCL( $\mathcal{A}', \omega$ ) = Success then            $\triangleright$  Refining recursively
10:    return Success
11:    REVERT( $\mathcal{A}, \pi$ )                             $\triangleright$  Failure, undo resolver application
12:     $\mathcal{A} \leftarrow \mathcal{A} \cup \{\otimes\}$                  $\triangleright$  Flaw was not resolved
13:  return Failure                                 $\triangleright$  Revert to last non-deterministic choice
```

the algorithm backtracks to previous resolver choices to explore all the possible plans and ensure completeness.

5.1.1.3 Existing PSP Planners

Related works already tried to explore new ideas to make PSP an attractive alternative to regular state-based planners like the appropriately named “Reviving partial order planning” (Nguyen and Kambhampati 2001) and VHPOP (Younes and Simmons 2003). More recent efforts (Coles *et al.* 2011; Sapena *et al.* 2014) are trying to adapt the powerful heuristics from state-based planning to PSP’s approach. An interesting approach of these last efforts is found in (Shekhar and Khemani 2016) with meta-heuristics based on offline training on the domain. Yet, we clearly note that only a few papers lay the emphasis upon plan quality using PSP (Ambite and Knoblock 1997; Say *et al.* 2016).

5.1.2 Plan Repair & Reuse

In online planning, the plan is computed frequently from a changing initial state. This means that the previous plan is very often available. In order to take advantage of the effort invested in previous plans, it is tempting to simply reuse the existing plan instead of replanning from scratch. Most work on the field focus on monitoring execution and find ways to make resilient plans.

In such a case, a lot of the planning model can be subject to uncertainty. Indeed, the execution of an action can fail because an external event changed a precondition required to do it or the model itself can be inaccurate.

The idea of reusing plan emerged early on (Nebel and Koehler 1995) but with a caveat: often repairing needed more effort than replanning. So plan repair became more of a gamble and needed incentives to reuse an existing plan given an application. For example, such process is useful for multi-agent planning where a significant change of plan is expansive among agents (Ephrati and Rosenschein 1993; Alami *et al.* 1995;

Sugawara 1995; Borrajo 2013; Luis and Borrajo 2014). This motivation for plan repair comes from plan merging and is needed in cooperative environments.

The question of the efficiency of replanning vs. repairing has been since studied extensively under several aspects (Van Der Krogt and De Weerdt 2005; Fox et al. 2006). The emergence of diverse planning and requirement on plan stability of execution monitoring gave new research on the subject. At this point, most of the literature focuses on a case-based planning, where a plan library is already provided and the planner must select a plan and repair it to fit a given case (Gerevini et al. 2013; Borrajo et al. 2015).

A recent work of Zhuo and Kambhampati (2017) gives an interesting approach to the problem by questioning the domain. Indeed, the need to re-plan can be an opportunity to revised issues in the current model and to improve it by adding newly found a solution to complete the given planning model.

In our case, we focus on PSP and how to make plans repair efficiently using that technique. Classical PSP algorithms don't take as an input an existing plan but can be enhanced to fit plan to repair, as for instance in (Van Der Krogt and De Weerdt 2005). Usually, PSP algorithms take a problem as an input and use a loop or a recursive function to refine the plan into a solution. We can't solely use the refining recursive function to be able to use our existing partial plan. This causes multiples side effects if the input plan is suboptimal. This problem was already explored as of LGP-adapt (Borrajo 2013) that explains how reusing a partial plan often implies replanning parts of the plan.

5.1.3 HTN

"HTN planners differ from classical planners in what they plan for and how they plan for it. In an HTN planner, the objective is not to achieve a set of goals but instead to perform some set of tasks."

Ghallab et al.
(2004)

Planning using HTN gives a completely different approach to the problem and its formulation. In this formalism, actions are composite tasks and there is no goal other than to complete the root task. One can find similarities with our general planning formalism and it isn't a coincidence. Indeed, HTN is more general than planning and therefore one need to be able to allow for this level of expressivity.

This expressivity comes at a cost. HTN problems are on a complexity category that is significantly harder than regular STRIPS planning:

From Bercher and Höller (2018) HTN
Tutorial at ICAPS
2018

Table 5.1: HTN Expressivity associated with their respective complexity classes
{#tbl:htn}

<i>Restrictions</i>	<i>Complexity</i>
Classical	PSPACE
Task Insertion	NEXPTIME
Totally Ordered	EXPTIME
Regular	PSPACE
Acyclic	NEXPTIME

<i>Restrictions</i>	<i>Complexity</i>
Tail-recursive	EXPSPACE

HTN is often combined with classical approaches since it allows for a more natural expression of domains making expert knowledge easier to encode. These kinds of planners are named **decompositional planners** when no initial plan is provided (Fox 1997). Most of the time the integration of HTN simply consists in calling another algorithm when introducing a composite operator during the planning process. The DUET planner by Gerevini *et al.* (2008) does so by calling an instance of a HTN planner based on the task insertion called SHOP2 (Nau *et al.* 2003) to decompose composite actions. Some planners take the integration further by making the decomposition of composite actions into a special step in their refinement process. Such works include the discourse generation oriented DPOCL (Young and Moore 1994) and the work of Kambhampati *et al.* (1998) generalizing the practice for decompositional planners.

In our case, we chose a class of hierarchical planners based on Plan-Space Planning (PSP) algorithms (Bechon *et al.* 2014; Dvorak *et al.* 2014; Bercher *et al.* 2014) as a reference approach. The main difference here is that the decomposition is integrated into the classical POCL algorithm by only adding new types of flaws. This allows keeping all the flexibility and properties of POCL while adding the expressivity and abstraction capabilities of HTN.

5.2 LOLLIPOP

That first planner is a prototype destined to test the feasibility of a plan to repair approach using PSP. This is done through the addition of new flaw types in a classical POCL algorithm.

Another issue is caused by the need to get any existing partial plan as an input. This plan can contain any problems and inconsistency since there are trivial ways to ensure consistence between the plan given and the new initial state or goal.

5.2.1 Operator Graph

In order to make POCL faster we experimented a means to build a dependency graph for operators in the domain, computed at domain compilation time. The idea is to extract a partial plan from this graph by pruning it based on the provided goal and use that plan to quick start the planning process. First we need to define the operator graph.

Definition 40 (Operator Graph). An operator graph g_O of a set of operators O is a labeled directed graph that binds two operators with the causal link $o_1 \xrightarrow{f} o_2$ if and only if there exists at least one fluent so that $(f \in \text{eff}(o_1)) \wedge f \models \text{pre}(o_2)$.

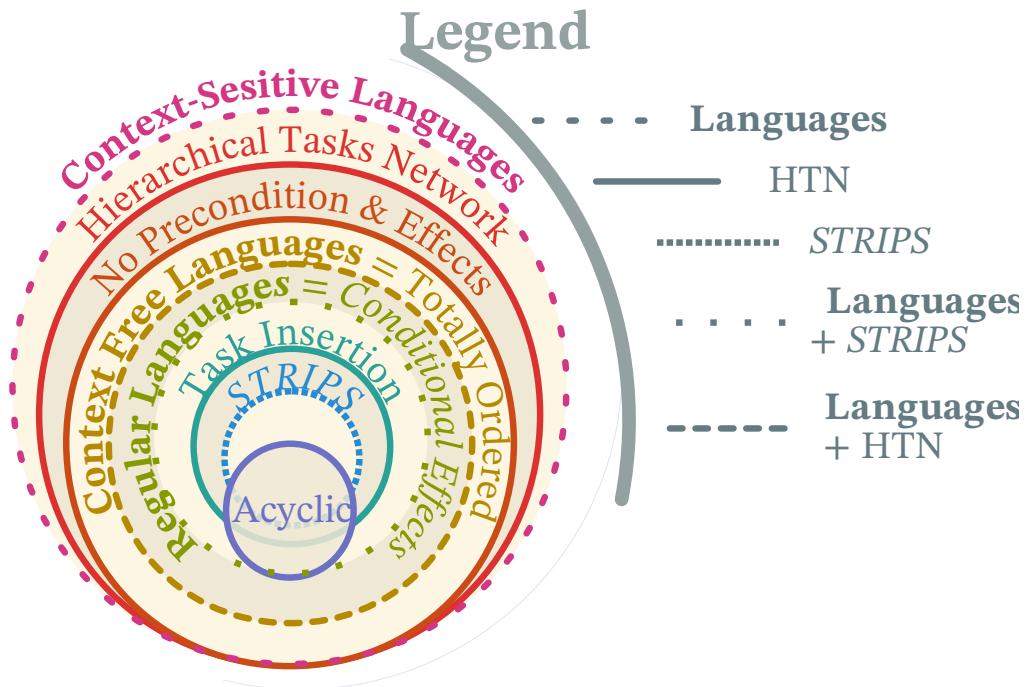


Figure 5.5: Venn diagram of the expressivity class of HTN paradigms.

This definition was inspired by the notion of domain causal graph as explained in (Göbelbecker *et al.* 2010) and originally used as a heuristic in (Helmert *et al.* 2011). Causal graphs have fluents as their nodes and operators as their edges. Operator graphs are the opposite: an *operator dependency graph* for a set of actions. A similar structure was used in (Peot and Smith 1994) that builds the operator dependency graph of goals and uses precondition nodes instead of labels. We call *co-dependent* operators that form a cycle. If the cycle is made of only one operator (self-loop), then it is called *auto-dependent*.

While building this operator graph, we need a **providing map** that indicates, for each fluent, the list of operators that can provide it. This is a simpler version of the causal graphs that is reduced to an associative table easier to maintain. The list of providers can be sorted to drive resolver selection (as detailed in section ??). A **needing map** is also built but is only used for operator graph generation. We note $g_{\mathcal{D}}$ the operator graph built with the set of operators in the domain \mathcal{D} .

Example 31. In the figure 5.6, we illustrate the application of this mechanism on our example from figure ?? Continuous lines correspond to the *domain operator graph* computed during domain compilation time.

The generation of the operator graph is detailed in algorithm 6. It explores the operator space and builds a providing and a needing map that gives the provided and needed fluents for each operator. Once done it iterates on every precondition and searches for a satisfying cause to add the causal links to the operator graph.

To apply the notion of operator graphs to planning problems, we just need to add the initial and goal steps to the operator graph. In figure 5.6, we depict this insertion with

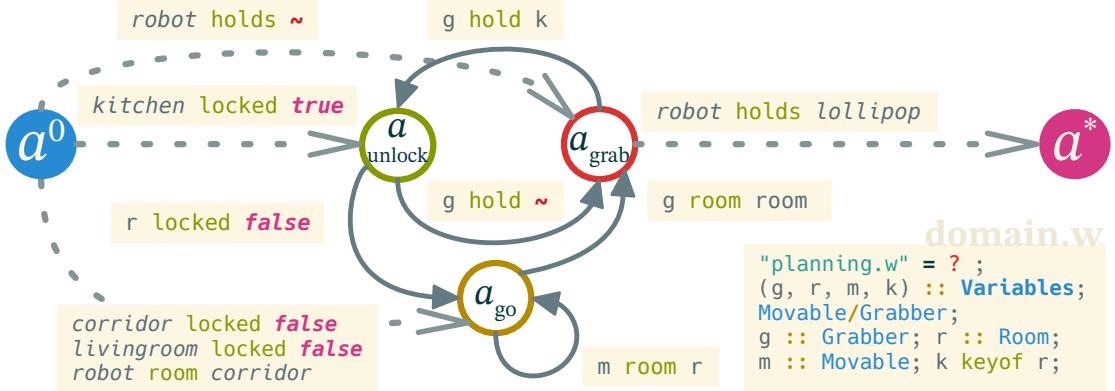


Figure 5.6: Diagram of the operator graph of example domain. Full arrows represent the domain operator graph and dotted arrows the dependencies added to inject the initial and goal steps.

Algorithm 6 Operator graph generation and update algorithm

```

function ADDVERTEX(Action a)
  CACHE(a)                                ▷ Update of the providing and needing map
  if binding then                      ▷ boolean that indicates if the binding was requested
    BIND(a)
  function CACHE(Action a)
    for all f ∈ eff(a) do            ▷ Adds a to the list of providers of f
      ADD( $A_p$ , f, a)
    ...
    function BIND(Action a)
      for all f ∈ pre(a) do
        if f ∈  $A_p$  then
          for all π ∈ GET( $A_p$ , f) do
            Link l ← GETEDGE(π, a)           ▷ Create the link if needed
            ADDCAUSE(l, f)                  ▷ Add the fluent as a cause
        ...
        ▷ Same operation with needing and effects
  
```

our previous example using dotted lines. However, since operator graphs may have cycles, they can't be used directly as input to POCL algorithms to ease the initial back chaining. Moreover, the process of refining an operator graph into a usable one could be more computationally expensive than POCL itself.

In order to give a head start to the LOLLIPOP algorithm, we propose to build operator graphs differently with the algorithm detailed in algorithm 7. A similar notion was already presented as "basic plans" in (Sebastia *et al.* 2000). These "basic" partial plans use a more complete but slower solution for the generation that ensures that each selected steps are *necessary* for the solution. In our case, we built a simpler solution that can solve some basic planning problems but that also makes early assumptions (since our algorithm can handle them). It does a simple and fast backward construction of a partial plan driven by the providing map. Therefore, it can be tweaked with the powerful heuristics of state search planning.

Algorithm 7 Safe operator graph generation algorithm

```

function SAFE(Action  $\omega$ )
    Stack<Action>  $open \leftarrow [a^*]$ 
    Stack<Action>  $closed \leftarrow \emptyset$ 
    while  $open \neq \emptyset$  do
        Action  $a \leftarrow \text{POCL}(open)$                                  $\triangleright$  Remove  $a$  from  $open$ 
        PUSH( $closed, a$ )
        for all  $f \in \text{pre}(a)$  do
            Actions  $A_p \leftarrow \text{GETPROVIDING}(\pi, f)$                  $\triangleright$  Sorted by usefulness
            if  $A_p = \emptyset$  then                                          $\triangleright$  (see section ??)
                 $S \leftarrow S \setminus \{\pi\}$ 
                continue
            Action  $a' \leftarrow \text{GETFIRST}(\pi)$ 
            if  $a' \in closed$  then
                continue
            if  $a' \notin S$  then
                PUSH( $open, a'$ )
                 $S \leftarrow S \cup \{a'\}$ 
            Link  $l \leftarrow \text{GETEDGE}(a', a)$                                  $\triangleright$  Create the link if needed
            ADDCAUSE( $l, f$ )                                               $\triangleright$  Add the fluent as a cause

```

This algorithm is useful since it is specifically used on goals. The result is a valid partial plan that can be used as input to POCL algorithms.

5.2.2 Negative Refinements

The classical POCL algorithm works upon a principle of positive plan refinements. The two standard flaws (subgoals and threats) are fixed by *adding* steps, causal links, or variable binding constraints to the partial plan. Online planning needs to be able to *remove* parts of the plan that are not necessary for the solution. Since we assume that the input partial plan is quite complete, we need to define new flaws to optimize and fix this plan. These flaws are called *negative* as their resolvers apply subtractive refinements on partial plans.

Definition 41 (Alternative). An alternative is a negative flaw that occurs when there is a better provider choice for a given link. An alternative to a causal link $a_p \xrightarrow{f} a_n$ is a provider a_b that has a better *utility value* than a_p .

The **utility value** of an operator is a measure of usefulness being the base of our ranking mechanism detailed in section ???. It uses the incoming and outgoing degrees of the operator in the domain operator graph to measure its usefulness.

Finding an alternative to an operator is computationally expensive. It requires searching a better provider for every fluent needed by a step. To simplify that search, we select only the best provider for a given fluent and check if the one used is the same. If not, we add the alternative as a flaw. This search is done only on updated steps for online planning. Indeed, the safe operator graph mechanism is guaranteed to only choose the best provider (algorithm 7 at line 12). Furthermore, subgoals won't introduce new fixable alternatives as they are guaranteed to select the best possible provider.

Definition 42 (Orphan). An orphan is a negative flaw that occurs when a step in the partial plan (other than the initial or goal step) is not participating in the plan. Formally a_o is an orphan if and only if $a_o \neq a^0 \wedge a_o \neq a^* \wedge (|\chi_{\leftarrow}(a_o)| = 0) \vee \{= (\emptyset) : \chi_{\leftarrow}(a_o)\}$.

With $\chi_{\leftarrow}(a_o)$ being the set of *outgoing causal links* of a_o in π . This last condition checks for *dangling orphans* that are linked to the goal with only bare causal links (introduced by threat resolution).

The solution to an alternative is a negative refinement that simply removes the targeted causal link. This causes a new subgoal as a side effect, which will focus on its resolver by its rank (explained in section ??) and then pick the first provider (the most useful one). The resolver for orphans is the negative refinement that is meant to remove a step and its incoming causal link while tagging its providers as potential orphans.

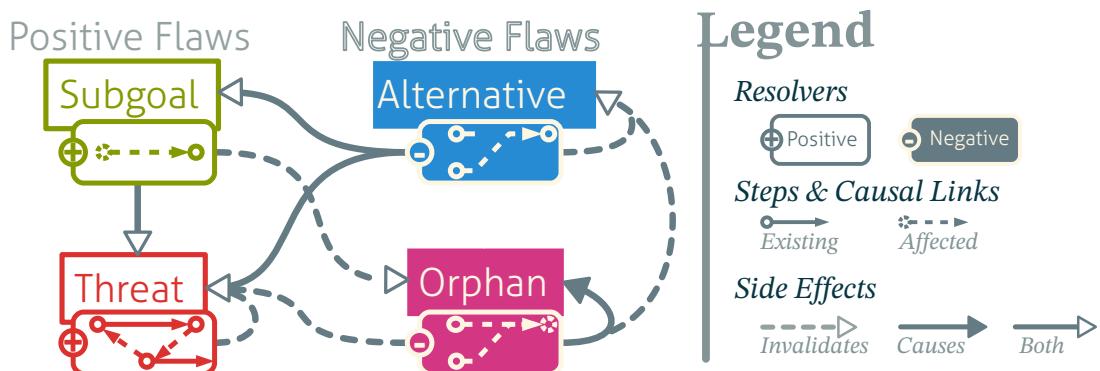


Figure 5.7: Schema representing flaws with their signs, resolvers and side effects relative to each other

The side effects mechanism also needs an upgrade since the new kinds of flaws can interfere with one another. This is why we extend the side effect definition (cf. definition ??) with a notion of sign.

Definition 43 (Signed Side Effects). A signed side effect is either a regular *causal side effect* or an *invalidating side effect*. The sign of a side effect indicates if the related flaw needs to be added or removed from the agenda.

The figure 5.7 exposes the extended notion of signed resolvers and side effects. When treating positive resolvers, nothing needs to change from the classical method. When dealing with negative resolvers, we need to search for extra subgoals and threats. Deletion of causal links and steps can cause orphan flaws that need to be identified for removal.

In the method described in (Peot and Smith 1993), a **invalidating side effect** is explained under the name of *DEnd* strategy. In classical POCL, it has been noticed that threats can disappear in some cases if subgoals or other threats were applied before them. For our mechanisms, we decide to gather under this notion every side effect that removes the need to consider a flaw. For example, orphans can be invalidated if a subgoal selects the considered step. Alternatives can remove the need to compute further subgoal of an orphan step as orphans simply remove the need to fix any flaws that concern the selected step.

These interactions between flaws are decisive for the validity and efficiency of the whole model, that is why we aim to drive flaw selection in a rigorous manner.

5.2.3 Usefulness Heuristic

FIXME: Redo this section and be clear about what it is and when it is computed. Needing an execution schema too with phases and stuff

Resolvers and flaws selection are the keys to improving performances. Choosing a good resolver helps to reduce the branching factor that accounts for most of the time spent on running POCL algorithms (Kambhampati 1994). Flaw selection is also important for efficiency, especially when considering negative flaws which can conflict with other flaws.

Conflicts between flaws occur when two flaws of opposite sign target the same element of the partial plan. This can happen, for example, if an orphan flaw needs to remove a step needed by a subgoal or when a threat resolver tries to add a promoting link against an alternative. The use of side effects will prevent most of these occurrences in the agenda but a base ordering will increase the general efficiency of the algorithm.

Based on the figure 5.7, we define a base ordering of flaws by type. This order takes into account the number of flaw types affected by causal side effects.

1. **Alternatives** will cut causal links that have a better provider. It is necessary to identify them early since they will add at least another subgoal to be fixed as a related flaw.

2. **Subgoals** are the flaws that cause most of the branching factor in POCL algorithms. This is why we need to make sure that all open conditions are fixed before proceeding on finer refinements.
3. **Orphans** remove unneeded branches of the plan. Yet, these branches can be found out to be necessary for the plan to meet a subgoal. Since a branch can contain many actions, it is preferable to leave the orphan in the plan until they are no longer needed. Also, threats involving orphans are invalidated if the orphan is resolved first.
4. **Threats** occur quite often in the computation. Searching and solving them is computationally expensive since the resolvers need to check if there are no paths that fix the flaw already. Many threats are generated without the need of resolver application (Peot and Smith 1993). That is why we rank all related subgoals and orphans before threats because they can add causal links or remove threatening actions that will fix the threat.

Resolvers need to be ordered as well, especially for the subgoal flaws. Ordering resolvers for a subgoal is the same operation as choosing a provider. Therefore, the problem becomes "how to rank operators?". Usually, each operator has an assigned cost in the domain, but more often than not, costs are hard to estimate manually. In our case we need an automated way to rank operators. The most relevant information on an operator is how useful it may be to other actions in the plan and how hard is it to realize.

Since this may be computationally expensive to compute while planning, the evaluation of the cost of an operator is done offline using the operator graph.

The first metric to compute this heuristic is the degree of the operator.

Definition 44 (Degree of an operator). Degrees are a measurement of the usefulness of an operator. Such a notion is derived from the incoming and outgoing degrees of a node in the operator graph.

We note $|\chi_{\rightarrow}(a)|$ being the *outgoing degree* of a in the directed graph formed by π and $|\chi_{\leftarrow}(a)|$ being the *incoming degree* of a in the directed graph formed by π respectively the outgoing and incoming degrees of an operator in a plan π . These represent the number of causal links that goes out or toward the operator. We call proper degree of an operator $|eff(a)|$ and $|pre(a)|$ the number of preconditions and effects that reflect its intrinsic usefulness.

There are several ways to use the degrees as indicators. The *utility value* increases with every outgoing degree, since this reflects a positive participation in the plan. It decreases with every negative degree since actions with higher incoming degrees are harder to satisfy. The utility value bounds are useful when selecting special operators. For example, a user-specified constraint could be laid upon an operator to ensure it is only selected as a last resort. This operator will be affected with the smallest utility value possible. More commonly, the highest value is used for initial and goal steps to ensure their selection.

Our ranking mechanism is based on scores noted $\chi(a)$. A score is a tuple of metrics:

In this section χ is the connectivity of the operator graph g_0 .

- $\mathfrak{r}_1(a) = |\chi_{\leftarrow}(a)|$ is the positive degree of a in the domain operator graph. This will give a measurement of the predicted usefulness of the operator.
- $\mathfrak{r}_2(a) = |\otimes_a^\dagger|$ is the number of open conditions of a in the domain operator graph. This is symptomatic of action that can't be satisfied without a compliant initial step.
- $\mathfrak{r}_3(a) = |pre(a)|$ is the proper negative degree of a . Having more preconditions will likely add subgoals.
- $\mathfrak{r}_4(a) = \{\min_{+\infty}(n) : n \in \mathbb{N} \wedge (a \rightarrow a) \in \chi^n \wedge \chi^n \neq \chi^+\}$ is the size of the shortest cycle involving a in the operator graph or $+\infty$ if there is none. Having this value at 1 is usually symptomatic of a *toxic operator* (cf. definition 46). Having an operator behaving this way can lead to backtracking because of operator instantiation.

The computation of the cost of the operator is done by multiplying the score tuple with a weighted parameter tuple α given by the user. The cost is then:

$$\mathfrak{c}(a) = - \sum_{i=1}^4 \alpha_i \mathfrak{r}_i(a)$$

In practice, α_1 is positive, and the rest is negative. It is also better to make sure that $-1 \leq \alpha_4 \leq 0$ so that the penalties goes down as the cycles gets bigger.

This respects the criteria of having a bound for the *utility value* as it ensures that it remains positive with 0 as a minimum bound and $+\infty$ for a maximum. The initial and goal steps have their utility values set to the upper bound to ensure their selection over other steps.

Choosing to compute the resolver selection at operator level has some positive consequences on the performances. Indeed, this computation is much lighter than approaches with heuristics on plan space (Shekhar and Khemani 2016) as it reduces the overhead caused by real time computation of heuristics on complex data. In order to reduce this overhead more, the algorithm sorts the providing associative array to easily retrieve the best operator for each fluent. This means that the evaluation of the heuristic is done only once for each operator. This reduces the overhead and allows for faster results on smaller plans.

5.2.4 Algorithm

The LOLLIPOP algorithm uses the same refinement algorithm as described in algorithm 5. The differences reside in the changes made on the behavior of resolvers and side effects. In line 7 of algorithm 5, LOLLIPOP algorithm applies negative resolvers if the selected flaw is negative. In line ??, it searches for both signs of side effects. Another change resides in the initialization of the solving mechanism and the domain as detailed in algorithm 8. This algorithm contains several parts. First, the `domainInit` function corresponds to the code computed during the domain compilation time. It will prepare the rankings, the operator graph, and its caching mechanisms. It will also use strongly connected component detection algorithm to detect cycles. These cycles

are used during the base score computation (line 11). We add a detection of illegal fluents and operators in our domain initialization (line 5). Illegal operators are either inconsistent or toxic.

Definition 45 (Inconsistent operators). An operator a is contradictory iff $\exists f \{f, \neg f\} \in eff(o) \vee \{f, \neg f\} \in pre(o)$.

Definition 46 (Toxic operators). Toxic operators have effects that are already in their preconditions or empty effects. An operator o is toxic iff $pre(o) \cap eff(o) \neq \emptyset \vee eff(o) = \emptyset$.

Toxic actions can damage a plan as well as make the execution of POCL algorithm longer than necessary. This is fixed by removing the toxic fluents ($pre(a) \not\subseteq eff(a)$) and by updating the effects with $eff(a) = eff(a) \setminus pre(a)$. If the effects become empty, the operator is removed from the domain.

The lollipopInit function is executed during the initialization of the solving algorithm. We start by realizing the scores, then we add the initial and goal steps in the providing map by caching them. Once the ranking mechanism is ready, we sort the providing map. With the ordered providing map, the algorithm runs the fast generation of the safe operator graph for the problem's goal.

The last part of this initialization (line 21) is the agenda population that is detailed in the populate function. During this step, we perform a search of alternatives based on the list of updated fluents. Online updates can make the plan outdated relative to the domain. This forms liar links :

Definition 47 (Liar links). A liar link is a link that doesn't hold a fluent in the preconditions or effect of its source and target. We note:

$$a_i \xrightarrow{f} a_j \mid f \notin eff(a_i) \cap pre(a_j)$$

A liar link can be created by the removal of an effect or preconditions during online updates (with the causal link still remaining).

We call lies the fluents that are held by links without being in the connected operators. To resolve the problem, we remove all lies. We delete the link altogether if it doesn't bear any fluent as a result of this operation. This removal triggers the addition of orphan flaws as side effects.

While the list of updated operators is crucial for solving online planning problems, a complementary mechanism is used to ensure that LOLLIPOP is complete. User provided plans have their steps tagged. If the failure has backtracked to a user-provided step, then it is removed and replaced by subgoals that represent each of its participation in the plan. This mechanism loops until every user provided steps have been removed.

Algorithm 8 LOLLIPOP initialization (preprocessing) mechanisms

```

function DOMAININIT(Operators  $A$ )
    operatorgraph  $g_O$ 
    Score  $S$ 
    for all Operator  $a \in A$  do
        if ISILLEGAL( $a$ ) then
             $A \leftarrow A \setminus \{a\}$ 
            continue
        ADDVERTEX( $a, g_O$ )
        CACHE( $p, a$ )
    Cycles  $C \leftarrow$  STRONGLYCONNECTEDCOMPONENT( $g_O$ )
     $S \leftarrow$  BASESCORES( $A, \mathcal{D}^{\Pi}$ )
     $i \leftarrow$  INAPPLICABLES( $\mathcal{D}^{\Pi}$ )
     $e \leftarrow$  EAGERS( $\mathcal{D}^{\Pi}$ )
function LOLLIPOPINIT(Problem  $\mathcal{P}$ )
    REALIZE( $S, \mathcal{P}$ )
    CACHE(providing,  $I$ )
    CACHE(providing,  $G$ )
    SORT(providing,  $S$ )
    if  $L = \emptyset$  then
         $\mathcal{P}^{\Pi} \leftarrow$  SAFE( $\mathcal{P}$ )
        POPULATE( $a, \mathcal{P}$ )
    function POPULATE(Agenda  $a$ , Problem  $\mathcal{P}$ )
        for all Update  $u \in U$  do
            Fluents  $F \leftarrow eff(u.new) \setminus eff(u.old)$ 
            for all Fluent  $f \in F$  do
                for all Operator  $o \in$  BETTER(providing,  $f, o$ ) do
                    for all Link  $l \in L^+(o)$  do
                        if  $f \in l$  then
                            ADDALTERNATIVE( $a, f, o, l_{\leftarrow}, \mathcal{P}$ )
             $F \leftarrow eff(u.old) \setminus eff(u.new)$ 
            for all Fluent  $f \in F$  do
                for all Link  $l \in L^+(u.new)$  do
                    if ISLIAR( $l$ ) then
                         $L \leftarrow L \setminus \{l\}$ 
                        ADDORPHANS( $a, u, \mathcal{P}$ )
            ...
            for all Operator  $o \in S$  do
                ADDSUBGOALS( $a, o, \mathcal{P}$ )
                ADDTHREATS( $a, o, \mathcal{P}$ )

```

5.2.5 Theoretical and Empirical Results

As proven in (Penberthy *et al.* 1992), the classical POCL algorithm is *sound* and *complete*.

First, we define some new properties of partial plans. The following properties are taken from the original proof. We present them again for convenience.

Definition 48 (Full Support). A partial plan π is fully supported if each of its steps $o \in S$ is fully supported. A step is fully supported if each of its preconditions $f \in \text{pre}(o)$ is supported. A precondition is fully supported if there exists a causal link l that provides it. We note:

$$\Downarrow \pi \equiv \begin{aligned} & \forall o \in S \forall f \in \text{pre}(o) \exists l \in L_{\pi}^-(o) : \\ & (f \in l \wedge \nexists t \in S(l_{\rightarrow} > t > o \wedge \neg f \in \text{eff}(t))) \end{aligned}$$

with $L_{\pi}^-(o)$ being the incoming causal links of o in π and l_{\rightarrow} being the source of the link.

Definition 49 (Partial Plan Validity). A partial plan is a **valid solution** of a problem \mathcal{P} if it is *fully supported* and *contains no cycles*. The validity of π regarding a problem \mathcal{P} is noted $\pi \models (\mathcal{P} \equiv \Downarrow \pi \wedge (C(\pi) = \emptyset))$ with $C(\pi)$ being the set of cycles in π .

5.2.5.1 Proof of Soundness

In order to prove that this property applies to LOLLIPOP, we need to introduce some hypothesis:

- operators updated by online planning are known.
- user provided steps are known.
- user provided plans don't contain illegal artifacts. This includes toxic or inconsistent actions, lying links and cycles.

Based on the definition 49 we state that:

$$\left(\begin{array}{l} \forall \text{pre} \in \text{pre}(G) : \\ \Downarrow \text{pre} \wedge \forall o \in L_{\pi}^-(G)_{\rightarrow} \forall \text{pre}' \in \text{pre}(o) : \\ \Downarrow \text{pre}' \wedge C_o(\pi) = \emptyset \end{array} \right) \implies \pi \models \mathcal{P}$$

{#eq:recursivevalidity} where $L_{\pi}^-(G)_{\rightarrow}$ is the set of direct antecedents of G and $C_o(\pi)$ is the set of fluents containing o in π .

This means that π is a solution if all preconditions of G are satisfied. We can satisfy these preconditions using operators if and only if their preconditions are all satisfied and if there is no other operator that threatens their supporting links.

First, we need to prove that equation ?? holds on LOLLIPOP initialization. We use our hypothesis to rule out the case when the input plan is invalid. The algorithm 7 will only solve open conditions in the same way subgoals do it. Thus, safe operator graphs are valid input plans.

Since the soundness is proven for regular refinements and flaw selection, we need to consider the effects of the added mechanisms of LOLLIPPOP. The newly introduced refinements are negative, they don't add new links:

$$\forall f \in \mathcal{F}(\pi) \forall r \in r(f) : C_\pi(f.n) = C_{f(\pi)}(f.n)$$

{#eq:nocycle} with $\mathcal{F}(\pi)$ being the set of flaws in π , $r(f)$ being the set of resolvers of f , $f.n$ being the needer of the flaw and $f(\pi)$ being the resulting partial plan after the application of the flaw. Said otherwise, an iteration of LOLLIPPOP won't add cycles inside a partial plan.

The orphan flaw targets steps that have no path to the goal and so can't add new open conditions or threats. The alternative targets existing causal links. Removing a causal link in a plan breaks the full support of the target step. This is why an alternative will always insert a subgoal in the agenda corresponding to the target of the removed causal link. Invalidating side effects also doesn't affect the soundness of the algorithm since the removed flaws are already solved. This makes:

$$\forall f \in \mathcal{F}^-(\pi) : \Downarrow \pi \implies \Downarrow f(\pi)$$

{#eq:conssupport} with $\mathcal{F}^-(\pi)$ being the set of negative flaws in the plan π . This means that negative flaws don't compromise the full support of the plan.

Equation ?? lead to equation ?? being valid after the execution of LOLLIPPOP. The algorithm is sound.

5.2.5.2 Proof of Completeness

The soundness proof shows that LOLLIPPOP's refinements don't affect the support of plans in terms of validity. It was proven that POCL is complete. There are several cases to explore to transpose the property to LOLLIPPOP:

Lemma (Conservation of Validity). *If the input plan is a valid solution, LOLLIPPOP returns a valid solution.*

Proof. With equation ?? and the proof of soundness, the conservation of validity is already proven. \square

Lemma (Reaching Validity with incomplete partial plans). *If the input plan is incomplete, LOLLIPPOP returns a valid solution if it exists.*

Proof. Since POCL is complete and the equation ?? proves the conservation of support by LOLLIPPOP, then the algorithm will return a valid solution if the provided plan is an incomplete plan and the problem is solvable. \square

Lemma (Reaching Validity with empty partial plans). *If the input plan is empty and the problem is solvable, LOLLIPPOP returns a valid solution.*

Proof. This is proven using BEFORE and POCL's completeness. However, we want to add a trivial case to the proof: $\text{pre}(G) = \emptyset$. In this case the line ?? of the algorithm 5 will return a valid plan.

□

Lemma (Reaching Validity with a dead-end partial plan). *If the input plan is in a dead-end, LOLLIPPOP returns a valid solution.*

Proof. Using input plans that can be in an undetermined state is not covered by the original proof. The problem lies in the existing steps in the input plan. Still, using our hypothesis we add a failure mechanism that makes LOLLIPPOP complete. On failure, the needer of the last flaw is deleted if it wasn't added by LOLLIPPOP. User defined steps are deleted until the input plan acts like an empty plan. Each deletion will cause corresponding subgoals to be added to the agenda. In this case, the backtracking is preserved and all possibilities are explored as in POCL. □

Since all cases are covered, this proves the property of completeness.

5.2.5.3 Experimental Results

The experimental results focused on the properties of LOLLIPPOP for online planning. Since classical POCL is unable to perform online planning, we tested our algorithm considering the time taken for solving the problem for the first time. We profiled the algorithm on a benchmark problem containing each of the possible issues described earlier.

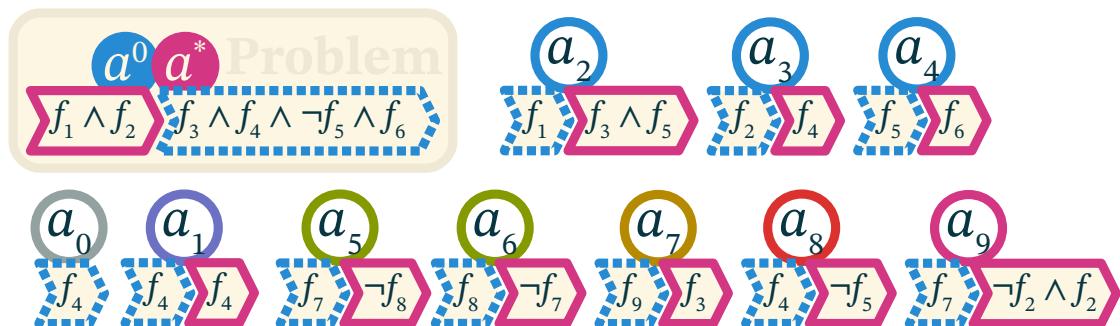


Figure 5.8: Domain used to compute the results. First line is the initial and goal steps along with the useful actions. Second line contains a threatening action t , two co-dependent actions n and l , a useless action u , a toxic action v , a dead-end action w and an inconsistent action x

In figure 5.8, we expose the planning domain used for the experiments. During the domain initialization, the actions u and v are eliminated from the domain since they serve no purpose in the solving process. The action x is stripped of its negative effect because it is inconsistent with the effect f_2 .

As the solving starts, Lollipop computes a safe operator graph (full lines in figure 5.9). As we can see, this partial plan is nearly complete already. When the main refining function starts it receives an agenda with only a few flaws remaining.

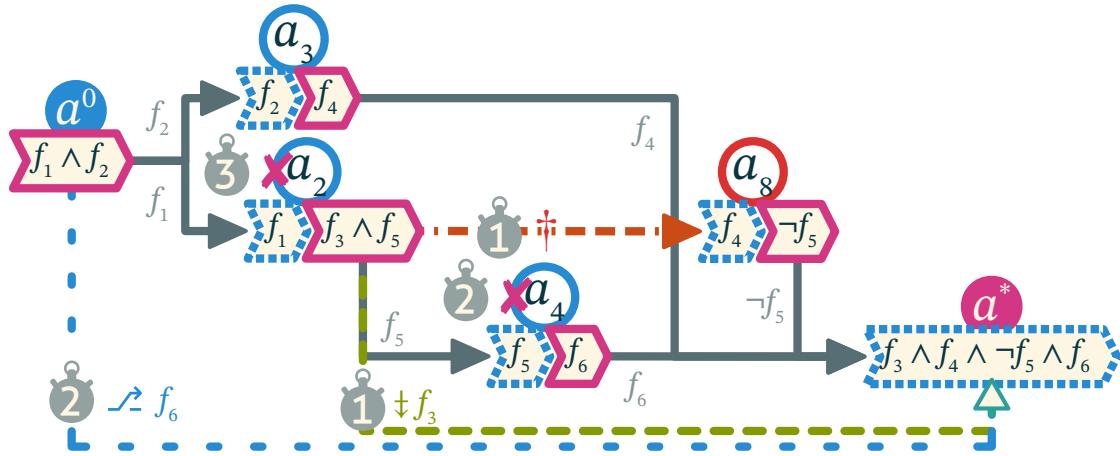


Figure 5.9: In full lines the initial safe operator graph. In thin, sparse and irregularly dotted lines respectively a subgoal, alternative and threat caused causal link.

FIXME: names to indexes

Then the main refinement function starts (time markers 1). LOLLIPOP selects as resolver a causal link from a to satisfy the open condition of the goal step. Once the first threat between a and t is resolved the second threat is invalidated. On a second execution, the domain changes for online planning with 6 added to the initial step. This solving (time markers 2) adds as flaw an alternative on the link from c to the goal step. A subgoal is added that links the initial and goal step for this fluent. An orphan flaw is also added that removes c from the plan. Another solving takes place as the goal step doesn't need 3 as a precondition (time markers 3). This causes the link from a to be cut since it became a liar link. This adds a as an orphan that gets removed from the plan even if it was hanging by the bare link to t .

The measurements exposed in table ?? were made with an Intel® Core™ i7-4720HQ with a 2.60GHz clock. Only one core was used for the solving. The same experiment done only with the chronometer code gave a result of 70ns of errors. We can see an increase of performance in the online runs because of the way they are conducted by LOLLIPOP.

Table 5.2: Average times of 1,000 executions on the problem. The first column is for a simple run on the problem. Second and third columns are times to re-plan with one and two changes done to the domain for online planning.

Experiment	<i>Single</i>	<i>Online 1</i>	<i>Online 2</i>
Time (ms)	0.86937	0.38754	0.48123

TODO: section that explains why LOLLIPOP failed

5.3 HEART

5.3.1 Domain Compilation

In order to simplify the input of the domain, the causes of the causal links in the methods are optional. If omitted, the causes are inferred by unifying the preconditions and effects with the same mechanism as in the subgoal resolution in our POCL algorithm. Since we want to guarantee the validity of abstract plans, we need to ensure that user provided plans are solvable. We use the following formula to compute the final pre-conditions and effects of any composite action a : $\text{pre}(a) = \bigcup_{l \in L^+(I_m)}^{m \in \text{methods}(a)} \text{causes}(l)$ and $\text{eff}(a) = \bigcup_{l \in L^-(G_m)}^{m \in \text{methods}(a)} \text{causes}(l)$. An instance of the classical POCL algorithm is then run on the problem $\mathcal{P}_a = \langle \mathcal{D}, C_{\mathcal{P}}, a \rangle$ to ensure its coherence. The domain compilation fails if POCL cannot be completed. Since our decomposition hierarchy is acyclic ($a \notin A_a$, see definition 51) nested methods cannot contain their parent's action as a step.

5.3.2 Abstraction in POCL

In order to properly introduce the changes made for using HTN domains in POCL, we need to define a few notions.

Transposition is needed to define decomposition.

Definition 50 (Transposition). In order to transpose the causal links of an action a' with the ones of an existing step a in a plan π , we use the following operation:

$$a \triangleright_{\pi}^- a' = \left\{ \phi^-(l) \xrightarrow{\text{causes}(l)} a' : l \in L_{\pi}^-(a) \right\}$$

It is the same with $a' \xrightarrow{\text{causes}(l)} \phi^+(l)$ and L^+ for $a \triangleright^+ a'$. This supposes that the respective preconditions and effects of a and a' are equivalent. When not signed, the transposition is generalized: $a \triangleright a' = a \triangleright^- a' \cup a \triangleright^+ a'$.

Example 32. $a \triangleright^- a'$ gives all incoming links of a with the a replaced by a' .

Definition 51 (Proper Actions). Proper actions are actions that are “contained” within an entity (either a domain, plan or action). We note this notion $A_a = A_a^{\text{lv}(a)}$ for an action a . It can be applied to various concepts:

- For a *domain* or a *problem*, $A_{\mathcal{P}} = A_{\mathcal{D}}$.
- For a *plan*, it is $A_{\pi}^0 = S_{\pi}$.
- For an *action*, it is $A_a^0 = \bigcup_{m \in \text{methods}(a)} S_m$. Recursively: $A_a^n = \bigcup_{b \in A_a^0} A_b^{n-1}$. For atomic actions, $A_a = \emptyset$.

Example 33. The proper actions of *make(drink)* are the actions contained within its methods. The set of extended proper actions adds all proper actions of its single proper composite action *infuse(drink, water, cup)*.

Definition 52 (Abstraction Level). This is a measure of the maximum amount of abstraction an entity can express, defined recursively by:

We use Iverson brackets here, see notations in table ??.

$$lv(x) = \left(\max_{a \in A_x} (lv(a)) + 1 \right) [A_x \neq \emptyset]$$

Example 34. The abstraction level of any atomic action is 0 while it is 2 for the composite action *make(drink)*. The example domain (in listing ??) has an abstraction level of 3.

The most straightforward way to handle abstraction in regular planners is illustrated by Duet (Gerevini *et al.* 2008) by managing hierarchical actions separately from a task insertion planner. We chose to add abstraction in POCL in a manner inspired by the work of Bechon *et al.* (2014) on a planner called HiPOP. The difference between the original HiPOP and our implementation of it is that we focus on the expressivity and the ways flaw selection can be exploited for partial resolution. Our version is lifted at runtime while the original is grounded for optimizations. All mechanisms we have implemented use POCL but with different management of flaws and resolvers. The original algorithm 5 is left untouched.

One of those changes is that resolver selection needs to be altered for subgoals. Indeed, as stated by the authors of HiPOP: the planner must ensure the selection of high-level operators in order to benefit from the hierarchical aspect of the domain. Otherwise, adding operators only increases the branching factor. Composite actions are not usually meant to stay in a finished plan and must be decomposed into atomic steps from one of their methods.

Definition 53 (Decomposition Flaws). They occur when a partial plan contains a non-atomic step. This step is the needer a_n of the flaw. We note its decomposition $a_n \oplus$.

- **Resolvers:** A decomposition flaw is solved with a **decomposition resolver**. The resolver will replace the needer with one of its instantiated methods $m \in \text{methods}(a_n)$ in the plan π . This is done by using transposition such that: $a_n \oplus_m^\pi = \langle S_m \cup (S_\pi \setminus \{a\}), a_n \triangleright^- I_m \cup a_n \triangleright^+ G_m \cup (L_\pi \setminus L_\pi(a_n)) \rangle$.
- **Side effects:** A decomposition flaw can be created by the insertion of a composite action in the plan by any resolver and invalidated by its removal:

$$\bigcup_{\substack{f \in \text{pre}(a_m) \\ a_m \in S_m}} \pi' \nmid_f a_m \quad \bigcup_{\substack{l \in L_{\pi'} \\ a_b \in S_{\pi'}}} a_b \otimes l \quad \bigcup_{\substack{lv(a_c) \neq 0 \\ a_c \in S_m}} a_c \oplus$$

Example 35. When adding the step *make(tea)* in the plan to solve the subgoal that needs tea being made, we also introduce a decomposition flaw that will need this composite step replaced by its method using a decomposition resolver. In order to decompose a composite action into a plan, all existing links are transposed to the initial and goal step of the selected method, while the composite action and its links are removed from the plan.

The main differences between HiPOP and HEART in our implementations are the functions of flaw selection and the handling of the results (one plan for HiPOP and a plan per cycle for HEART). In HiPOP, the flaw selection is made by prioritizing the decomposition flaws. Bechon *et al.* (2014) state that it makes the full resolution faster. However, it also loses opportunities to obtain abstract plans in the process.

5.3.3 Planning in cycle

The main focus of our work is toward obtaining **abstract plans** which are plans that are completed while still containing composite actions. In order to do that the flaw selection function will enforce cycles in the planning process.

Definition 54 (Cycle). A cycle is a planning phase defined as a triplet $c = \langle lv(c), agenda, \pi_{lv(c)} \rangle$ where: $lv(c)$ is the maximum abstraction level allowed for flaw selection in the *agenda* of remaining flaws in partial plan $\pi_{lv(c)}$. The resolvers of subgoals are therefore constrained by the following: $a_p \downarrow_f a_n : lv(a_p) \leq lv(c)$.

During a cycle all decomposition flaws are delayed. Once no more flaws other than decomposition flaws are present in the agenda, the current plan is saved and all remaining decomposition flaws are solved at once before the abstraction level is lowered for the next cycle: $lv(c') = lv(c) - 1$. Each cycle produces a more detailed abstract plan than the one before.

Abstract plans allow the planner to do an approximate form of anytime execution. At any given time the planner is able to return a fully supported plan. Before the first cycle, the plan returned is $\pi_{lv(a_0)}$.

Example 36. In our case using the method of intent recognition of Sohrabi *et al.* Sohrabi *et al.* (2016), we can already use $\pi_{lv(a_0)}$ to find a likely goal explaining an observation (a set of temporally ordered fluents). That can make an early assessment of the probability of each goal of the recognition problem.

For each cycle c , a new plan $\pi_{lv(c)}$ is created as a new method of the root operator a_0 . These intermediary plans are not solutions of the problem, nor do they mean that the problem is solvable. In order to find a solution, the HEART planner needs to reach the final cycle c_0 with an abstraction level $lv(c_0) = 0$. However, these plans can be used to derive meaning from the potential solution of the current problem and give a good approximation of the final result before its completion.

Example 37. In the figure 5.10, we illustrate the way our problem instance is progressively solved. Before the first cycle c_2 , all we have is the root operator and its plan π_3 . Then within the first cycle, we select the composite action *make(tea)* instantiated from the operator *make(drink)* along with its methods. All related flaws are fixed until all that is left in the agenda is the abstract flaws. We save the partial plan π_2 for this cycle and expand *make(tea)* into a copy of the current plan π_1 for the next cycle. The solution of the problem will be stored in π_0 once found.

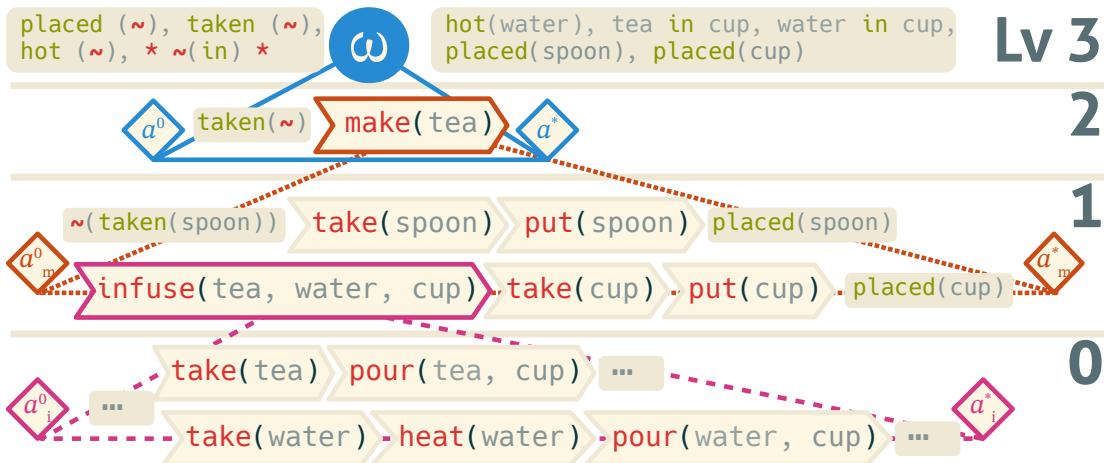


Figure 5.10: Illustration of how the cyclical approach is applied on the example domain. Atomic actions that are copied from a cycle to the next are omitted.

5.3.4 Properties of Abstract Planning

In this section, we prove several properties of our method and resulting plans: HEART is complete, sound and its abstract plans can always be decomposed into a valid solution.

The completeness and soundness of POCL has been proven in (Penberthy *et al.* 1992). An interesting property of POCL algorithms is that flaw selection strategies do not impact these properties. Since the only modification of the algorithm is the extension of the classical flaws with a decomposition flaw, all we need to explore, to update the proofs, is the impact of the new resolver. By definition, the resolvers of decomposition flaws will take into account all flaws introduced by its resolution into the refined plan. It can also revert its application properly.

Lemma (Decomposing preserves acyclicity). *The decomposition of a composite action with a valid method in an acyclic plan will result in an acyclic plan. Formally $\forall a_s \in S_\pi : a_s \not\sim_\pi a_s \implies \forall a'_s \in S_{a \oplus_\pi^m} : a'_s \not\sim_{a \oplus_\pi^m} a'_s$.*

Proof. When decomposing a composite action a with a method m in an existing plan π , we add all steps S_m in the refined plan. Both π and m are guaranteed to be cycle free by definition. We can note that $\forall a_s \in S_m : (\nexists a_t \in S_m : a_s > a_t \wedge \neg f \in \text{eff}(a_t)) \implies f \in \text{eff}(a)$. Said otherwise, if an action a_s can participate a fluent f to the goal step of the method m then it is necessarily present in the effects of a . Since higher level actions are preferred during the resolver selection, no actions in the methods are already used in the plan when the decomposition happens. This can be noted $\exists a \in \pi \implies S_m \cup S_\pi$ meaning that in the graph formed both partial plans m and π cannot contain the same edges therefore their acyclicity is preserved when inserting one into the other. □

Lemma (Solved decomposition flaws cannot reoccur). *The application of a decomposition resolver on a plan π , guarantees that $a \notin S_{\pi'}$ for any partial plan refined from π without reverting the application of the resolver.*

Proof. As stated in the definition of the methods (definition 32): $a \notin A_a$. This means that a cannot be introduced in the plan by its decomposition or the decomposition of its proper actions. Indeed, once a is expanded, the level of the following cycle $c_{lv(a)-1}$ prevents a to be selected by subgoal resolvers. It cannot either be contained in the methods of another action that are selected afterward because otherwise following definition 52 its level would be at least $lv(a) + 1$.

□

Lemma (Decomposing to abstraction level 0 guarantees solvability). *Finding a partial plan that contains only decomposition flaws with actions of abstraction level 1, guarantees a solution to the problem.*

Proof. Any method m of a composite action $a : lv(a) = 1$ is by definition a solution of the problem $\mathcal{P}_a = \langle \mathcal{D}, C_{\mathcal{P}}, a \rangle$. By definition, $a \notin A_a$, and $a \notin A_{a \oplus_{\pi}^m}$ (meaning that a cannot reoccur after being decomposed). It is also given by definition that the instantiation of the action and its methods are coherent regarding variable constraints (everything is instantiated before selection by the resolvers). Since the plan π only has decomposition flaws and all flaws within m are guaranteed to be solvable, and both are guaranteed to be acyclical by the application of any decomposition $a \oplus_{\pi}^m$, the plan is solvable.

□

Lemma (Abstract plans guarantee solvability). *Finding a partial plan π that contains only decomposition flaws, guarantees a solution to the problem.*

Proof. Recursively, if we apply the previous proof on higher level plans we note that decomposing at level 2 guarantees a solution since the method of the composite actions are guaranteed to be solvable.

□

From these proofs, we can derive the property of soundness (from the guarantee that the composite action provides its effects from any methods) and completeness (since if a composite action cannot be used, the planner defaults to using any action of the domain).

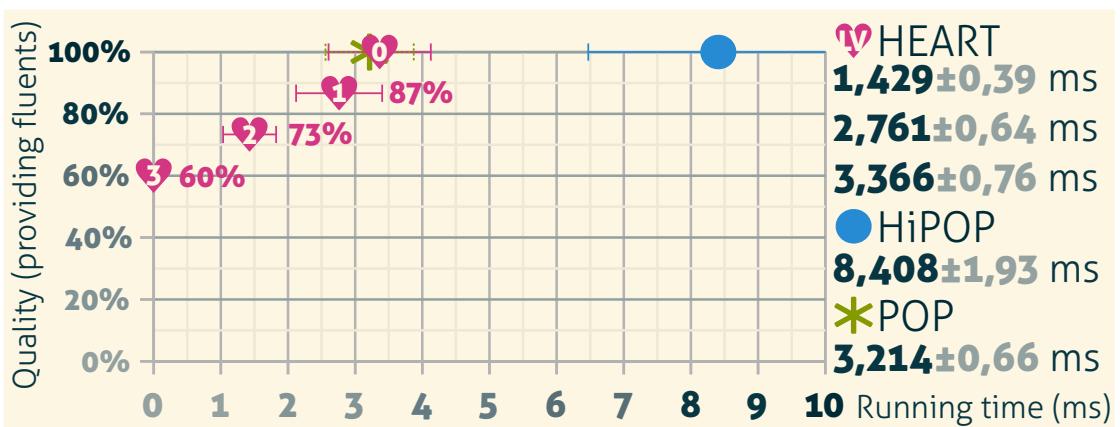


Figure 5.11: Evolution of the quality with computation time.

5.3.5 Computational Profile

In order to assess its capabilities, HEART was evaluated on two criteria: quality and complexity. All tests were executed on an Intel® Core™ i7-7700HQ CPU clocked at 2.80GHz. The Java process used only one core and was not limited by time or memory (32 GB that wasn't entirely used up). Each experiment was repeated between 700 and 10000 times to ensure that variations in speed were not impacting the results.

Figure 5.11 shows how the quality is affected by the abstraction in partial plans. The tests are made using our example domain (see listing ??). The quality is measured by counting the number of providing fluents in the plan $|\bigcup_{a \in S_\pi} \text{eff}(a)|$. This metric is actually used to approximate the probability of a goal given observations in intent recognition ($P(G|O)$ with noisy observations, see (Sohrabi *et al.* 2016)). The percentages are relative to the total number of unique fluents of the complete solution. These results show that in some cases it may be more interesting to plan in a leveled fashion to solve HTN problems. For the first cycle of level 3, the quality of the abstract plan is already of 60%. This is the quality of the exploitation of the plan *before any planning*. With almost three quarters of the final quality and less than half of the complete computation time, the result of the first cycle is a good quality/time compromise.

In the second test, we used generated domains. These domains consist of an action of abstraction level 5. This action has a single method containing a number of actions of levels 4. We call this number the width of the domain. All needed actions are built recursively to form a tree shape. Atomic actions only have single fluent effects. The goal is the effect of the higher level action and the initial state is empty. These domains do not contain negative effects. Figure 5.12 shows the computational profile of HEART for various levels and widths. We note that the behavior of HEART seems to follow an exponential law with the negative exponent of the trend curves seemingly being correlated to the actual width. This means that computing the first cycles has a complexity that is close to being *linear* while computing the last cycles is of the same complexity as classical planning which is at least *P-SPACE* (depending on the expressivity of the domain) (Erol *et al.* 1995).

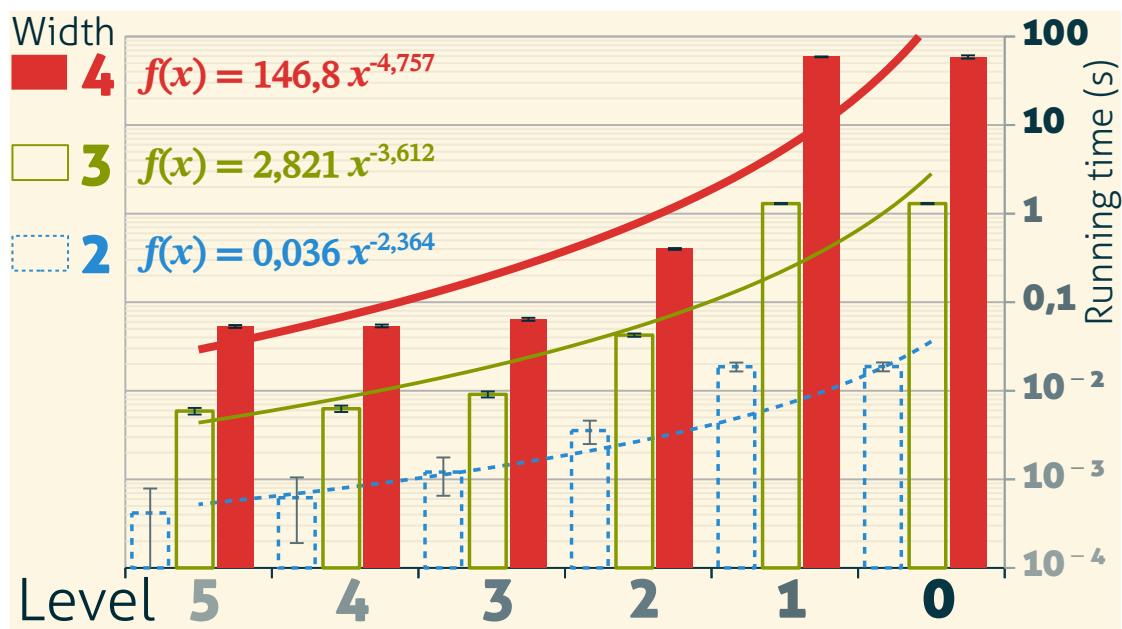


Figure 5.12: Impact of domain shape on the computation time by levels. The scale of the vertical axis is logarithmic. Equations are the definition of the trend curves.

6 Toward Intent Recognition

Since the original goal of this thesis was on intent recognition of dependent persons, we need to explain some more about that specific domain. The problem is to infer the goals of an external agent through only observations without intervention. In the end, the idea is to infer that goal confidently enough to start assisting the other agent without explicit instructions.

6.1 Domain problems

This field was widely studied. Indeed, at the end of the last century, several works started using *abduction* to infer intents from observational data. Of course this comes as a challenge since there is a lot of uncertainty involved. We have no reliable information on the possible goals of the other agent and we don't have the same knowledge of the world. Also, observations can sometimes be incomplete or misleading, the agent can abandon goals or pursue several goals at once while also doing them sub-optimally or even fail at them. To finish, sometimes agents can actively try to hide their intents to the viewer.

6.1.1 Observations and inferences

As explained above, we can only get close to reality and any progress in relevance and detail is exponentially expensive in terms of computing and memory resources. That is why any system will maintain a high degree of abstraction that will cause errors inherent in this approximation.

This noise phenomenon can impact the activity and situation recognition system and therefore seriously impact the intention recognition and decision-making system with an amplification of the error as it is processed. It is also important to remember that this phenomenon of data noise is also present in inhibition and that the lack of perception of an event is as disabling as the perception of the wrong event.

It is possible to protect recognition systems in an appropriate way, but this often implies a restriction on the levels of possibilities offered by the system such as specialized recognition or recognition at a lower level of relevance.

6.1.2 Cognitive inconsistencies

In the field of personal assistance, activity recognition is a crucial element. However, it happens that these events are very difficult to recognize. The data noise mentioned above can easily be confused with an omission on the part of the observed agent. This dilemma is also present during periods of inactivity, the system can start creating events from scratch to fill what it may perceive as inhibition noise.

These problems are accompanied by others related to the behavior of the observed agent. For example, they may perform unnecessary steps, restart ongoing activities or suddenly abandon them. It is added that other aspects of observations can make automated inferences such as ambiguous actions or the agent performing an action that resembles another complicated.

However, some noise problems can be easily detected by simple cognitive processes such as impossible sequences (e. g. closing a closed door). Contextual analyses provide a partial solution to some of these problems.

6.1.3 Sequentiality

Since our recognition is based on a highly temporal planning aspect, we must take into account the classic problems of sequentiality.

A first problem is to determine the end of one plan and the beginning of another. Indeed, it is possible that some transitions between two planes may appear to be a plane in itself and therefore may cause false positives. Another problem is that of intertwined planes. A person can do two things at once, such as answering the phone while cooking. An action in an intertwined plan can then be identified as a discontinuation of activity or a logical inconsistency. A final problem is that of overloaded actions. Not only can an agent perform two tasks simultaneously, but also perform an action that contributes to two activities. These overloaded actions make the process of intention recognition complex because they are close to data noise.

6.2 Existing approaches

The problem of intention recognition has been strongly addressed from many angles. It is therefore not surprising that there are many paradigms in the field. The first studies on the subject highlight the fact that intention recognition problems are problems of abductive logic or graph coverage. Since then, many models have competed in imagination and innovation to improve the field. These include constraint system-based models that provide a solution based on pre-established rules and compiled plan libraries, those that use state or action networks that then launch algorithms on this data, and reverse planning systems.

6.2.1 Constraint

One of the approaches to intention recognition is the one that builds a system around a strong logical constraint. There is often a time constraint system that is complemented by various extensions to cover as many sequential problems as possible.

6.2.1.1 Deductive Approach

In order to solve a problem of intention recognition, abductive logic can be used. Contrary to deductive logic, the goal is to determine the objective from the observed actions. Among the first models introduced is Goldman *et al.* (1999)'s model, which uses the principle of action to construct a logical representation of the problem. This paradigm consists in creating logical rules as if the action in question was actually carried out, but in hypothesizing the predicates that concretize the action and thus being able to browse the research space thus created in order to find all the possible plans containing the observed actions and concretizing defined intentions. This model is strongly based on first-order logic and SWI Prolog logic programming languages. Although revolutionary for the time, this system pale in comparison to recent systems, particularly in terms of prediction accuracy.

6.2.1.2 Algebraic Approach

Some paradigms use algebra to determine possible plans from observed actions. In particular, we find the model of Bouchard *et al.* (2006) which extends the subsumption relationship from domain theory to the description of action and sequence of action in order to introduce it as an order relationship in the context of the construction of a lattice composed of possible plans considering the observed actions. This model simply takes into account the observed actions and selects any plan from the library of plans that contains at least one observed action. Then this paradigm will construct all the possible plans that correspond to the Cartesian product of the observed actions with the actions contained in the selected plans (while respecting their order). This system makes it possible to obtain a subsumption relationship that corresponds to the fact that the plans are more or less general. Unfortunately, this relationship alone does not provide any information on which plan is most likely.

That is why Roy *et al.* (2011) created a probabilistic extension of this model. This uses frequency data from a system learning period to calculate the influence probabilities of each plane in the recognition space. This makes it possible to calculate probabilistic intervals for each plan, action as well as for a plan to know a given action. In order to determine the probability of each plane knowing the upper bound of the lattice (plane containing all observed actions) the sum of the conditional probabilities of the plane for each observed action divided by the number of observed actions is made. This gives a probability interval for each plane allowing the ordinates. This model has the advantage of considering many possible plans but has the disadvantage of seeing a computational explosion as soon as the number of observed actions increases and the context is not taken into account.

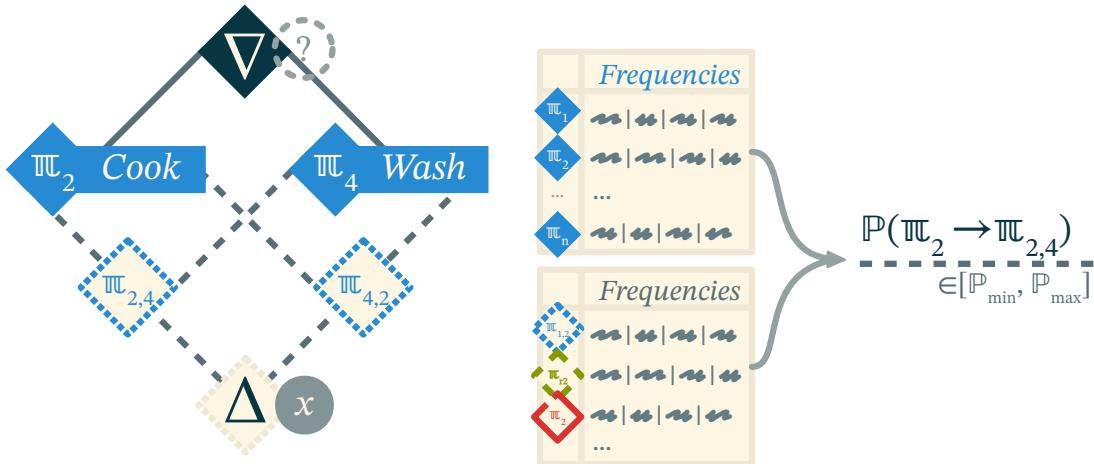


Figure 6.1: The lattice formed by observations (top), matching plans, possible hypothesis and problem (bottom)

6.2.1.3 Grammatical Approach

Another approach is that of grammar. Indeed, we can consider actions as words and sequences as sentences and thus define a system that allows us to recognize shots from incomplete sequences. Vidal *et al.* (2010) has therefore created a system of intention recognition based on grammar. It uses the evaluated grammar system to specify measurements from observations. These measures will make it possible to select specific plans and thus return a hierarchical hypothesis tree with the actions already carried out, the future and the plans from which they are derived. This model is very similar to first order logic-based systems, and uses a SWI Prolog type logic language programming system. Given the scope of maritime surveillance, this model, although taking very well into account the context and the evolution of the measures, is only poorly adapted to an application in assistance, particularly in the absence of a system for discriminating against results plans.

6.2.1.4 Linear programming approach

Another class of approaches is that of diverting standard problem-solving tools to solve the problem of intention recognition. It is therefore possible, by modifying traditional algorithms or by transforming a problem, to ensure that the solution of the tool corresponds to the one sought.

Inoue and Inui (2011) develops the idea of a model that uses linear programming to solve the recognition problem. Indeed, observations are introduced in the form of causes in relation to hypotheses, in a first-order logic predicate system. Each atom is then weighed and introduced into a process of problem transformation by feedback and the introduction of order and causality constraints in order to force the linear program towards optimal solutions by taking into account observations. Although ingenious, this solution does not discriminate between possible plans and is very difficult

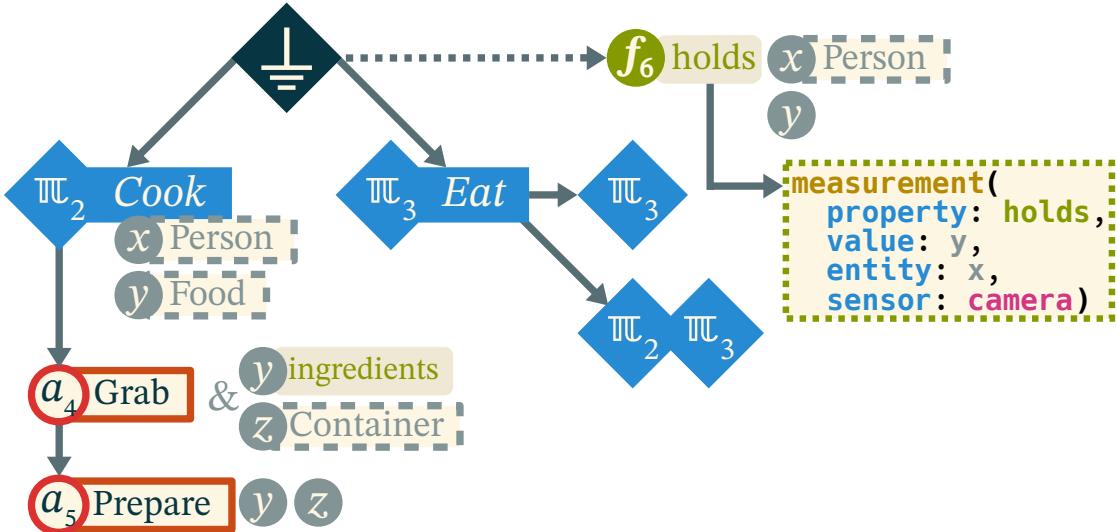


Figure 6.2: The valued grammar used for intent recognition

to apply to real-time recognition situations, mainly because of the problem transformation procedure required each time the problem is updated.

6.2.1.5 Markovian Logic Approach

Another constraint paradigm is the one presented by Raghavana *et al.* using a Markovian extension of first-order logic. The model consists of a library of plans represented in the form of Horn clauses indicating which actions imply which intentions. The aim is therefore to reverse the implications in order to transform the deduction mechanism into an abduction mechanism. Exclusionary constraints and a system of weights acquired through learning must then be introduced to determine the most likely intention. Once again, despite the presence of a system of result discrimination, there is no consideration of context and abductive transformation remains too cumbersome a process for real-time recognition.

6.2.2 Networks

6.2.2.1 And/Or trees approach

As in its early days, intention recognition can still be modeled in the form of graphs. Very often in intention recognition, trees are used to exploit the advantages of acyclicity in resolution and path algorithms. In the prolific literature of Geib *et al.* we find the model at the basis of PHATT (Geib 2002) which consists of an AND/OR tree representing a HTN that contains the intentions as well as their plans or methods. A prior relationship is added to this model and it is through this model that constraints are placed on the execution of actions. Once an action is observed, all the successors of

the action are unlocked as potential next observed action. We can therefore infer by hierarchical path the candidate intentions for the observed sequence.

Since this model does not allow discrimination of results, Geib and Goldman (2005) then adds probabilities to the explanations of the observations. The degree of coverage of each possible goal is used to calculate the probability of each goal. That is, the goal with the plan containing the most observed action and the least unobserved action will be the most likely. This is very ingenious, as the coverage rate is one of the most reliable indicators. However, the model only takes into account temporality and therefore has no contextual support. The representation in the form of a tree also makes it very difficult to be flexible in terms of the plans, which are then fixed a priori.

6.2.2.2 HTN Approach

The HTN model is often used in the field, such as the hierarchical tree form used by Avrahami-Zilberbrand *et al.* (2005). The tree consists of nodes that represent various levels of action and intent. A hierarchical relationship links these elements together to define each intention and its methods. To this tree is added an anteriority relationship that constrains the execution order. This paradigm uses time markers that guarantee order to use an actualization algorithm that also updates a hypothesis tree containing possible intentions for each observation.

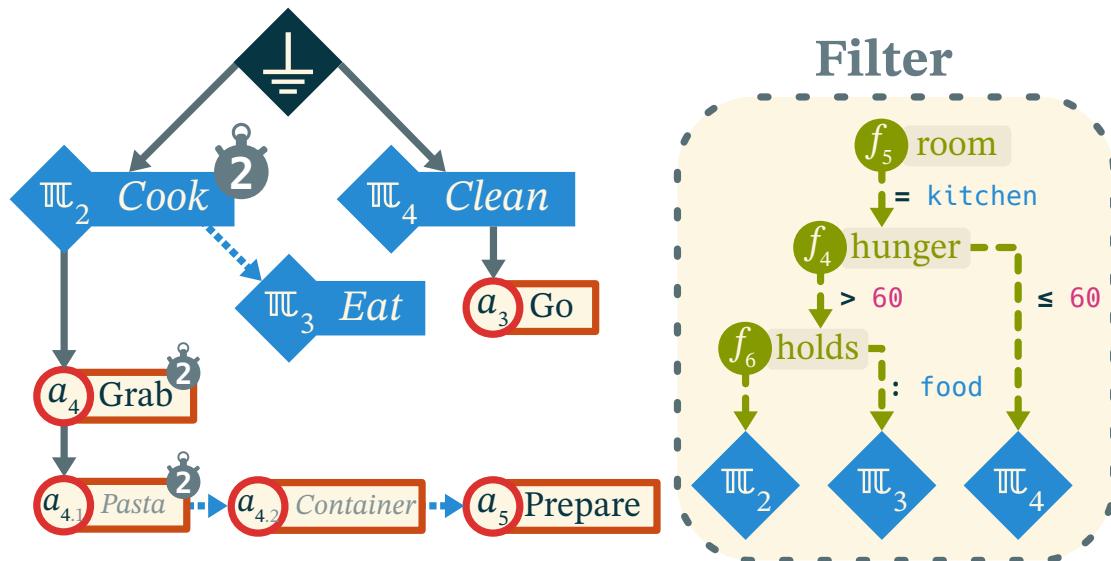


Figure 6.3: The HTN and decision tree used for intent recognition

A probabilistic extension of the Avrahami-Zilberbrand and Kaminka (2006) applies a hierarchical hidden Markov model to the action tree. Using three types of probability that of plan tracking, execution interleaving and interruption, we can calculate the probability of execution of each plan according to the observed sequence. The logic and contextual model filtered on the possible plans upstream leaving us with few calculations to order these plans.

This contextual model uses a decision tree based on a system of world properties. Each property has a finite (and if possible very limited) number of possible values. This allows you to create a tree containing for each node a property and an arc for each value. This is combined with other nodes or leaves that are actions. While running through the tree during execution, the branches that do not correspond to the current value of each property are pruned. Once a leaf is reached, it is stored as a possible action. This considerably reduces the research space but requires a balanced tree that is not too large or restrictive.

6.2.2.3 Hidden Markovian Approach

When we approach stochastic models, we very often find Markovian or Bayesian models. These models use different probabilistic tools ranging from simple probabilistic inference to the fusion of stochastic networks. It can be noted that probabilities are often defined by standard distributions or are isomorphic to weighted systems.

A stochastic model based on THR is the one presented by Blaylock and Allen (2006). This creates hierarchical stacks to categorize abstraction levels from basic actions to high level intentions. By chaining a hidden Markov model to these stacks, the model is able to affect a probability of intention according to the observed action.

6.2.2.4 Bayesian Approach

Another stochastic paradigm is the one of Han and Pereira (2013). It uses Bayesian networks to define relationships between causes, intentions and actions in a given field. Each category is treated separately in order to reduce the search space. The observed actions are then selected from the action network and extracted. The system then uses the intention network to build a temporary Bayesian network using the NoisyOR method. The network created is combined in the same way with the network of causes and makes it possible to have the intention as well as the most probable cause according to the observations.

6.2.2.5 Markovian network approach

The model of Kelley *et al.* (2012) (based on (Hovland *et al.* 1996)) is a model using hidden Markov networks. This stochastic network is built here by learning data from robotic perception systems. The goal is to determine intent using past observations. This model uses the theory of mind by invoking that humans infer the intentions of their peers by using a projection of their own.

Another contextual approach is the one developed for robotics by Hofmann and Williams (2007). The stochastic system is completed by a weighting based on an analysis of vernacular corpuses. We can therefore use the context of an observation to determine the most credible actions using the relational system built with corpus

analysis. This is based on the observation of the objects in the scene and their condition. This makes common sense actions much more likely and almost impossible actions leading to semantic contradictions.

6.2.2.6 Bayesian Theory of Mind

This principle is also used as the basis of the paradigm of Baker and Tenenbaum (2014) which forms a Bayesian theory of the mind. Using a limited representation of the human mind, this model defines formulas for updating beliefs and probabilities a posteriori of world states and actions. This is constructed with sigmoid distributions on the simplex of inferred beliefs. Then the probabilities of desire are calculated in order to recover the most probable intention. This has been validated as being close to the assessment of human candidates on simple intention recognition scenarios.

6.3 Inverted planning

Another way to do intent recognition do not rely on having a plan library at all by using inverted planning. In fact, intent recognition is the opposite problem as planning. In planning we compute the plan from the goal and in intent recognition we seek the goal from the plan. This means that planning is a *deduction* problem while intent recognition is an *abduction* problem. It is therefore possible to transform an intent recognition problem into a planning one.

More intuitively, this transformation relies on the *theory of mind*. This notion of psychology states that one of the easiest ways to predict the mind of another agent is by projecting our own way of thinking onto the target. The familiar way to understand this is to ask the question, "what would I do if I were them ?". This is obviously imperfect since we don't have the complete knowledge of the other mind but is often good enough at basic predictions.

This theory is based on the Belief, Desire and Intention (BDI) model. In our case the belief part is akin to the knowledge database, the desires are the set of possible goals (weighted by costs) and the intent is the plan that achieve a selected goal.

A good analysis of this way of thinking in the context of intent recognition can be found in **CITATION** (Baker)'s work on the subject.

To get further, the work of **CITATION** (Ramirez) is the founding paper on the principle of transforming the intent recognition problem into a planning one. That work was later improved by **CITATION** (Chen) in order to support multiple and concurrent goals at once.

In order to do that Ramirez rediscovers an old tool called constraints encoding into planning **CITATION**. This allows to force the selection of operators in a given a certain order or adding arbitrary constraints on the solution.

The encoding on itself is quite straight forward : Adding artificial fluents to derive an action's behavior considering its selection. In the case of Ramirez's work, the fluents

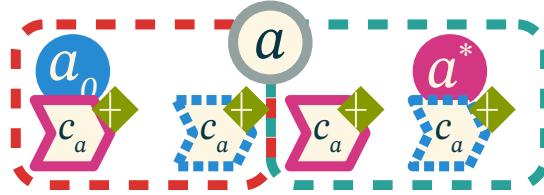


Figure 6.4: Representation of the encoding of constraints using extra fluents.

ensure that the observed actions are selected at the start. The resulting plan is then compared to another plan computed while avoiding the observed action. The difference in cost is proportional to the likelihood of the goal to be pursued.

This problem transformation was more recently improved significantly by **CITATION** (Sohrabi). Indeed, their work allows for using observed fluents instead of actions. This modification allows for a more accurate and flexible prediction with less advanced observations. It also takes into account the missing and noisy observation to affect negatively the likelihood of a goal. Along with the use of diverse planning, this technique allows for seamless multi-goal recognition.

In order to see this technique used in practice we refer to the works of **CITATION** (Tamadupula).

6.3.1 Probabilities and approximations

Now that the intuition is covered, we need to prove that the result of the planning process is indeed correlated to the probability of the agent pursuing that plan knowing the observations. But first we need to formalize how probabilities work.

An *event* is a fixed fluent, a logical proposition that can occur. The likelihood of an event happening ranges from 0 (impossible) to 1 (certain). This is represented by a relation named **probability** of any event e noted $\mathbb{P}(e) \in [0, 1]_{\mathbb{R}}$. This means that probabilities are real numbers restricted between 0 and 1.

Definition 55 (Conditional probabilities). Conditional probabilities are probabilities of an event assuming that another related event happened. This allows to evaluate the ways in which events are affecting one another. The probability of the event A occurring knowing that B occurred is written:

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}$$

with, $\mathbb{P}(A \cap B)$ being the probability that both events occur. In the case of two *independent* events $\mathbb{P}(A|B) = \mathbb{P}(A)$.

We note the set of goals that can be pursued G and the temporal sequence of observations \mathcal{O} . Using conditional probabilities, we seek to have a measure of $\mathbb{P}(g|\mathcal{O})$ for any goal $g \in G$.

In that section we explain how inverted planning does that computation.

For any set of observations \mathcal{O} the probability of the set is the product of the probability of any observation $o \in \mathcal{O}$. We can then note $\mathbb{P}(\mathcal{O}) = \prod_{o \in \mathcal{O}} \mathbb{P}(o)$.

We assume that the observed agent is pursuing one of the known goals. The event of an agent pursuing a specific goal is noted g . This means that $\mathbb{P}(G) = \sum_{g \in G} \mathbb{P}(G) = 1$ because the event is considered certain.

Using *conditional probabilities* we can also note $\mathbb{P}(G|\pi) = 1$ for a valid plan π that achieves any goals $g \in G$.

Theorem 4 (Bayes). *Bayes's theorem allows to find the probability of an event based on prior knowledge of other factors related to said event. It is another basic way to compute conditional probabilities as follows:*

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(B|A)\mathbb{P}(A)}{\mathbb{P}(B)}$$

In the Bayesian logic, one should start with a *prior* probability of an event and actualize it with any new information to make it more precise. In our case, we suppose that $\mathbb{P}(G)$ is given or computed by an external tool.

From the direct application of Bayes's theorem and the previous assumptions, we have :

$$\mathbb{P}(\pi|\mathcal{O}) = \frac{\mathbb{P}(\mathcal{O}|\pi)\mathbb{P}(\pi)}{\mathbb{P}(\mathcal{O})} = \frac{\mathbb{P}(\mathcal{O}|\pi)\mathbb{P}(\pi|G)\mathbb{P}(G)}{\mathbb{P}(\mathcal{O})}$$

{#eq:plan-obs}

Using the event π transitively we can simplify to:

$$\mathbb{P}(G|\mathcal{O}) = \frac{\mathbb{P}(\mathcal{O}|G)\mathbb{P}(G)}{\mathbb{P}(\mathcal{O})}$$

{#eq:goal-obs}

Theorem 5 (Total Probability). *If we have a countable set of disjoint events E we can compute the probability of all events happening as the sum of the probabilities of each event:*

$$\mathbb{P}(E) = \sum_{e \in E} \mathbb{P}(e)$$

Since we consider that we have all likely plans for a given goal we can neglect the very improbable ones and assert that the events of any given plans being acted are independent from one another. Also, the total probability of all plans is certain. From the total probability formula:

$$\mathbb{P}(\mathcal{O}|G) = \sum_{\pi \in \Pi_G} \mathbb{P}(\mathcal{O}|\pi)\mathbb{P}(\pi|G)$$

{#eq:obs-goal}

In equation ??, we have $\mathbb{P}(G)$ and $\mathbb{P}(\mathcal{O})$ known from prior knowledge. Along with equation ??, we can say that:

$$\mathbb{P}(G|\mathcal{O}) = \alpha \mathbb{P}(\mathcal{O}|G) \mathbb{P}(G)$$

With α being a normalizing constant. Also, using the previous formula we can assert that:

$$\mathbb{P}(g|\mathcal{O}) \propto \sum_{\pi \in \Pi_g} \mathbb{P}(\pi|\mathcal{O})$$

This means that if the cost of the plan is related to its likeliness of being pursued knowing the observation sequence, we can evaluate the probability of any goal being pursued. This allows for Sohrabi's problem transformation to work.

That transformation is simply affecting the cost of a plan by dissuading any missed or added fluents while rewarding correct predicted fluents relative to the observations. This process is illustrated in figure 6.5.

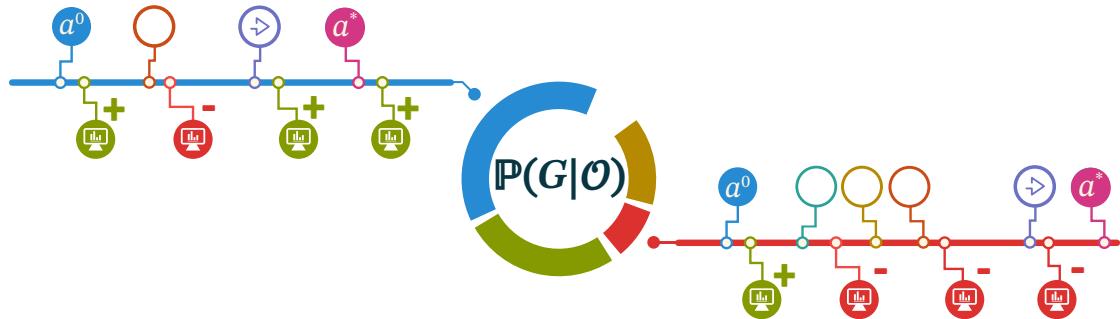


Figure 6.5: Illustration of how plan costs are affecting the resulting probabilities.

6.4 Intent recognition Using Abstract Plans

The initial objective of this thesis was to make intent predictions use abstract plans and repairs. Since plan repair is very susceptible to the heuristic, it is not a reliable tool for general and uncertain planning and will perform worse on all cases not handled well by a given heuristic.

This is not the case of abstract HTN planning since the plan generated is valid while incomplete. The idea here is to quantify the quality of an abstract plan to weigh its plausibility and the likelihood of missing fluents using Sohrabi's method.

The unforeseen problem is that, while in theory this seems like a very efficient approach, in practice it can lack a lot of performance since turning a sequence of observed fluents into a backward chaining heuristic is tricky at best.

In this section, some idea of how that could be done is explored along with perspectives for further works regarding the subject.

6.4.1 Linearization

In order to use the approach of Sohrabi, we need total ordered plans. This is quite an issue since our planner generates partial order plans. Each of these plans have one or several *linearizations*: totally ordered plans that correspond to all possible orders of the plan.

The idea here is to merge parallel actions into one using graph quotient. This is the same mechanism behind sheaves. To do this we use the fact that there are no threats in valid plans and therefore parallel actions have compatible preconditions and effects. This allows to merge several actions into one that is equivalent in terms of fluents and cost.

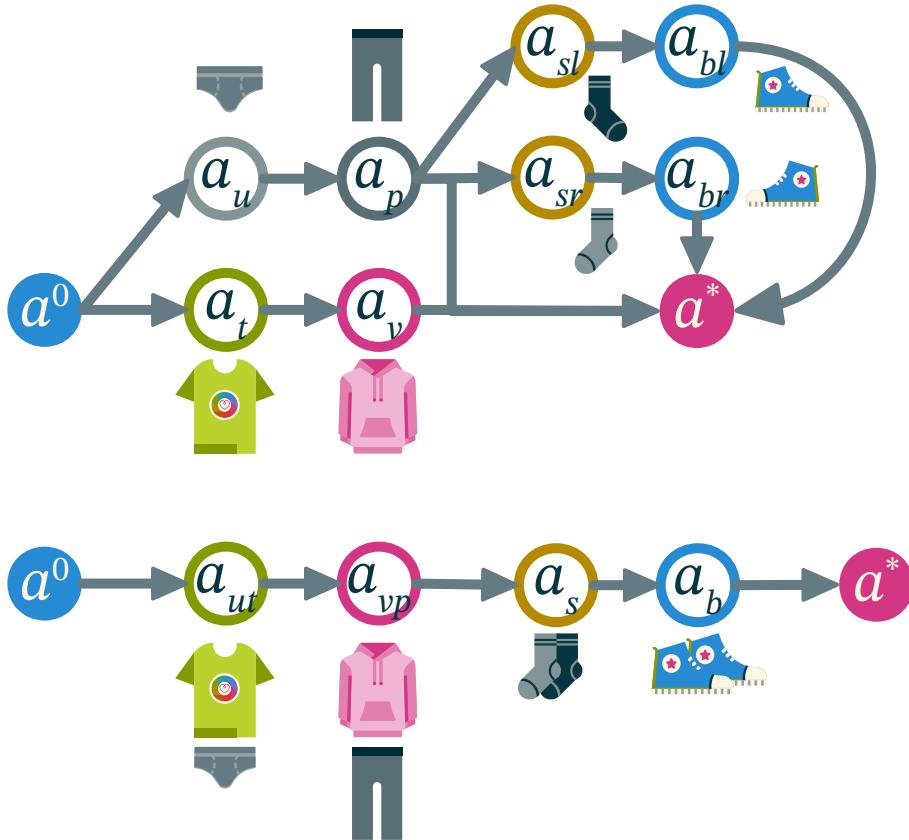


Figure 6.6: Example of linearization of partial order plan.

7 Conclusion and Perspectives

7.1 Fondation

7.1.1 Issues with current fondations

Like stated, these three aspect of knowledge representation makes defining a fondation of mathematics way more difficult than it may appear. In this section, we analyse some of the existing fondation of mathematics and their dependancies.

7.1.1.1 Set theory

From the book releasing the complete formulation of the ZFC theory, we can find that:

*"the discipline underlying ZF will be the **first-order predicate calculus**. The primitive symbols of this set theory, taken from logic, are the connectives, quantifiers and variables mentioned above and, possibly, also the symbol of equality (discussed below), as well as such auxiliary symbols as commas, parentheses and brackets."*

Fraenkel et al.
(1973, 67,22)

This means that ZFC is using FOL as a host language. In turn FOL is using formal grammar as its host language. And at last, formal grammar are defined using set theory. This means that there is a dependancy cycle in the definition of ZFC.

This is significative as this theory is the ground work for most of the classical mathematics still used to this day. Moreover, several assumptions are made in the theory that are not explicitely stated. For example, when naming an element or a set we suppose the principle of identity as we can distinguish them from one another. Also, the notion of set isn't properly defined as well as the classical FOL quantifiers that are used in even the very first formula of the theory.

7.1.1.2 Type theory

Russell and Whitehead (1978)

TODO? Read the book since typed Lambda calculus is posterior to type theory

Type theory -> Typed Lambda calculus (no definition of type) -> Lambda claculus -> set theory or formal grammar

7.1.1.3 Category theory

Awodey (2010)

TODO? Is that what I found ?

Category theory -> Typed Lamda calculus

7.1.1.4 Proof theory

TODO?

Proof Theory -> Hilbert calculus -> Mathematic logic -> (FOL | Set | etc)

7.1.2 Reverse mathematics

Another radical way to approach the fondations of mathematics is called reverse mathematics. In this field, the goal is to find the minimal set of axioms needed to prove any theorem. This is not directly meant to be a fondation of mathematics but helps approaching it by finding groups of axioms that prove all theorems of a given field of mathematics.

This is the approach that should be used to find the most adequate set of axioms while keeping in mind the constraints any fondation of mathematics has.

7.1.3 Possible paradox in proposed theory

TODO: Speak about Russel paradox in fonctional form: A fonction that associates any function that doesn't associate itself. This arrise from the complement operation.

7.2 Knowledge representation

Listing the contributions there are a couple that didn't make the cut. It is mainly ideas or projects that were too long to check or implement and needed more time to complete. SELF is still a prototype, and even if the implementation seemed to perform well on a simple example, no benchmarks have been done on it. It might be good to make a theoretical analysis of OWL compared to SELF along with some benchmark results.

On the theoretical parts there are some works that seems worthy of exposure even if unfinished.

7.2.1 Literal definition using Peano's axioms

The only real exceptions to the axioms and criteria are the first statement, the comments and the liberals.

For the first statement, there is yet to find a way to express both inclusion, the equality relation and solution quantifier. If such a convenient expression exists, then the language can become almost entirely self described.

Comments can be seen as a special kind of container. The difficult part is to find a clever way to differentiate them from regular containers and to ignore their content in the regular grammar. It might be possible to at first describe their structure but then they become parseable entities and fail at their purpose.

Lastly, and perhaps the most complicated violation to fix: literals. It is quite possible to define literals by structure. First we can define boolean logic quite easily in SELF as demonstrated by listing 7.1.

```
1 ~(false) = true;
2 (false, true) :: Boolean;
3 true =?; //conflicts with the first statement!
4 *a : ((a | true) = true);
5 *a : ((false | a) = a);
6 *a : ((a & false) = false);
7 *a : ((true & a) = a);
```

Listing 7.1: Possible definition of boolean logic in SELF.

Starting with line 1, we simply define the negation using the exclusive quantifier. From there we define the boolean type as just the two truth values. And now it gets complicated. We could either arbitrarily say that the false literal is always parameters of the exclusion quantifier or that it comes first on either first two statements but that would just violate minimalism even more. We could use the solution quantifier to define truth but that collides with the first statement definition. There doesn't seem to be a good answer for now.

From line 4 going on, we state the definition of the logical operators \wedge and \vee . The problem with this is that we either need to make a native property for those operators or the inference to compute boolean logic will be terribly inefficient.

We can use Peano's axioms (1889) to define integers in SELF. The attempt at this definition is presented in listing 7.2.

```
1 0 :: Integer;
2 *n : (++(n) :: Integer);
3 (*m, *n) : ((m=n) : (++m = ++n));
4 *n : (++n ~= 0);
5 *n : ((n + 0) = n);
6 (*n, *m) : ((n + ++m) = ++(n + m));
7 *n : ((n * 0) = 0);
8 (*n, *m) : ((n * ++m) = (n + (n * m)));
```

Listing 7.2: Possible integration of the Peano axioms in SELF.

We got several problems doing so. The symbols `*` and `/` are already taken in the default file and so would need replacement or we should use the non-ASCII `×` and `÷` symbols for multiplication and division. Another more fundamental issue is as previously discussed for booleans: the inference would be excruciatingly slow or we should revert to a kind of parsing similar to what we have already under the hood. The last problem is the definition of digits and bases that would quickly become exceedingly complicated and verbose.

For floating numbers this turns out even worse and complicated and such a description wasn't even attempted for now.

The last part concerns textual laterals. The issue is the same as the one with comments but even worse. We get to interpret the content as literal value and that would necessitate a similar system as we already have and wouldn't improve the minimalist aspect of things much. Also we should define ways to escape characters and also to input escape sequences that are often needed in such case. And since SELF isn't meant for programming that can become very verbose and complex.

7.2.2 Advanced Inference

The inference in SELF is very basic. It could be improved a lot more by simply checking the consistency of the database on most aspects. However, such a task seems to be very difficult or very slow. Since that kind of inference is undecidable in SELF, it would be all a research problem just to find a performant inference algorithm.

Another kind of inference is more about convenience. For example, one can erase singlets (containers with a single value) to make the database lighter and easier to maintain and query.

7.2.3 Queries

We haven't really discussed quarries in SELF. They can be made using the main syntax and the solution quantifiers but efficiency of such queries is unknown. Making an efficient query engine is a big research project on its own.

For now a very simplified query API exists in the prototype and seems to perform well but further tests are needed to assess its scalability capacities.

7.3 General Automated Planning

7.4 Planning Improvements

7.4.1 Heuristics using Semantics

7.4.2 Macro-Action learning

8 References

- Alami, R., F. Robert, F. Ingrand, and S. Suzuki
Multi-robot cooperation through incremental plan-merging, *Robotics and Automation, 1995. Proceedings., 1995 IEEE International Conference on*, IEEE, 1995, 3,2573–2579.
- Alessandro, W., and I. Piumarta
OMeta: An object-oriented language for pattern matching, *Proceedings of the 2007 symposium on Dynamic languages*, 2007.
- Aljazzar, H., and S. Leue
K*: A heuristic search algorithm for finding the k shortest paths, *Artificial Intelligence*, 175 (18), 2129–2154, 2011.
- Ambite, J. L., and C. A. Knoblock
Planning by Rewriting: Efficiently Generating High-Quality Plans., DTIC Document, 1997.
- American Heritage Dictionary
Formal (adj.), *American Heritage Dictionary of the English Language*, 2011a.
- American Heritage Dictionary
Circularity (n.d.), *American Heritage Dictionary of the English Language*, 2011b.
- Asimov, I.
The gods themselves, Greenwich, Connecticut: Fawcett Crest, 1973.
- Avrahami-Zilberbrand, D., and G. A. Kaminka
Hybrid symbolic-probabilistic plan recognizer: Initial steps, *Proceedings of AAAI workshop on modeling others from observations (MOO-06)*, 2006.
- Avrahami-Zilberbrand, D., G. A. Kaminka, and H. Zarosim
Fast and complete plan recognition: Allowing for duration, interleaved execution, and lossy observations, *IJCAI Workshop on Modeling Others from Observations*, 2005.
- Awodey, S.
Category theory, 2nd ed. Oxford logic guides 52Oxford ; New York: Oxford University Press, 2010.
- Baader, F.
The description logic handbook: Theory, implementation and applications, Cambridge university press, 2003.
- Babli, M., E. Marzal, and E. Onaindia
On the use of ontologies to extend knowledge in online planning, *KEPS 2018*, 54, 2015.

Backus, J. W.

The syntax and semantics of the proposed international algebraic language of the Zurich ACM-GAMM conference, *Proceedings of the International Conference on Information Processing*, 1959, 1959.

Baker, C. L., R. R. Saxe, and J. B. Tenenbaum

Bayesian theory of mind: Modeling joint belief-desire attribution, *Proceedings of the thirty-second annual conference of the cognitive science society*, 2011, 2469–2474.

Baker, C. L., and J. B. Tenenbaum

Modeling Human Plan Recognition using Bayesian Theory of Mind, 2014.

Bechon, P., M. Barbier, G. Infantes, C. Lesire, and V. Vidal

HiPOP: Hierarchical Partial-Order Planning, *European Starting AI Researcher Symposium*, IOS Press, 2014, 264,51–60.

Becket, R., and Z. Somogyi

DCGs+ memoing= packrat parsing but is it worth it?, *International Symposium on Practical Aspects of Declarative Languages*, Springer, 2008, 182–196.

Beckett, D., and T. Berners-Lee

Turtle - Terse RDF Triple Language, W3C Team Submission W3C, March 2011.

Bercher, P., and D. Höller

Tutorial: An Introduction to Hierarchical Task Network (HTN) Planning 2018.

Bercher, P., S. Keen, and S. Biundo

Hybrid planning heuristics based on task decomposition graphs, *Seventh Annual Symposium on Combinatorial Search*, 2014.

Blaylock, N., and J. Allen

Fast hierarchical goal schema recognition, *Proceedings of the National Conference on Artificial Intelligence*, Menlo Park, CA; Cambridge, MA; London; AAAI Press; MIT Press; 1999, 2006, 21,796.

Borrajo, D.

Multi-agent planning by plan reuse, *Proceedings of the 2013 international conference on Autonomous agents and multi-agent systems*, International Foundation for Autonomous Agents and Multiagent Systems, 2013, 1141–1142.

Borrajo, D., A. Roubíčková, and I. Serina

Progress in Case-Based Planning, *ACM Computing Surveys*, 47 (2), 1–39, January 2015.
doi:[10.1145/2674024](https://doi.org/10.1145/2674024).

Bouchard, B., S. Giroux, and A. Bouzouane

A Smart Home Agent for Plan Recognition of Cognitively-impaired Patients., *Journal of Computers*, 1 (5), 53–62, 2006.

Brenner, M.

A multiagent planning language, *Proc. Of the Workshop on PDDL, ICAPS*, 2003, 3,33–38.

Bürckert, H.-J.

Terminologies and rules, *Workshop on Information Systems and Artificial Intelligence*, Springer, 1994, 44–63.

- Cantor, G.
On a Property of the Class of all Real Algebraic Numbers., *Crelle's Journal for Mathematics*, 77, 258–262, 1874.
- Cantor, G.
Beiträge zur Begründung der transfiniten Mengenlehre, *Mathematische Annalen*, 46 (4), 481–512, 1895.
- Chomsky, N.
Three models for the description of language, *IRE Transactions on information theory*, 2 (3), 113–124, 1956.
- Ciesielski, K.
Set Theory for the Working Mathematician, Cambridge University Press, August 1997.
- Coles, A., A. Coles, M. Fox, and D. Long
Popf2 : A forward-chaining partial order planner, *IPC*, 65, 2011.
- Collins English Dictionary
Abstraction (n.d.), *Collins English Dictionary Complete and Unabridged*, 2014.
- D'Alfonso, D.
Generalized Quantifiers: Logic and Language, 2011.
- Dvorak, F., A. Bit-Monnot, F. Ingrand, and M. Ghallab
A flexible ANML actor and planner in robotics, *Planning and Robotics (PlanRob) Workshop (ICAPS)*, 2014.
- Ephrati, E., and J. S. Rosenschein
Multi-agent planning as the process of merging distributed sub-plans, *In Proceedings of the Twelfth International Workshop on Distributed Artificial Intelligence (DAI-93*, Citeseer, 1993.
- Erol, K., J. A. Hendler, and D. S. Nau
UMCP: A Sound and Complete Procedure for Hierarchical Task-network Planning, *Proceedings of the International Conference on Artificial Intelligence Planning Systems*, University of Chicago, Chicago, Illinois, USA: AAAI Press, June 1994, 2,249–254.
- Erol, K., D. S. Nau, and V. S. Subrahmanian
Complexity, decidability and undecidability results for domain-independent planning, *Artificial intelligence*, 76 (1-2), 75–88, 1995.
- Fikes, R. E., and N. J. Nilsson
STRIPS: A new approach to the application of theorem proving to problem solving, *Artificial intelligence*, 2 (3-4), 189–208, 1971.
- Ford, B.
Packrat parsing:: Simple, powerful, lazy, linear time, functional pearl, *ACM SIGPLAN Notices*, ACM, 2002, 37,36–47.
- Ford, B.
Parsing expression grammars: A recognition-based syntactic foundation, *ACM SIGPLAN Notices*, ACM, 2004, 39,111–122.

- Fox, M.
Natural hierarchical planning using operator decomposition, *European Conference on Planning*, Springer, 1997, 195–207.
- Fox, M., A. Gerevini, D. Long, and I. Serina
Plan Stability: Replanning versus Plan Repair., *ICAPS*, 2006, 6,212–221.
- Fox, M., and D. Long
PDDL+: Modeling continuous time dependent effects, *Proceedings of the 3rd International NASA Workshop on Planning and Scheduling for Space*, 2002, 4,34.
- Fraenkel, A. A., Y. Bar-Hillel, and A. Levy
Foundations of set theory, vol. 67 Elsevier, 1973.
- Geib, C. W.
Problems with intent recognition for elder care, *Proceedings of the AAAI-02 Workshop "Automation as Caregiver*, 2002, 13–17.
- Geib, C. W., and R. P. Goldman
Partial observability and probabilistic plan/goal recognition, *Proceedings of the International workshop on modeling other agents from observations (MOO-05)*, 2005.
- Gerevini, A. E., A. Roubíčková, A. Saetti, and I. Serina
On the plan-library maintenance problem in a case-based planner, *International Conference on Case-Based Reasoning*, Springer, 2013, 119–133.
- Gerevini, A., U. Kuter, D. S. Nau, A. Saetti, and N. Waisbrot
Combining Domain-Independent Planning and HTN Planning: The Duet Planner, *Proceedings of the European Conference on Artificial Intelligence*, 2008, 18,573–577.
- Ghallab, M., C. Knoblock, D. Wilkins, A. Barrett, D. Christianson, et al.
PDDL – The planning domain definition language, 1998.
- Ghallab, M., D. Nau, and P. Traverso
Automated planning: Theory & practice, Elsevier, 2004.
- Ghallab, M., D. Nau, and P. Traverso
Automated Planning and Acting, Cambridge University Press, 2016.
- Goldman, R. P., C. W. Geib, and C. A. Miller
A new model of plan recognition, *Proceedings of the Fifteenth conference on Uncertainty in artificial intelligence*, Morgan Kaufmann Publishers Inc., 1999, 245–254.
- Göbelbecker, M., T. Keller, P. Eyerich, M. Brenner, and B. Nebel
Coming Up With Good Excuses: What to do When no Plan Can be Found, *Proceedings of the International Conference on Automated Planning and Scheduling*, AAAI Press, May 2010, 20,81–88.
- Gradel, E., M. Otto, and E. Rosen
Two-variable logic with counting is decidable, *Logic in Computer Science*, 1997. *LICS'97. Proceedings.*, 12th Annual IEEE Symposium on, IEEE, 1997, 306–317.
- Grünwald, P.
A minimum description length approach to grammar inference, in S. Wermter, E. Riloff, and G. Scheler (eds.), *Connectionist, Statistical and Symbolic Approaches to Learning*

for Natural Language Processing, Lecture Notes in Computer Science; Springer Berlin Heidelberg, 1996, 203–216.

Han, T. A., and L. M. Pereira

Context-dependent incremental decision making scrutinizing the intentions of others via Bayesian network model construction, *Intelligent Decision Technologies*, 7 (4), 293–317, 2013.

Hart, D., and B. Goertzel

Opencog: A software framework for integrative artificial general intelligence, *AGI*, 2008, 468–472.

Hehner, E. C.

A practical theory of programming, Springer Science & Business Media, 2012.

Helmert, M., G. Röger, and E. Karpas

Fast downward stone soup: A baseline for building planner portfolios, *ICAPS 2011 Workshop on Planning and Learning*, Citeseer, 2011, 28–35.

Henglein, F., and U. T. Rasmussen

PEG parsing in less space using progressive tabling and dynamic analysis, *Proceedings of the 2017 ACM SIGPLAN Workshop on Partial Evaluation and Program Manipulation*, ACM, 2017, 35–46.

Hernández, D., A. Hogan, and M. Krötzsch

Reifying RDF: What works well with wikidata?, *SSWS@ ISWC*, 1457, 32–47, 2015.

Hoffmann, J.

FF: The fast-forward planning system, *AI magazine*, 22 (3), 57, 2001.

Hofmann, A. G., and B. C. Williams

Intent Recognition for Human-Robot Interaction., *Interaction Challenges for Intelligent Assistants*, 2007, 60–61.

Horrocks, I., P. F. Patel-Schneider, and F. V. Harmelen

From SHIQ and RDF to OWL: The Making of a Web Ontology Language, *Journal of Web Semantics*, 1, 2003, 2003.

Hovland, G. E., P. Sikka, and B. J. McCarragher

Skill acquisition from human demonstration using a hidden markov model, *Robotics and Automation, 1996. Proceedings., 1996 IEEE International Conference on*, ieee, 1996, 3,2706–2711.

Höschele, M., and A. Zeller

Mining input grammars with AUTOGRAM, *Proceedings of the 39th International Conference on Software Engineering Companion*, IEEE Press, 2017, 31–34.

Hutton, G., and E. Meijer

Monadic parser combinators, 1996.

Inoue, N., and K. Inui

ILP-Based Reasoning for Weighted Abduction., *Plan, Activity, and Intent Recognition*, 2011.

- Kambhampati, S.
 Design Tradeoffs in Partial Order (Plan space) Planning., *AIPS*, 1994, 92–97.
- Kambhampati, S., A. Mali, and B. Srivastava
 Hybrid planning for partially hierarchical domains, *AAAI/IAAI*, 1998, 882–888.
- Kelley, R., A. Tavakkoli, C. King, A. Ambardekar, and M. Nicolescu
 Context-based bayesian intent recognition, *Autonomous Mental Development, IEEE Transactions on*, 4 (3), 215–225, 2012.
- Klyne, G., and J. J. Carroll
Resource Description Framework (RDF): Concepts and Abstract Syntax, Language Specification W3C RecommendationW3C, 2004.
- Korzybski, A.
Science and sanity; an introduction to non-Aristotelian systems and general semantics., Lakeville, Conn.: International Non-Aristotelian Library Pub. Co.; distributed by Institute of General Semantics, 1958.
- Kovacs, D. L.
BNF Description of PDDL 3.1, Unpublished manuscript from the IPC-2011 websiteIPC, 2011.
- Kovács, D. L.
 A multi-agent extension of PDDL3. 1, 2012.
- Kovács, D. L., and T. P. Dobrowiecki
 Converting MA-PDDL to extensive-form games, *Acta Polytechnica Hungarica*, 10 (8), 27–47, 2013.
- Kovitz, B.
 Terminology - What do you call graphs that allow edges to edges?, *Mathematics Stack Exchange*, January 2018.
- Kunen, K.
Set theory an introduction to independence proofs, vol. 102Elsevier, 1980.
- Lindström, P.
 First Order Predicate Logic with Generalized Quantifiers, 1966.
- Loff, B., N. Moreira, and R. Reis
 The Computational Power of Parsing Expression Grammars, *International Conference on Developments in Language Theory*, Springer, 2018, 491–502.
- Luis, N., and D. Borrajo
 Plan merging by reuse for multi-agent planning, *Distributed and Multi-Agent Planning*, 38, 2014.
- McDermott, D.
OPT Manual Version 1.7. 3 (Reflects Opt Version 1.6. 11) DRAFT*, 2005.
- McDermott, D., and D. Dou
 Representing disjunction and quantifiers in RDF, *International Semantic Web Conference*, Springer, 2002, 250–263.

- Motik, B.
On the properties of metamodeling in OWL, *Journal of Logic and Computation*, 17 (4), 617–637, 2007.
- Nau, D. S., T.-C. Au, O. Ilghami, U. Kuter, J. W. Murdock, et al.
SHOP2: An HTN planning system, *J. Artif. Intell. Res.(JAIR)*, 20, 379–404, 2003.
- Nebel, B., and J. Koehler
Plan reuse versus plan generation: A theoretical and empirical analysis, *Artificial Intelligence*, 76 (1), 427–454, 1995.
- Nguyen, X., and S. Kambhampati
Reviving partial order planning, *IJCAI*, 2001, 1,459–464.
- Paulson, L.
A semantics-directed compiler generator, *Proceedings of the 9th ACM SIGPLAN-SIGACT symposium on Principles of programming languages - POPL '82*, Albuquerque, Mexico: ACM Press, 1982, 224–233. doi:[10.1145/582153.582178](https://doi.org/10.1145/582153.582178).
- Peano, G.
Arithmetices principia: Nova methodo exposita, Fratres Bocca, 1889.
- Pednault, E. P.
ADL: Exploring the Middle Ground Between STRIPS and the Situation Calculus., *Kr*, 89, 324–332, 1989.
- Penberthy, J. S., D. S. Weld, and others
UCPOP: A Sound, Complete, Partial Order Planner for ADL, *Kr*, 92, 103–114, 1992.
- Peot, M. A., and D. E. Smith
Threat-removal strategies for partial-order planning, *AAAI*, 1993, 93,492–499.
- Peot, M. A., and D. E. Smith
Postponing Threats in Partial-Order Planning 1994.
- Raghavana, S., P. Singla, and R. J. Mooneya
Plan Recognition using Statistical Relational Models.
- Ramírez, M., and H. Geffner
Plan recognition as planning, *Proceedings of the International Conference on International Conference on Automated Planning and Scheduling*, AAAI Press, 2009, 19,1778–1783.
- RAMOUL, A.
Mixed-initiative planning system to assist the management of complex IT systems 2018.
- Renggli, L., S. Ducasse, T. Gîrba, and O. Nierstrasz
Practical Dynamic Grammars for Dynamic Languages, *Workshop on Dynamic Languages and Applications*, Malaga, Spain, 2010, 4.
- Richter, S., and M. Westphal
The LAMA planner: Guiding cost-based anytime planning with landmarks, *Journal of Artificial Intelligence Research*, 39 (1), 127–177, 2010.

Roy, P. C., A. Bouzouane, S. Giroux, and B. Bouchard
Possibilistic activity recognition in smart homes for cognitively impaired people, *Applied Artificial Intelligence*, 25 (10), 883–926, 2011.

Russell, B.
Mysticism and Logic and Other Essays, London: George Allen and Unwin, 1917.

Russell, B., and A. N. Whitehead
Principia mathematica, Cambridge University Press, 1978.

Sanner, S.
Relational dynamic influence diagram language (rddl): Language description Unpublished ms. Australian National University Unpublished ms. Australian National University, 2010.

Sapena, O., E. Onaindia, and A. Torreno
Combining heuristics to accelerate forward partial-order planning, *CSTPS*, 25, 2014.

Say, B., A. A. Cire, and J. C. Beck
Mathematical programming models for optimizing partial-order plan flexibility, *22nd European Conference of Artificial Intelligence*, 2016.

Schwarzentruber, F.
Hintikka's World: Agents with Higher-order Knowledge, *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence*, Stockholm, Sweden: International Joint Conferences on Artificial Intelligence Organization, July 2018, 5859–5861. doi:[10.24963/ijcai.2018/862](https://doi.org/10.24963/ijcai.2018/862).

Sebastia, L., E. Onaindia, and E. Marzal
A Graph-based Approach for POCL Planning, *ECAI*, 2000, 531–535.

Shekhar, S., and D. Khemani
Learning and Tuning Meta-heuristics in Plan Space Planning, *arXiv preprint arXiv:1601.07483*, 2016.

Silberschatz, A.
Port directed communication, *The Computer Journal*, 24 (1), 78–82, January 1981. doi:[10.1093/comjnl/24.1.78](https://doi.org/10.1093/comjnl/24.1.78).

Sjöberg, J., and H. Nissar
Automated scheduling: Performance in different scenarios, 2015.

Sohrabi, S., A. V. Riabov, and O. Udrea
Plan Recognition as Planning Revisited, *Proceedings of the International Joint Conference on Artificial Intelligence*, Vol. 25, 2016.

Souto, D. C., M. V. Ferro, and M. A. Pardo
Dynamic Programming as Frame for Efficient Parsing, *Proceedings SCCC'98. 18th International Conference of the Chilean Society of Computer Science (Cat. No.98EX212)(SCCC)*, November 1998, 68. doi:[10.1109/SCCC.1998.730784](https://doi.org/10.1109/SCCC.1998.730784).

Sugawara, T.
Reusing Past Plans in Distributed Planning., *ICMAS*, 1995, 360–367.

- Takesaki, M.
Theory of operator algebras II, vol. 125 Springer Science & Business Media, 2013.
- Tan, X., and M. Gruninger
The Complexity of Partial-Order Plan Viability Problems., *ICAPS*, 2014.
- Thiébaux, S., J. Hoffmann, and B. Nebel
In defense of PDDL axioms, *Artificial Intelligence*, 168 (1-2), 38–69, 2005.
- Tolksdorf, R., L. Nixon, F. Liebsch, D. Minh Nguyen, and E. Paslaru Bontas
Semantic web spaces, 2004.
- Unicode Consortium
The Unicode Standard, Version 11.0, Core Specification 11.0 Mountain View, CA, June 2018a.
- Unicode Consortium
Unicode Character Database, *About the Unicode Character Database*, June 2018b.
- Van Der Krog, R., and M. De Weerdt
Plan Repair as an Extension of Planning., *ICAPS*, 2005, 5, 161–170.
- Van Harmelen, F., V. Lifschitz, and B. Porter
Handbook of knowledge representation, vol. 1 Elsevier, 2008.
- Vepstas, L.
Hypergraph edge-to-edge, *Wikipedia*, May 2008.
- Vepštās, L.
Sheaves: A Topological Approach to Big Data, 2008.
- Vidal, N., P. Taillibert, and S. Aknine
Online behavior recognition: A new grammar model linking measurements and intents, *2010 22nd IEEE International Conference on Tools with Artificial Intelligence*, IEEE, 2010, 2, 129–137.
- W3C
Examples for OWL. Eds. P. F. Patel-Schneider and I. Horrocks 2004a.
- W3C
RDF Semantics, W3C, 2004b.
- W3C
RDF Vocabulary Description Language 1.0: RDF Schema February 2004c.
- Younes, H. akan L., and M. L. Littman
PPDDL 1.0: An extension to PDDL for expressing planning domains with probabilistic effects, *Techn. Rep. CMU-CS-04-162*, 2004.
- Younes, H. akan L., and R. G. Simmons
VHPOP : Versatile heuristic partial order planner, *JAIR*, 405–430, 2003.
- Young, R. M., and J. D. Moore
DPOCL: A principled approach to discourse planning, *Proceedings of the Seventh International Workshop on Natural Language Generation*, Association for Computational Linguistics, 1994, 13–20.

Zhuo, H. H., and S. Kambhampati

Model-lite planning: Case-based vs. Model-based approaches, *Artificial Intelligence*, 246, 1–21, 2017.

Apendix

8.1 Chapter 1

SAMIR:

- Cite slow start 1.2
- System : implementation -> Model : theory
- Examples
- theorems or not
- $s : D \rightarrow D$ notation in table
- Double arrow too for \pm
- to port : porter une notion in smth
- 1.2.2 : criterion
- Example of parsing
- 1.2.6.1 : light defeasible logic
- lacks properties and proofs: what are the advantages

Symbol	Name	Arity	Arguments	Type	Definition
=	Equal	2	expr, expr	boolean	$= x : T$
\neq	Not Equal	2	expr, expr	boolean	$\neq x : \perp$
:	Such that	2	element, element	Specification	of Specification
?	Predicate	n	expr	boolean	arbitrary
T	True	0	NA	boolean	
\perp	False	0	NA	boolean	
\neg	Negation	1	boolean	boolean	$\neg T = \perp$
\wedge	Conjunction	2	boolean, boolean	boolean	$a \wedge b \vdash (a = b = T)$
\vee	Disjunction	2	boolean, boolean	boolean	$(a \vee b) \vdash (a = b = \perp)$
\otimes	Logic operator	n	boolean	boolean	arbitrary
\vdash	Entailment	2	boolean, boolean	boolean	
\forall	Universality	1	var	modifier	
\exists	Existentiality	1	var	modifier	
$\exists!$	Unicity	1	var	modifier	
\nexists	Exclusivity	1	var	modifier	$\exists \neg$
\S	Solution	1	var	modifier	
[]	Iverson's brackets	1	boolean int	[\perp] = 0 \wedge [T] = 1	

Symbol	Name	Arity	Arguments	Type	Definition
{}	Set	n	elements	set	
\emptyset	Empty set	0	NA	set	$\emptyset = \{\}$
\in	Member	2	element, boolean	set	
\subset	Subset	2	set	boolean	$\mathcal{S} \subset \mathcal{T} \vdash ((e \in \mathcal{S} \vdash e \in \mathcal{T}) \wedge \mathcal{S} \neq \mathcal{T})$
\cup	Union	n	set	set	$\mathcal{S} \cup \mathcal{T} = \{e : e \in \mathcal{S} \vee e \in \mathcal{T}\}$
\cap	Intersection		set	set	$\mathcal{S} \cap \mathcal{T} = \{e : e \in \mathcal{S} \wedge e \in \mathcal{T}\}$
\setminus	Difference	2	set	set	$\mathcal{S} \setminus \mathcal{T} = \{e : e \in \mathcal{S} \wedge e \notin \mathcal{T}\}$
\times	Cartesian product	n	set	set	$\mathcal{S} \times \mathcal{T} = \{(e_{\mathcal{S}}, e_{\mathcal{T}}) : e_{\mathcal{S}} \in \mathcal{S} \wedge e_{\mathcal{T}} \in \mathcal{T}\}$
\parallel	Cardinal	1	set	int	
\wp	Powerset	1	set	set	axiom of Power set
\circ	Function	n	function	function	
	composition				
σ	Selection	2	predicate	set	$\sigma_i(\mathcal{S}) = \{e : ?(e) \wedge e \in \mathcal{S}\}$
				set	
π	Projection	2	function, set	set	$\pi_f(\mathcal{S}) = \{f(e) : e \in \mathcal{S}\}$
				set	
\mapsto	Substitution		var, func-	expr	$(e \mapsto f(e))(e(e)) = e(f(e))$
			tion,		
			expr		
$\langle \rangle$	Tuple	n	element	tuple	
\rightarrow	Association		element	tuple	
ϕ	Incidence/Adjacency		various	various	
χ	Transitivity	1	relation	relation	
\div	Quotient	2	function, graph	graph	
μ	Meta	1	entity	entity	
ν	Name	1	entity	string	
ρ	Parameter	1	entity	list	
\otimes	Flaws	1	plan	set	
\odot	Resolvers	1	flaw	set	
\downarrow	Partial support	2	link, action		
\downarrow	Full support	2	plan, action		
\prec	Precedance	2	action	boolean	
\succ	Succession	2	action	boolean	
\Rightarrow^*	Shortest path	?			
h	Heuristic	1	element	float	
γ	Constraints	1	element	set	
$\$$	Cost	1	element	float	
d	Duration	1	element	time	

Symbol	Name	Arity	Arguments	Type	Definition
--------	------	-------	-----------	------	------------