



**HAL**  
open science

# Dialogue Modeling in a Dynamic Framework

Maria Boritchev

► **To cite this version:**

Maria Boritchev. Dialogue Modeling in a Dynamic Framework. Computation and Language [cs.CL]. Université de Lorraine; École doctorale IAEM Lorraine - Informatique, Automatique, Électronique - Électrotechnique, Mathématiques de Lorraine, 2021. English. NNT : 2021LORR0199 . tel-03541628

**HAL Id: tel-03541628**

**<https://hal.science/tel-03541628v1>**

Submitted on 24 Jan 2022

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Modélisation dynamique des dialogues

## Dialogue Modeling in a Dynamic Framework

### THÈSE

présentée et soutenue publiquement le 22 novembre 2021

pour l'obtention du

**Doctorat de l'Université de Lorraine**

(mention informatique)

par

Maria Boritchev

#### Composition du jury

<i>Président :</i>	Miguel Couceiro	Université de Lorraine, France
<i>Rapportrice :</i>	Farah Benamara Zitoune	IRIT, Université Paul Sabatier, Toulouse, France
<i>Rapporteur :</i>	Jonathan Ginzburg	Université Paris Diderot – Paris 7, France
<i>Examinatrices :</i>	Ellen Breitholtz	University of Gothenburg, Sweden
	Iris Eshkol-Taravella	Université Paris Nanterre, France
<i>Encadrants :</i>	Maxime Amblard	LORIA, Université de Lorraine, Nancy, France
	Philippe de Groote	LORIA, Inria, Nancy, France

Mis en page avec la classe thesul.

## Acknowledgments

When I think about this thesis, I mostly have in mind work conducted with other researchers. Of course, this is work with both my supervisors: Maxime Amblard, who was there (almost) every week for the last four years; and Philippe de Groote, whom I grew to work with in the past years. It is work with my fellow Ph.D. student colleagues, in particular my scientific half-twin Pierre Ludmann, whose judgment I trust and fear; Clément Beysson-Cabaret, whose friendly and overthinking gaze I felt over my shoulder at every step of the writing, and Laurine Jeannot, my friend and overqualified rubber duck<sup>1</sup>. William Babonnaud, Chuyuan Li, and Siyana Pavlova, we did not (yet?) have much time to work together, but I really hope this time will come.

I wrote Chapter 1 bearing in mind all the discussions with Clément Beysson, Pierre Ludmann, William Babonnaud, Laurine Jeannot, Siyana Pavlova, Laurine Huber and Esteban Marquer. Chapter 2 and Chapter 4 make me very thankful for the (former) Master students I got the opportunity to work with in the past years: Maria Andrea Cruz Blandon, Gosse Minemma, Aria Nourbakhsh, Léa Dieudonnat, Marta Carletti, Yiting Tsai, Axel Didier and Alena Yakavets. Chapter 5 resulted from an intense year of work spent with both my supervisors, after a build-up of three years. Chapter 3 could not have existed without the work of amazing transcriptors, many of whom are already cited above, the main one being Julien Botzanowski. This chapter also owes a lot to Pierre Lefebvre, whose engineering work has given me hope that clean, reproducible research was possible. Chapter 5 starts with a `tikz` picture for which I have to thank Amandine Escalier, who has contributed to most of the `tikz` illustrations in this thesis. The whole thesis contains numerous examples, translated from French to English, for double-checking which I have to thank Sally Lanar. Finally, each time I felt that my motivation was dwindling, I knew I could count on Marie Bouchez and her everlasting thesis enthusiasm.

This seems as a good place to thank Farah Benamara Zitoune and Jonathan Ginzburg for their many helpful and interesting comments and questions. Thank you as well to Iris Eshkol-Taravella and Miguel Couceiro. I also wanted to cite here the researchers with whom I had the chance to talk during the past years, whom I admire and look up to: Ruth Kempson, Ellen Breitholtz, Chris Howes, thank you for your time and your consideration.

It is impossible for me to look back at these years of work without thinking about the great work environment I had the chance to work within. Loria is an incredible place where to conduct one's work, and for me this is in particular thanks to Isabelle Herlich, Floriane Marchand, Isabelle Legrand, Caroline Maillard, and Tarek Bekhouche. *Merci pour votre présence, le labo ne serait vraiment pas le même sans vous.*

These years have been filled with a lot of work, but life still managed to find a way. My first thought goes to the little human beings who have made these four years a little more magical: Viktoria, Roman, Mélusine, Charlotte, Malo, Alashela, Nathanaël, Sofia, and the ones yet to come. Thank you to my oldest friend, Marie Bilu, for always supporting me and for the crocodiles. Thank you to Alice Tourenq and Catherine Nguyen, for all our old traditions, and all the new ones to come. Julien Botzanowski, there is already a lot about you above, *merci pour toute l'heurganizzazione*. Justine Martin, thank you for being here, always. Anastasia Shimorina, thank you for always keeping me grounded! Friends from Lyon, old and new ones, thank you for being there then and being here now. NLP master 2 students of 2017, you were the best possible people to end one's studies with. *Les agents, merci pour tout.*

Parents, brother, thank you for reading my work since kindergarten and up till now. I hope you will enjoy this one! Also, thank you for proofreading these acknowledgments.

*Messages.*

---

<sup>1</sup>[https://en.wikipedia.org/wiki/Rubber\\_duck\\_debugging](https://en.wikipedia.org/wiki/Rubber_duck_debugging)



# Contents

<b>List of Definitions</b>	<b>vii</b>
<b>List of Examples</b>	<b>ix</b>
<b>List of Figures</b>	<b>xiii</b>
<b>List of Tables</b>	<b>xv</b>

<b>Une introduction</b>	<b>1</b>
-------------------------	----------

<b>An introduction</b>	
------------------------	--

<b>Part I State of the art</b>	<b>13</b>
--------------------------------	-----------

<b>Chapter 1</b>	
<b>Formal semantics</b>	<b>15</b>

1.1	From natural language to logic . . . . .	16
1.2	Logic & linguistic phenomena for natural language modeling . . . . .	16
1.2.1	Syntax, semantics, articulations . . . . .	17
1.2.2	Quantification, negation . . . . .	23
1.2.3	Inter- and intra-sentential interactions . . . . .	23
1.2.4	Towards dialogue modeling . . . . .	25
1.3	Frameworks & formalisms for sentence & discourse semantics . . . . .	26
1.3.1	Montague semantics and its descendants . . . . .	26
1.3.2	FOL inspired approaches, anchored in linguistics . . . . .	29
1.4	Discussion and comparison in a dialogue modeling perspective . . . . .	31
1.5	Continuation Style Dynamic Semantics . . . . .	34
1.5.1	Formal set-up . . . . .	35
1.5.2	CSDS – an example . . . . .	36

1.5.3	Abstract Categorical Grammar . . . . .	39
1.6	Quantificational event semantics . . . . .	41
1.6.1	Event semantics – First order logic representations . . . . .	41
1.6.2	Neo-Davidsonian Event Semantics . . . . .	42
1.6.3	Compositional Neo-Davidsonian Event Semantics . . . . .	42
1.7	Inquisitive Semantics . . . . .	46
1.8	Conclusion . . . . .	48

<b>Chapter 2</b>
------------------

<b>Dialogue and formal semantics</b>	<b>51</b>
--------------------------------------	-----------

2.1	Gricean maxims . . . . .	52
2.2	Studies of dialogue . . . . .	53
2.2.1	Speech Act Theory and annotation schema . . . . .	53
2.2.2	Dialogue context & the common ground . . . . .	55
2.2.3	KoS . . . . .	56
2.2.4	SDRT . . . . .	58
2.3	Questions & Answers in dialogue . . . . .	59
2.3.1	Typologies of questions and answers . . . . .	60
2.3.2	Clarification Requests . . . . .	65
2.4	Question semantics . . . . .	65
2.4.1	Questions in inquisitive semantics . . . . .	68
2.5	Semantics of <i>wh</i> -questions . . . . .	69
2.5.1	Semantic representations of <i>wh</i> -questions . . . . .	70
2.6	<i>Wh</i> -words – thematic roles correspondence . . . . .	74
2.6.1	French Interrogatives . . . . .	77
2.7	Conclusion of the chapter . . . . .	78

**Part II Real-life data** **81**

<b>Chapter 3</b>
------------------

<b>Real-life corpora</b>	<b>83</b>
--------------------------	-----------

3.1	Introduction . . . . .	83
3.2	A corpora herbarium . . . . .	85
3.2.1	An overview of corpora of English . . . . .	86
3.2.2	An overview of corpora of French . . . . .	88
3.3	Dialogues in Games (DinG) . . . . .	97

3.3.1	A game of <i>Catan</i> . . . . .	99
3.3.2	Corpus description . . . . .	100
3.3.3	Collection process . . . . .	101
3.3.4	Data processing . . . . .	102
3.4	DinG in the corpora landscape . . . . .	109
3.5	Conclusion of the chapter . . . . .	111

<b>Chapter 4</b>	
<b>Dialogue annotation</b>	<b>115</b>

4.1	A linguistic-based classification . . . . .	116
4.1.1	Classification schema . . . . .	116
4.1.2	Annotation experiments – English, French, Italian, and Mandarin Chinese . . . . .	120
4.1.3	The need for a simpler schema . . . . .	126
4.2	Towards an automation of the classification . . . . .	126
4.2.1	The annotation schemas . . . . .	126
4.2.2	Annotation experiments – English, Dutch, Spanish . . . . .	128
4.2.3	Results and evaluation . . . . .	129
4.3	Annotation experiments on DinG . . . . .	130
4.3.1	Automatic annotation . . . . .	130
4.3.2	Human annotation . . . . .	132
4.3.3	Discussion of chosen examples . . . . .	134
4.3.4	Results of the human annotation . . . . .	137
4.4	Conclusion and future work . . . . .	140

**Part III Semantic processing of dialogue** **141**

<b>Chapter 5</b>	
<b>Semantic processing &amp; dialogue</b>	<b>143</b>

5.1	Negotiation phases . . . . .	145
5.2	Relative response set . . . . .	147
5.2.1	Semantic representation of questions . . . . .	148
5.2.2	The relevant response set . . . . .	150
5.2.3	Answers & the effect on the relative response set . . . . .	151
5.2.4	Model review and conclusions . . . . .	154
5.3	Semantic representation of <i>wh</i> -questions . . . . .	155



5.3.1	Syntax and semantics of <i>wh</i> -interrogatives . . . . .	156
5.3.2	Controlling <i>wh</i> -extraction and quantifier raising . . . . .	161
5.3.3	Discussion and shortcomings . . . . .	163
5.3.4	Model review and conclusions . . . . .	164
5.4	Dynamic interactions and the dialogue context . . . . .	164
5.4.1	Dialogue gameboards . . . . .	165
5.4.2	Formal settings . . . . .	166
5.4.3	Dialogue participant’s epistemic state . . . . .	169
5.4.4	Dialogue dynamics and context updating . . . . .	170
5.4.5	A complete example: Cheryl’s birthday . . . . .	173
5.4.6	Model review and conclusions . . . . .	175
5.5	Conclusion of the chapter . . . . .	176

<b>Conclusion</b>	<b>179</b>
-------------------	------------

<b>Appendixes</b>	<b>183</b>
<b>Appendix A DinG – complete statistical data</b>	<b>185</b>
<b>Appendix B An Abstract Categorical Grammar for <i>wh</i>-questions</b>	<b>191</b>
<b>Index</b>	<b>201</b>
<b>Bibliography</b>	<b>203</b>

# List of Definitions

1	Definition (Syntax of FOL)	17
2	Definition (Basic types, Montague Semantics)	20
3	Definition (Basic types, CSDS)	35
5	Definition (Sentence composition)	35
6	Definition (Context aggregation operator and oracles)	36
7	Definition (Higher order linear signature)	40
8	Definition (Lexicon between two higher order linear signatures)	40
9	Definition (ACG)	40
10	Definition (Basic types, NDES)	43
11	Definition (Basic syntax-semantics interface)	44
12	Definition (Gricean maxims)	53
13	Definition (Dialogue gameboard (DGB), [Ginzburg, 2012])	56
14	Definition (Closed question, [Aarts et al., 2014])	61
15	Definition ( <i>Wh</i> -question, [Ginzburg and Sag, 2000])	61
16	Definition (Alternative question, [Aarts et al., 2014])	61
17	Definition (Simple answer, [Ginzburg, 2012])	62
18	Definition (Aboutness, [Ginzburg, 2012])	63
19	Definition (Strongly exhaustive answer, [Ginzburg, 2012])	63
20	Definition (Potentially resolving answer, [Ginzburg, 2012])	64
21	Definition (Categorial principle, [Groenendijk and Stokhof, 1984])	67
22	Definition (Propositional principle, [Groenendijk and Stokhof, 1984])	67
23	Definition (Imperative epistemic principle, [Groenendijk and Stokhof, 1984])	67
24	Definition (Mention-all <i>wh</i> -questions, [Ciardelli et al., 2018])	69
25	Definition (Mention-some <i>wh</i> -questions, [Ciardelli et al., 2018])	70
26	Definition (Single-match <i>wh</i> -questions, [Ciardelli et al., 2018])	70
27	Definition (Informal definition of the semantic value of a question)	72
28	Definition (Questions in KoS, [Maraev et al., 2018])	72
29	Definition (Inclusive and exclusive disjunctive questions)	116
30	Definition (Action auxiliary questions)	117
31	Definition (Epistemic auxiliary questions)	117
32	Definition (Simple normal form of NDES FOL formulae)	147
33	Definition (Co-predicated variables)	150
34	Definition (Dialogue context, [Boritchev and de Groote, 2020])	170



# List of Examples

1	Example (Un dialogue minimal) . . . . .	2
2	Example (Un deuxième dialogue minimal) . . . . .	3
3	Example (A minimal dialogue) . . . . .	8
4	Example (A second minimal dialogue) . . . . .	9
5	Example (A FOL formula) . . . . .	18
6	Example (A minimal grammar) . . . . .	19
7	Example (A simple syntactic tree) . . . . .	19
8	Example (A discourse domain and interpretation for FOL) . . . . .	19
9	Example (Some semantic representations) . . . . .	20
10	Example (A simple syntactic tree with types) . . . . .	22
11	Example (Some semantic representations, with types) . . . . .	22
12	Example (Computation of the semantic representation of “Brutus stabbed Caesar”) . . . . .	22
13	Example (An anaphora) . . . . .	24
14	Example (A problematic anaphora) . . . . .	24
15	Example (Donkey-sentence) . . . . .	24
17	Example (Semantic representation of a sentence, MS) . . . . .	41
18	Example (Semantic representation of a sentence, Davidsonian event semantics) . . . . .	42
19	Example (Semantic representation of a sentence, NDES) . . . . .	42
20	Example (Deep syntactic tree for the sentence “Brutus stabbed Caesar”) . . . . .	44
21	Example (Deep syntactic tree for the sentence “Brutus stabbed Caesar”, with types) . . . . .	45
22	Example (Computation of the semantic representation of “Brutus stabbed Caesar”) . . . . .	45
23	Example (Ellipsis in dialogue) . . . . .	52
24	Example (A simple dialogue) . . . . .	55
25	Example (Rhetorical question, DinG6) . . . . .	59
26	Example (Answer’s span, SCoSE [Norricks, 2017]) . . . . .	60
27	Example (Simple answer, DinG6) . . . . .	60
28	Example (Closed question, DinG6) . . . . .	61
29	Example ( <i>Wh</i> -question, DinG6) . . . . .	61
30	Example (Alternative question, DinG3) . . . . .	62
31	Example (Simple answer, DinG3) . . . . .	62
32	Example (Aboutness, DinG3) . . . . .	63
33	Example (Aboutness, DinG8) . . . . .	63
34	Example (Strongly exhaustive answer, fabricated) . . . . .	64
35	Example (Potentially resolving answer (self), DinG5) . . . . .	64
36	Example (Question/question, DinG2) . . . . .	64
37	Example (Equivalent interrogatives, fabricated) . . . . .	66

38	Example (Entailed questions, adapted from [Groenendijk and Stokhof, 1984]) . . .	67
39	Example (Not entailed questions, adapted from [Groenendijk and Stokhof, 1984])	67
40	Example (What tea is Charlie drinking?) . . . . .	69
41	Example (Examples of <i>wh</i> -questions, adapted from [Ciardelli et al., 2018]) . . . . .	70
42	Example (Semantic representations of (4a)) . . . . .	70
43	Example (Semantic representations of (4b)) . . . . .	71
44	Example (Semantic representations of (4c)) . . . . .	71
45	Example (Questions in KoS, [Maraev et al., 2018]) . . . . .	73
46	Example (Fragment of conversation with an FEC, [Ginzburg, 2012]) . . . . .	73
48	Example (S-BNC, K62, Walsall Local Studies Centre, interview) . . . . .	86
49	Example (SCoSE part 1, Addie and Brianne) . . . . .	87
50	Example (STAC, s1-league2-game4) . . . . .	87
51	Example (CFPP2000 [02-01] 02-01_Lea_Samvarian_F_55_2e) . . . . .	89
52	Example (CLAPI, Interaction between four friends during an aperitif) . . . . .	90
53	Example (C-ORAL-ROM, ffamcv02) . . . . .	91
54	Example (CRFP, PRI-AUX-1) . . . . .	91
55	Example (TCOF, conversation_mat_08) . . . . .	92
56	Example (TUFS, conversation about differences between China and France) . . . . .	93
57	Example (Rhapsodie, D0001) . . . . .	94
58	Example (FQB_CAF-1, « Qu'est ce qu'une aide au logement ? ») . . . . .	94
59	Example (ESLO1, ESL01_ENT_001_C) . . . . .	95
60	Example (ESLO2, ESL02_REPAS_1247_C) . . . . .	95
61	Example (SLAM) . . . . .	96
62	Example (A dialogue from DinG) . . . . .	97
63	Example (Excerpt from DinG transcription, DinG6) . . . . .	108
64	Example (DinG8, almost standard French) . . . . .	112
65	Example (DinG6, complex utterance construction) . . . . .	112
66	Example (Examples of inclusive and exclusive disjunctive questions) . . . . .	117
67	Example (Completion suggestion, SCoSE) . . . . .	127
68	Example ( <i>Wh</i> -question followed by a 'no', DinG4) . . . . .	131
69	Example (English translation of the example presented in figure 4.5) . . . . .	132
70	Example (English translation of the example presented in figure 4.6) . . . . .	133
71	Example (Request, DinG6) . . . . .	134
72	Example (N/A question, DinG5) . . . . .	134
73	Example (« ou » VS « où », DinG4) . . . . .	135
74	Example (Implicit <i>wh</i> -question, DinG2) . . . . .	135
75	Example (A negative yes/no- or disjunctive question, DinG1) . . . . .	136
76	Example (A disjunctive or <i>wh</i> -question, DinG2) . . . . .	136
77	Example (Utterance transcribed with two '?', DinG2) . . . . .	137
78	Example (Yes/no-question with an embedded disjunction, DinG1) . . . . .	138
79	Example (Negotiation Phase, DinG6) . . . . .	145
80	Example (Minimal negotiation phase, fabricated) . . . . .	147
81	Example (Declarative content of a polar question) . . . . .	148
82	Example (Semantic representation of a polar question) . . . . .	148
83	Example (Semantic representation of a <i>wh</i> -question) . . . . .	149

---

84	Example (Relative response set of <b>Yellow</b> <sub>1001</sub> (example 83)) . . . . .	150
85	Example (Different answers to <b>Yellow</b> <sub>1001</sub> (example 83), fabricated) . . . . .	151
86	Example ( $\mathcal{R}$ corresponding to (9a), example 85) . . . . .	152
87	Example ( $\mathcal{R}$ corresponding to (9b), example 85) . . . . .	153
88	Example ( $\mathcal{R}$ corresponding to (9c), example 85) . . . . .	153
89	Example ( $\mathcal{R}$ corresponding to (9d), example 85) . . . . .	153
90	Example ( $\mathcal{R}$ corresponding to (9e), example 85) . . . . .	154
91	Example (Illustration of the necessity for predicate inferences, fabricated) . . . . .	155



# List of Figures

1.1	Syntactic tree of the sentence “Brutus stabbed Caesar” . . . . .	19
1.2	Genealogical graph of the formal semantics frameworks and formalisms presented in this thesis . . . . .	32
1.3	Schematic view of a sentence in CSDS perspective . . . . .	35
1.4	Schematic view of an ACG . . . . .	40
1.5	Schematic view of ACGs used for NLP . . . . .	41
1.6	Propositions with informative and inquisitive content, projection operators . . . .	48
2.1	Visualization of possible worlds . . . . .	68
2.2	Visualization of the interrogative sentences from example 40 . . . . .	69
3.1	XML code corresponding to PS5B2_102, example 48, S-BNC . . . . .	86
3.2	.trs file corresponding to <b>spk3</b> ’s speech turn, example 52 . . . . .	90
3.3	.trs file corresponding to <b>spk1</b> then <b>spk2</b> ’s speech turns, example 55 . . . . .	92
3.4	FQB, « Qu’est ce qu’une aide au logement ? » . . . . .	94
3.5	.trs file corresponding to example 59 . . . . .	95
3.6	.trs file corresponding to example 60 . . . . .	96
3.7	A photo of the game board during a game of <i>Catan</i> . . . . .	99
3.8	Quantitative data, DinG . . . . .	101
3.9	Recording process – schematic view . . . . .	103
3.10	Data collection and processing pipeline . . . . .	104
3.11	View of ELAN 5.6-FX in segmentation mode, working on DinG2 . . . . .	105
3.12	View of ELAN 5.6-FX in transcription mode, working on DinG2 . . . . .	107
4.1	Statistic distribution of question forms across English, Italian, French and Mandarin Chinese . . . . .	122
4.2	Statistic distribution of answer forms across English, Italian, French and Mandarin Chinese . . . . .	123
4.3	Statistic distribution of question functions across English, Italian, French and Mandarin Chinese . . . . .	124
4.4	Statistic distribution of answer functions across English, Italian, French and Mandarin Chinese . . . . .	125
4.5	Automatic annotation of DinG – <i>wh</i> -question . . . . .	132
4.6	Automatic annotation of DinG – <i>wh</i> -question (erroneous) . . . . .	133
4.7	Distribution of the question annotation tags for DinG, on average, in percentage	138
5.1	Subdivision of dialogue in negotiation phases . . . . .	146
5.2	Syntactic structures of two possible readings for sentence (10) . . . . .	157



*List of Figures*

---

5.3	$\lambda$ -terms corresponding to the syntactic structures presented in figure 5.2 . . . . .	158
5.4	An abstract categorial grammar model of the the syntax-semantics interface . . .	159
5.5	Grammatical architecture of our model . . . . .	162
5.6	Subdivision of dialogue in negotiation phases . . . . .	167

# List of Tables

1.1	Major parts-of-speech (POS)	18
1.2	Major syntactic constituents	18
1.3	Semantic representation of “Brutus”, “stabbed”, “Caesar”	20
1.4	Syntactic categories and corresponding semantic types	21
1.5	Semantic representation of “Brutus stabbed Caesar”	22
1.6	Semantic formalisms and their features	33
1.7	Some commonly used thematic roles with their definitions	43
1.8	Semantic roles defined in ISO 24617-4 – DIT++	50
2.1	DAMSL annotation schema	54
2.2	<i>Wh</i> -words and corresponding queries	61
2.3	Taxonomy of answers	62
2.4	Categories of clarification requests	66
2.5	Commonly used thematic roles	74
2.6	New set of thematic roles	76
2.7	<i>Wh</i> -words and thematic roles pairing	77
2.8	English to French correspondence	78
3.1	UniC – English	84
3.2	UniC – French	84
3.3	English corpora – summary	86
3.4	CÉFC copora – data	88
3.5	DinG data – observations per game, average on whole corpus and coefficients of variation	100
3.6	Interrater agreement for transcription before/after noise tags and pauses removal	106
3.7	Interrater agreement for transcription before/after super-annotation	108
3.8	Quantitative corpora description	109
3.9	Qualitative corpora description	110
3.10	Comparison of the characteristics of (oral) corpora transcriptions	111
4.1	Forms of questions	118
4.2	Forms of answers	118
4.3	Functions of questions	119
4.4	Functions of answers	119
4.5	Forms and functions of questions and answers	120
4.6	Inter-annotator agreement scores for the annotation of the corpus of 90 question-answer pairs from SCoSE	120

4.7	Linguistic cues for question annotation across English, French, Italian and Mandarin Chinese . . . . .	121
4.8	Statistic distribution of question forms across English, Italian, French and Mandarin Chinese . . . . .	121
4.9	Statistic distribution of question forms across English, Italian, French and Mandarin Chinese . . . . .	122
4.10	Statistic distribution of question functions across English, Italian, French and Mandarin Chinese . . . . .	123
4.11	Statistic distribution of answer functions across English, Italian, French and Mandarin Chinese . . . . .	125
4.12	Set of question tags . . . . .	127
4.13	Set of answer tags and corresponding question tags . . . . .	128
4.14	Statistic distribution of question tags across English, Spanish, and Dutch corpora	130
4.15	Statistic distribution of answer tags across English, Spanish, and Dutch corpora .	130
4.16	Annotation results for DinG’s questions annotations . . . . .	137
4.17	Inter-annotator agreement scores for DinG’s questions annotations . . . . .	139
4.18	Inter-annotator agreement scores for the first half of DinG’s questions annotations	139
5.1	Taxonomy of answers . . . . .	151
5.2	List of thematic roles . . . . .	156
5.3	Abstract Syntax Signature . . . . .	159
5.4	Surface Realization Lexicon . . . . .	160
5.5	Semantic Interpretation Lexicon . . . . .	160
5.6	Control Module . . . . .	162
A.1	Quantitative data, DinG corpus, DinG1 and DinG2 . . . . .	186
A.2	Quantitative data, DinG corpus, DinG3, DinG4 and DinG5 . . . . .	187
A.3	Quantitative data, DinG corpus, DinG6, DinG7 and DinG8 . . . . .	188
A.4	Quantitative data, DinG corpus, DinG9 and DinG10, total . . . . .	189

# Une introduction

Les dialogues sont à la base de la communication humaine : les conversations servent à transmettre des informations, à prendre des décisions, à alimenter les relations sociales. La plupart d'entre nous participent quotidiennement à des conversations, avec des personnes différentes, peut-être dans des langues différentes, en discutant certainement de sujets différents. Les dialogues peuvent prendre diverses formes, ils peuvent être oraux ou écrits, il est même possible d'avoir une conversation par le biais de messages vocaux. Pourtant, malgré toute cette diversité, il est possible de dégager au moins deux caractéristiques générales du dialogue :

- il implique deux participants (ou plus <sup>2</sup>) ;
- il implique un échange de sentiments, d'observations, d'opinions ou d'idées<sup>3</sup>.

Ensuite, si nous essayons de restreindre davantage la définition du dialogue, nous rencontrons plusieurs problèmes. Une première caractéristique supplémentaire à laquelle on peut penser est de dire que l'échange entre les deux participants doit se faire en utilisant une *langue (naturelle)* : des langues qui sont utilisées par les humains mais qui n'ont pas été construites par eux. Le traitement automatique des langues (TAL) et la linguistique informatique utilisent ce terme par opposition au *langages (construits)* ou au *langages formels*. L'anglais, le français, le coréen, le yoruba et le pirahã sont des langues naturelles, tandis que l'espéranto, l'elfique, Java, C++ et la logique des prédicats sont des langues construites. Si nous essayons de restreindre la définition du dialogue en ajoutant la condition qu'il doit se dérouler dans une langue naturelle, nous rencontrons des contre-exemples : il est en effet possible d'avoir une conversation en elfique (au moins pour certaines personnes sur certains sujets) et l'espéranto a été spécifiquement conçu comme un moyen de communication international, donc à des fins de dialogue. Une autre question intéressante est soulevée par l'utilisation des *emojis*. Les emojis sont (du moins en partie) des symboles conçus par les êtres humains pour représenter des expressions faciales et/ou des gestes naturels [Gawne and McCulloch, 2019] ; par conséquent, si nous devons définir une langue basée uniquement sur les emojis, il n'est pas clair si cette langue doit être considérée comme naturelle ou construite. L'utilisation des emojis dans les conversations écrites a fait l'objet de plusieurs études ces dernières années (voir par exemple [McCulloch, 2019] pour une discussion approfondie) : est-il possible de mener une conversation en utilisant uniquement des emojis ? quel est le rôle des emojis dans la communication écrite ?

Par ailleurs, les informatiennnes et informaticiens ne parlent pas en utilisant Java et les logiciennnes et logiciens n'utilisent généralement pas la logique des prédicats pour exprimer leurs

---

<sup>2</sup>Le terme *multilogue* se retrouve dans la littérature pour désigner un dialogue entre plus de deux personnes. Nous utiliserons de préférence le terme *dialogue* sans distinction pour une conversation impliquant deux participants ou plus, à moins que le passage de deux participants à plus soit significatif pour les phénomènes discutés.

<sup>3</sup>Source : <https://www.merriam-webster.com/dictionary/conversation>, visité le 04/08/21.

sentiments. Par conséquent, certains langages construits peuvent être utilisés à des fins de conversation, mais ce n'est pas le cas de tous. Lorsqu'il s'agit de langues naturelles, il existe d'autres contraintes : il y a des règles dans le dialogue qui concernent l'ordre dans lequel les énoncés peuvent être utilisés (il n'est généralement pas possible de répondre à une question avant qu'elle n'ait été posée ou de demander de développer un sujet avant qu'il n'ait été évoqué). Plus généralement, des séquences aléatoires de mots dans une langue naturelle donnée constituent rarement des énoncés valides dans une conversation. Ainsi, les questions suivantes se posent :

- Comment la langue peut-elle être utilisée pour transmettre des sentiments, des observations, des opinions ou des idées ?
- Quels sont les mécanismes – linguistiques, cognitifs, mais aussi logiques – qui interviennent dans l'échange et la discussion des différents sujets véhiculés par le dialogue ?
- Quelles sont les règles – linguistiques, cognitives, logiques, temporelles – qui régissent la construction du dialogue ?

Nous nous attelons à aborder ces questions du point de vue de la linguistique informatique : nous étudions les questions linguistiques en utilisant les méthodes et les outils de l'informatique. L'objectif de cette thèse est d'ouvrir la voie à un modèle formel des interactions de dialogue. Ce faisant, nous espérons approfondir la connaissance et la compréhension des phénomènes spécifiques aux interactions de dialogue. Au cours d'une conversation, les êtres humains explorent en collaboration des options et des opinions, clarifient et révisent des suggestions, et essaient finalement d'arriver à un accord ou un désaccord à la suite de négociations. Une façon d'interpréter ces négociations du point de vue de la logique mathématique est de les interpréter comme la négociation de valeurs de vérité qui devraient être le résultat de l'évaluation d'énoncés.

### Exemple 1 (Un dialogue minimal)

**A**<sub>1</sub> Est-ce que Charlie aime le thé ?

**B**<sub>2</sub> Oui

L'exemple 3 montre la négociation minimale d'une déclaration. L'énoncé en cours de discussion est la phrase déclarative « Charlie aime le thé ». **A** interroge la valeur de vérité de cette affirmation, en demandant à **B** son opinion/sa connaissance/sa croyance liée à la véracité de cette affirmation. La réponse de **B** confirme la véracité de l'affirmation, donnant ainsi *vrai* comme valeur de vérité de « Charlie aime le thé ».

Cette thèse est écrite du point de vue de la linguistique computationnelle sur la langue en général et les dialogues en particulier. La linguistique informatique a une riche tradition d'approches logiques à la représentation de la sémantique des énoncés des langues naturelles. Ces études sont regroupées sous le nom des approches de *sémantique formelle*. La modélisation de la sémantique des phrases dans les langues naturelles est une tâche qui peut être abordée de différentes manières. L'approche la plus populaire aujourd'hui provient de la sémantique distributionnelle, voir le manuel [Lappin and Fox, 2015] pour les approches modernes basées sur les vecteurs, dérivant des idées de [Harris, 1954]. L'approche que nous suivons dans cette thèse est appelée *sémantique formelle*, voir [Bos, 2011] pour une étude comparative et un aperçu. La plus connue des multiples approches formelles de la sémantique des phrases déclaratives est celle de Richard Montague [Montague, 1970]. Les approches montagoviennes sont *vériconditionnelles* et *compositionnelles* :

- elles attribuent une *valeur de vérité* à chaque phrase déclarative ;
- elles le font en *composant* les valeurs de vérité des constituants de la phrase.

L'objet de cette thèse étant le dialogue, nous devons rendre compte non seulement des phrases déclaratives, mais aussi des phrases interrogatives, celles-ci étant omniprésentes dans le dialogue. L'exemple 3, bien que minimal, montre que pour pouvoir rendre compte de la négociation de la véracité des énoncés dans le dialogue, il faut être capable d'interpréter la sémantique des phrases interrogatives. Pour ce type de phrase, il n'est pas clair comment les approches conditionnelles de vérité peuvent être utilisées. Que signifierait l'attribution d'une valeur de vérité à une question ? Une question en *oui/non*, telle que **A**<sub>1</sub> dans l'exemple 3, est-elle vraie ou fausse ?

## Motivations

De multiples phénomènes entrent en jeu dans tout dialogue. Prenons l'exemple 4, qui montre un dialogue un peu plus complexe que celui de l'exemple 3.

### Exemple 2 (Un deuxième dialogue minimal)

**C**<sub>1</sub> Charlie est une licorne

**D**<sub>2</sub> Elle préfère le café ou le thé ?

**C**<sub>3</sub> Oui

Pour comprendre le sens de **C**<sub>1</sub>, nous utilisons la propriété de compositionnalité du français : nous connaissons le sens de « Charlie », de « est », de « une », de « licorne ». Nous interprétons la signification de **C**<sub>1</sub> en composant les significations de tous ses constituants. Ensuite, l'énoncé introduit par **C**<sub>1</sub> peut être stocké dans un *contexte dialogique* virtuel, qui contient toutes les informations partagées par toutes les personnes participant au dialogue et qui est constamment mis à jour au fur et à mesure que la conversation se déroule. Une partie de l'information introduite par **C**<sub>1</sub> est que l'individu « Charlie » a la propriété de *être une licorne*. Si Charlie n'a pas été mentionné auparavant, **C**<sub>1</sub> introduit également le besoin d'une accommodation (connaissons-nous une ou un « Charlie » ? est-ce quelqu'un de complètement nouveau ?) du nom propre « Charlie » et une mise à jour des informations relatives à cet individu.

**D**<sub>2</sub> commence par un pronom. Afin d'interpréter la signification de ce constituant de **D**<sub>2</sub>, nous devons accéder au contexte, qui nous donne l'information qu'à ce moment précis, il ne contient qu'un seul individu : « Charlie ». Par conséquent, nous pouvons raisonnablement supposer que, dans le contexte de cette conversation, « Elle », utilisé dans **D**<sub>2</sub>, désigne « Charlie ». Ce type de phénomène d'interaction contextuelle est lié à la *dynamacité* des interactions discursives et dialogiques : au fur et à mesure que les phrases ou les énoncés s'accumulent, les liens et les références entre eux font de même.

**C**<sub>3</sub>, tout comme **B**<sub>2</sub> dans l'exemple 3, confirme la véracité de l'énoncé soulevé par la question posée dans **D**<sub>2</sub>. Le problème vient du fait que la question posée dans **D**<sub>2</sub> est une question disjonctive, elle énonce l'existence de deux alternatives ; la réponse attendue doit choisir une parmi les deux et non valider une valeur de vérité. Cela signifie qu'un modèle de dialogue doit prendre en compte un ensemble de phénomènes liés à la logique, en dehors des approches limitées aux conditions de vérité. L'autre conséquence de l'existence de phénomènes tels que celui mis en évidence dans la séquence **D**<sub>2</sub>-**C**<sub>3</sub> est qu'il apparaît que différents types de questions déclenchent différents types d'interprétations liées à la logique. Par conséquent, un modèle de dialogue doit être ancré dans une classification approfondie des différents types d'énoncés.

## Méthodologie & contributions

Notre objectif est d’esquisser un modèle compositionnel, logique et dynamique du dialogue, ancré dans les études linguistiques des énoncés de dialogue. Afin de poursuivre cet objectif, nous procédons comme suit :

- Nous commençons par restreindre nos études à de courts (sous-)dialogues, organisés autour d’un sujet de discussion unique ; nous appelons ces dialogues les *phases de négociation* [Boritchev and Amblard, 2021].
- Au sein des phases de négociation, nous nous concentrons sur les énoncés centraux permettant la négociation : les questions et les réponses. Nous nous penchons aussi sur les mécanismes qui sous-tendent l’articulation des questions et des réponses.
- La sémantique formelle des phrases déclaratives est un domaine bien établi ; nous l’utilisons comme base pour notre approche de la sémantique formelle des phrases interrogatives [Boritchev and Amblard, 2021].
- Nous menons une étude systématique des questions et des réponses dans les dialogues réels (provenant de corpus) en élaborant et en testant plusieurs schémas d’annotation sur des corpus dans différentes langues (anglais, français, italien, espagnol, néerlandais et chinois mandarin) [Amblard et al., 2019], [Cruz Blandon et al., 2019].
- Comme les dialogues réels présentent une grande variété de phénomènes linguistiques, nous restreignons le champ de notre étude. Pour ce faire, nous constituons le corpus Dialogues in Games (DinG) : un corpus composé de transcriptions de multilogues oraux spontanés entre joueurs francophones du jeu de société *Catane*.
- Les interactions dans DinG sont longues et complexes, ce qui les rend intéressantes pour notre étude des dialogues réels ; elles sont de plus très homogènes : toutes les interactions portent sur le jeu et tous les enregistrements sont très similaires. Par conséquent, nous pouvons utiliser des exemples de DinG et des typologies de questions et de réponses pour adapter notre modèle sémantique formel de dialogue [Boritchev and Amblard, 2020].
- Nous nous concentrons sur un type particulier de questions et produisons une méthode compositionnelle, dynamique et basée sur la logique pour calculer les représentations sémantiques de ces questions.
- Alors que nous façonnons notre modèle de dialogue, nous élargissons notre champ d’action pour modéliser les interactions entre les énoncés et le contexte du dialogue [Boritchev and de Groote, 2020].
- Enfin, en utilisant les études et les discussions développées tout au long de cette thèse, nous pouvons esquisser les principaux repères pour un modèle qui s’adapterait à la logique et aux phénomènes linguistiques rencontrés dans le dialogue à travers une méthode compositionnelle, dynamique et logique ancrée dans la sémantique formelle.

Cette liste résume le contenu de ce manuscrit. Partant d’une présentation de la sémantique formelle, cette thèse développe notre point de vue sur un modèle de dialogue dans un cadre dynamique avec une grande part de théories linguistiques et de travaux expérimentaux. Nous espérons que le résultat atteint un bon équilibre entre l’esquisse d’un modèle formellement cohérent et le respect et la prise en compte des données réelles, qui sont (terriblement) désordonnées et (incroyablement) intéressantes.

---

Nos contributions s’articulent autour de trois axes. La première contribution de cette thèse est la **production du corpus DinG** : en commençant par les grandes lignes du projet et tout au long du processus d’enregistrement, de transcription, d’annotation et d’analyse. Pendant que nous construisions DinG, nous avons également travaillé sur d’autres corpus du dialogue dans différentes langues, afin d’élargir notre connaissance des questions et des réponses dans le dialogue. La deuxième contribution de cette thèse est le **développement et test de plusieurs schémas d’annotation crosslingues pour les questions et les réponses dans le dialogue**. Une fois que nos études sur le dialogue réel ont commencé à progresser, nous avons pu également avancer dans nos approches de sémantique formelle du dialogue. La troisième contribution de cette thèse réside dans la création et l’étude de plusieurs modèles de sémantique formelle du dialogue dans un cadre dynamique. L’examen de ces modèles nous guide vers l’esquisse de repères majeurs pour un modèle de dialogue réel basé sur la sémantique formelle.

## Applications

Bien que notre travail ne débouche pas directement sur un outil clé en main, nous pensons néanmoins que les idées que nous développons tout au long de cette thèse peuvent être très utiles dans diverses applications qui vont des tâches classiques du TAL à la recherche en psychologie et en cognition.

Notre étude des questions et des réponses dans le dialogue nous conduit au développement de modèles qui peuvent être utilisés dans des tâches de questions/réponses en recherche d’information. Des approches hybrides, combinant des techniques d’apprentissage automatique et des représentations logiques telles que celles que nous produisons dans cette thèse, ont été utilisées avec succès pour améliorer les résultats de telles tâches, voir par exemple [Demner-Fushman and Lin, 2007] pour des réponses à des questions dans un contexte médical.

Notre objet d’étude étant le dialogue, une partie de notre travail est orientée vers l’étude de l’interaction humaine. De nos jours, le TAL est utilisé afin d’aider à l’étude et au diagnostic de maladies cognitives et psychiatriques telles que la dépression [Howes et al., 2014], la maladie d’Alzheimer [Nasreen et al., 2019], et la schizophrénie [Breitholtz et al., 2021]. L’une des motivations de l’approche par les données réelles dans notre travail provient de projets étudiant les interactions impliquant des patients atteints de schizophrénie, du point de vue de la sémantique computationnelle, tels que le projet *Dialogical Reasoning in Patients with Schizophrenia (DRiPS<sup>4</sup>)* et *Schizophrénie et Langage : Analyse and Modélisation (SLAM<sup>5</sup>)*. Les corpus utilisés dans ces projets sont difficiles à produire, à stocker et à utiliser, car ils concernent des patients atteints de schizophrénie, ce qui met en jeu le secret médical. Pour la même raison, ces corpus sont assez petits, de sorte que les modèles de TAL qui nécessitent beaucoup de données d’entraînement ne peuvent pas être facilement utilisés dans ces contextes.

Nous espérons que le travail que nous développons dans cette thèse pourra être utile au moins de deux manières différentes dans le cadre de ces études. Tout d’abord, le corpus que nous avons développé est très similaire dans sa constitution au corpus construit dans le cadre du projet appelé *Outils informatisés d’aide au Diagnostic des Maladies mentales (ODiM<sup>6</sup>)*. Le corpus construit pour ODiM contient des données très sensibles, tandis que le corpus DinG est distribué librement sous la licence Creative Commons Attribution ShareAlike (CC BY-SA 4.0). Il est disponible sur Gitlab : <https://gitlab.inria.fr/semagramme-public-projects/>

---

<sup>4</sup><https://www.christinehowes.com/research/drips>, visité le 16/07/21.

<sup>5</sup><https://team.inria.fr/semagramme/slam/>, visité le 16/07/21.

<sup>6</sup><https://team.inria.fr/semagramme/odim/>, visité le 16/07/21.



[resources/ding/](#). Par conséquent, les outils développés dans le cadre d'ODiM peuvent être testés et mis en valeur sur DinG, du moins en ce qui concerne les questions et les réponses dans le dialogue.

Ensuite, comme nous l'avons évoqué plus haut, il est difficile de produire des modèles de TAL pour des corpus impliquant des données personnelles, en particulier celles liées à la médecine, car ces corpus sont souvent trop petits par rapport à la quantité de données dont les réseaux neuronaux ont besoin pour être performants [Liu and Chen, 2019]. Nous espérons que les modèles de dialogue basés sur la logique que nous développons dans cette thèse pourront compenser le manque de données en ajoutant des sources d'information supplémentaires et en permettant l'utilisation de modèles hybrides du TAL.

## Plan de la thèse

Cette thèse est organisée en trois parties. La première partie présente les préliminaires linguistiques et logiques dont nous avons besoin pour définir précisément la portée et le but de notre travail. Le chapitre 1 porte sur la sémantique formelle du discours. Il contient les définitions des phénomènes linguistiques et logiques que nous considérons tout au long de cette thèse. Sur la base de ces définitions, nous donnons une présentation des principaux formalismes et cadres de la sémantique formelle du discours. Ensuite, nous les comparons par rapport aux phénomènes présentés et dans la perspective de la construction d'un modèle de sémantique formelle du dialogue dans un cadre dynamique. Le chapitre 2 se focalise sur le dialogue plutôt que le discours. Il présente les études existantes, en particulier formelles, du dialogue avant de plonger dans les définitions et les typologies des questions et des réponses. Ensuite, nous présentons les exigences pour un modèle formel des phrases interrogatives et donnons un compte rendu détaillé de la sémantique des questions. Nous terminons ce chapitre en nous concentrant sur les questions en *wh*-.

La deuxième partie de cette thèse se concentre sur les données réelles. Le chapitre 3 détaille la construction du corpus DinG avant de présenter le paysage des corpus existants de dialogues réels en anglais et en français, et de montrer comment DinG s'y intègre. Le chapitre 4 se concentre sur l'annotation des questions et des réponses dans les dialogues réels et présente plusieurs schémas d'annotation ainsi que des expériences interlinguistiques. Nous concluons le chapitre 4 en appliquant un schéma de classification de questions aux questions du corpus DinG.

La dernière partie de cette thèse présente nos recherches sur notre question principale : la construction d'un modèle formel du dialogue dans un cadre dynamique. Le chapitre 5 présente les différentes solutions que nous avons explorées pour la modélisation du dialogue, en utilisant la subdivision en phases de négociation du dialogue. Nous introduisons un outil mathématique qui rend compte du lien entre une question et sa réponse ; puis, nous présentons une méthode systématique et logique de calcul compositionnel des représentations sémantiques des questions en *wh*-. Enfin, nous détaillons les interactions entre les énoncés et le contexte du dialogue.

# An introduction

Dialogues are at the basis of human communication – conversations are used to convey information, to make decisions, to feed social relationships. Most of us take part in conversations daily, with different people, perhaps in different languages, definitely discussing different subjects. Dialogues can take various forms, they can be oral or written, it is even possible to have a conversation through voice notes. Yet, through all this variety, it is possible to outline at least two general characteristics of dialogue:

- It involves two participants (or more<sup>7</sup>).
- It involves an exchange of sentiments, observations, opinions, or ideas<sup>8</sup>.

Then, if we try to narrow the definition of dialogue further, we come across several problems. A first additional characteristic that can be thought of is to say that the exchange between the two participants should be made using *natural languages*: languages that are used by humans but have not been constructed by them. Natural Language Processing (NLP) and computational linguistics use this term as opposed to *constructed* or *formal languages*. English, French, Korean, Yoruba, and Pirahã are natural languages, while Esperanto, Elvish, Java, C++, and predicate logic are constructed languages. If we try to restrict the definition of dialogue by adding the condition that it should happen through a natural language, we encounter counter-examples: it is in fact possible to have a conversation in Elvish (at least for some people on some subjects) and Esperanto was specifically designed as an international mean of communication, therefore for dialogue purposes. Another interesting question is raised by the use of *emojis*. Emojis are (at least partly) human-designed symbols that represent natural facial expressions and/or gestures [Gawne and McCulloch, 2019]; therefore, if we were to define a language based on emojis only, it is not clear whether that language should be considered as a natural one or a constructed one. The use of emojis in written conversations has been the object of several studies in the past years (see for example [McCulloch, 2019] for a thorough discussion): is it possible to conduct a conversation using emojis only? what is the role of emojis in written communication?

On the other hand, computer scientists do not talk using Java and logicians generally do not use predicate logic to convey sentiments. Therefore, some constructed languages can be used for conversation purposes. When it comes to natural languages, there are other constraints. There are rules in dialogue when it comes to the order in which the utterances can be used (it is generally not possible to answer a question before it has been asked or to ask for a follow-up on a subject before it has been brought up). More generally, random sequences of words in a given natural language rarely constitute valid utterances in a conversation. Thus, the following questions arise:

---

<sup>7</sup>The term *multilogue* can be found in the literature to designate a dialogue between more than two people. We will preferably use the term *dialogue* without distinction for a conversation involving two participants or more, unless the shift from two participants to more is significant for the phenomena under discussion.

<sup>8</sup>Source: <https://www.merriam-webster.com/dictionary/conversation>, visited on 04/08/21.

- How can language be used to convey sentiments, observations, opinions, or ideas?
- What are the mechanisms – linguistic, cognitive, but also logic ones – that are involved in the exchange and discussion of various subjects conveyed by dialogue?
- What are the rules – linguistic, cognitive, logic, temporal ones – for dialogue construction?

We attempt to address these questions from the perspective of computational linguistics: we study linguistic questions using methods and tools of computer science. The aim of this thesis is to pave the way towards a formal model of dialogue interactions. In doing so, we hope to deepen the knowledge and understanding of phenomena that are specific to dialogue interactions. During a conversation, humans collaboratively explore options and opinions, clarify and revise suggestions, and ultimately try to come to an agreement or disagreement as a result of negotiations. A way to interpret these negotiations from a mathematical logic point of view is to interpret them as the negotiation of truth values that should be the result of the evaluation of statements.

### Example 3 (A minimal dialogue)

**A**<sub>1</sub> Does Charlie like tea?

**B**<sub>2</sub> Yes

Example 3 shows the minimal negotiation of a statement. The statement under discussion is “Charlie likes tea”. **A** interrogates the truth value of this statement, asking **B** for their opinion/knowledge/belief related to its truthfulness. **B**’s answer confirms the truthfulness of the statement, therefore setting the truth value of “Charlie likes tea” to **true**.

This thesis is written from the computational linguistics point of view on language in general and dialogues in particular. Computational linguistics has a rich tradition of logical approaches to the meaning of natural language sentences. Modeling the meaning of sentences in natural languages is a task that can be approached from different perspectives. The most popular approach nowadays comes from distributional semantics, see the [Lappin and Fox, 2015] handbook for the modern vector-based approaches, deriving from ideas of [Harris, 1954]. The approach we follow in this thesis is called *formal semantics*, see [Bos, 2011] for a survey and overview. The most well-known among the multiple formal accounts of declarative sentences is Richard Montague’s [Montague, 1970]. Montagovian approaches are *truth-conditional* and *compositional*:

- they assign a *truth value* to each declarative sentence;
- they do so by *composing* the sentence’s constituents truth values.

As the object of this thesis is dialogue, we need to be able to account for more than just declarative sentences, as interrogative sentences are ubiquitous in dialogue. Example 3, though minimal, shows that in order to be able to account for the negotiation of the truthfulness of statements in dialogue, we need to be able to interpret the semantics of interrogative sentences. For this type of sentence, it is not clear how truth-conditional approaches can be used. What assigning a truth value to a question would mean? Is a *yes/no*-question, such as **A**<sub>1</sub> in example 3, true or false?

---

## Motivations

Multiple phenomena come into play in any dialogue. Consider example 4, which shows a dialogue that is just a little bit more complex than the one in example 3.

### Example 4 (A second minimal dialogue)

$C_1$  Charlie is a unicorn

$D_2$  She prefers coffee or tea?

$C_3$  Yes

To understand the meaning of  $C_1$ , we use the compositionality property of English: we know the meaning of “Charlie”, of “is”, of “a”, of “unicorn”. We interpret the meaning of  $C_1$  by composing the meanings of all its constituents. Then, the statement introduced by  $C_1$  can be stored in a virtual *context of the conversation*, which contains all the information shared by the dialogue participants and is constantly being updated as the conversation unravels. Part of the information introduced by  $C_1$  is that the individual “Charlie” has the property of *being a unicorn*. If Charlie has not been mentioned before,  $C_1$  also introduces the need for an accommodation (do we know a “Charlie”? is it someone completely new?) of the proper noun “Charlie” and an update of the information relative to this individual.

$D_2$  starts with a pronoun. In order to interpret the meaning of this constituent of  $D_2$ , we need to access the context, that gives us the information that for now, it only contains one individual: “Charlie”. Therefore, we can reasonably assume that, in the context of this conversation, “She”, used in  $D_2$ , designates “Charlie”. This type of context-interacting phenomena is related to the *dynamicity* of discourse and dialogue interactions: as the sentences or utterances pile up, so do links and references between them.

$C_3$ , similarly to  $B_2$  in example 3, confirms the truthfulness of the statement raised by the question asked in  $D_2$ . The problem comes from the fact that the question asked in  $D_2$  is a disjunctive question, it states the existence of two alternatives; the expected answer should choose one among the two and not validate a truth content. This means that a model of dialogue should take into account a range of logic-related phenomena outside approaches restricted to truth conditions. The other consequence of the existence of phenomena such as the one showcased in the sequence  $D_2$ - $C_3$  is that it appears that different types of questions trigger different types of logic-related interpretations. Therefore, a model of dialogue needs to be rooted in a thorough classification of different types of utterances.

## Methodology & contributions

Our goal is to outline a compositional, logic-based, and dynamic model of dialogue rooted in linguistic studies of dialogue utterances. In order to achieve this goal, we proceed as follows:

- We start by restraining our studies to short (sub-)dialogues, organised around a unique subject of discussion; we call them *negotiation phases* [Boritchev and Amblard, 2021].
- Inside negotiation phases, we focus on the core utterances allowing the negotiation: questions and answers, along with the mechanisms behind the articulation of questions and answers.

- Formal semantics of declarative sentences is a well-established field; we use it as a base for our approach to formal semantic of interrogative sentences [Boritchev and Amblard, 2019].
- We conduct a systematic study of questions and answers in real-life dialogues by elaborating and testing several annotation schemas on corpora in different languages (English, French, Italian, Spanish, Dutch, and Mandarin Chinese) [Amblard et al., 2019], [Cruz Blandon et al., 2019].
- As the real-life dialogues present a great variety of linguistic phenomena, we restrain the scope of our study. To do so, we constitute the Dialogues in Games (DinG) corpus: a corpus composed of transcriptions of real-life, oral, spontaneous multi-party dialogues between French-speaking players of the board game *Catan*.
- Interactions in DinG are long and complex, which makes them interesting for our study of real-life dialogues, but they are also very homogeneous: all the interactions are about the game and all the recordings are very similar. Therefore, we can use examples from DinG and typologies of questions and answers to tailor our formal semantics model of dialogue [Boritchev and Amblard, 2020].
- We focus on a particular type of questions and produce a compositional, dynamic, and logic-based method for computing the semantic representations of these questions.
- As we are shaping our model of dialogue, we expand our focus to model the interactions between utterances and the dialogue context [Boritchev and de Groote, 2020].
- Finally, using the studies and discussions developed throughout this thesis, we can outline the major benchmarks for a model that would fit the logic and linguistic phenomena encountered in dialogue through a compositional, dynamic, and logic-based method rooted in formal semantics.

This list summarizes the content of the following pages. Starting from an overview of formal semantics, this thesis develops our take on a model of dialogue in a dynamic framework with a large side of linguistic theories and experimental work. We hope the result achieves a good balance between outlining a formally consistent model and respecting and accounting for real-life data, which is (frustratingly) messy and (incredibly) interesting.

Our original contributions are threefold. The first contribution of this thesis is the **production of the DinG corpus** from the ground up: starting with the outline of the project and all the way through the recording, transcription, annotation, and analysis process. While we were building DinG, we were also working on other dialogue-based corpora in various languages, to expand our knowledge of questions and answers in dialogue. The second contribution of this thesis is the **development and testing of several cross-lingual annotation schemas for questions and answers in dialogue**. Once our studies of real-life dialogue started to move forward, we could also move on with our formal semantics approaches to dialogue. The third contribution of this thesis lies in the **creation and investigation of several formal semantics models of dialogue in a dynamic framework**. The review of these models guides us towards the outline of major benchmarks for a formal semantics-based model of real-life dialogue.

---

## Applications

While our work does not directly result in a tool that could be shelved, we still think that the insights that we develop throughout this thesis can be very useful in various applications that range from classical NLP tasks to psychology and cognition research.

Our study of questions and answers in dialogue leads us to the development of models that can be used in question and answers tasks in information retrieval. Hybrid approaches, combining machine learning techniques and logic representations such as the ones we produce in this thesis, have successfully been used to improve the results of such tasks, see for example [Demner-Fushman and Lin, 2007] for question answering in a clinical setting.

As our object of study is dialogue, part of our work is directed towards human interaction studies. Nowadays NLP is used in order to help the study and diagnosis of cognitive and psychiatric diseases such as depression [Howes et al., 2014], Alzheimer’s disease [Nasreen et al., 2019], and schizophrenia [Breitholtz et al., 2021]. One of the motivations behind the real-life wing of our work comes from projects studying interactions involving patients with schizophrenia from a computational semantics point of view, such as Dialogical Reasoning in Patients with Schizophrenia (DRiPS<sup>9</sup>) and Schizophrenia and Language: Analysis and Modelisation (SLAM<sup>10</sup>). The corpora used in these projects are difficult to produce, store, and use, as they involve patients with schizophrenia, thus medical confidentiality is at stake. For the same reason, these corpora are quite small, so NLP models that need a lot of training data cannot be easily used in these settings.

We hope that the work we develop in this thesis can be useful at least in two different ways in the context of these studies. First, the corpus we developed is very similar in construction to the corpus built as part of the project called Outils informatisés d’aide au Diagnostic des Maladies mentales (ODiM<sup>11</sup>, Computerized tools to assist in the Diagnosis of Mental Illnesses). The corpus built for ODiM contains highly sensitive data while the DinG corpus is available on Gitlab: <https://gitlab.inria.fr/semagramme/DinG>, distributed under the Attribution Share-Alike Creative Commons license (CC BY-SA 4.0). Therefore the tools developed as part of ODiM can be tested and showcased on DinG, at least in regards to questions and answers in dialogue.

Then, as we have stated above, it is difficult to produce NLP models of corpora involving personal data, especially medical-related, as these corpora are often too small compared to the amount of data needed by neural networks to perform well [Liu and Chen, 2019]. We hope that the logic-based models of dialogue we develop in this thesis can compensate for the lack of data by adding supplementary information sources and enabling the use of hybrid NLP models.

## Thesis outline

This thesis is organized in three parts. The first part presents the linguistic and logic preliminaries that we need to precisely define the scope and the aim of our work. Chapter 1 focuses on formal semantics of discourse. It contains the definitions of the linguistic and logic phenomena we consider throughout this thesis. Based on these definitions, we give a presentation of the main formalisms and frameworks for formal semantics of discourse. Then, we compare them with regards to the presented phenomena and in perspective of the construction of a formal semantics model of dialogue in a dynamic framework. Chapter 2 shifts the focus from discourse to dialogue.

---

<sup>9</sup><https://www.christinehowes.com/research/drips>, visited on 16/07/21.

<sup>10</sup><https://team.inria.fr/semagramme/slam/>, visited on 16/07/21.

<sup>11</sup><https://team.inria.fr/semagramme/odim/>, visited on 16/07/21.

It presents the existing studies and formal accounts of dialogue before diving into the definitions and typologies of questions and answers. Then, we present the requirements for a formal model of interrogative sentences and give a detailed account of question semantics. We end this chapter by focusing on *wh*-questions.

The second part of this thesis concentrates on real-life data. Chapter 3 details the construction of the DinG corpus before presenting the landscape of existing corpora of real-life dialogues in English and French and showing how DinG fits in it. Chapter 4 focuses on the annotation of questions and answers in real-life dialogue and presents several annotation schemas together with cross-lingual experiments. We conclude Chapter 4 by applying a schema for question classification to questions from DinG.

The last part of this thesis presents our investigations on our main research question: the construction of a formal model of dialogue in a dynamic framework. Chapter 5 presents the different solutions we explored for dialogue modeling, using the negotiation phases subdivision of dialogue. We introduce a mathematical tool that accounts for the link between a question and its answer; then, we present a systematic method for compositional computation of logic-based representations of *wh*-questions. Finally, we detail the interactions between the utterances and the dialogue context.

## Part I

# State of the art





# 1

## Formal semantics

### Contents

---

<b>1.1</b>	<b>From natural language to logic</b>	<b>16</b>
<b>1.2</b>	<b>Logic &amp; linguistic phenomena for natural language modeling</b>	<b>16</b>
1.2.1	Syntax, semantics, articulations	17
1.2.2	Quantification, negation	23
1.2.3	Inter- and intra-sentential interactions	23
1.2.4	Towards dialogue modeling	25
<b>1.3</b>	<b>Frameworks &amp; formalisms for sentence &amp; discourse semantics</b>	<b>26</b>
1.3.1	Montague semantics and its descendants	26
1.3.2	FOL inspired approaches, anchored in linguistics	29
<b>1.4</b>	<b>Discussion and comparison in a dialogue modeling perspective</b>	<b>31</b>
<b>1.5</b>	<b>Continuation Style Dynamic Semantics</b>	<b>34</b>
1.5.1	Formal set-up	35
1.5.2	CSDS – an example	36
1.5.3	Abstract Categorical Grammar	39
<b>1.6</b>	<b>Quantificational event semantics</b>	<b>41</b>
1.6.1	Event semantics – First order logic representations	41
1.6.2	Neo-Davidsonian Event Semantics	42
1.6.3	Compositional Neo-Davidsonian Event Semantics	42
<b>1.7</b>	<b>Inquisitive Semantics</b>	<b>46</b>
<b>1.8</b>	<b>Conclusion</b>	<b>48</b>

---

The goal of this chapter is to present the fundamental ideas of formal semantics as a field of computational linguistics. We outline the links between logic and natural language before diving into a detailed account of logic and linguistic-related phenomena in section 1.2. The concepts introduced in this section are essential to our presentation of existing frameworks and formalisms, as they constitute important points of comparison.

The second part of this chapter focuses on three formalisms that are central to the work presented in this thesis: Neo-Davidsonian Event Semantics (NDES) [Parsons, 1990], as presented in [Champollion, 2017]), Continuation Style Dynamic Semantics (CSDS, [de Groote, 2006]) and Inquisitive Semantics (IS, [Ciardelli et al., 2018]). These three formalisms constitute the theoretical foundations of the models presented in Chapter 5.

## 1.1 From natural language to logic

Since the first linguistic works of Ferdinand de Saussure [De Saussure, 1972], natural language has been studied from different points of view, each focusing on specific features such as the sound of it (phonology), its aspects and forms (morphology), its organization (syntax), its meaning (semantics), its meaning in context (pragmatics). In computational semantics, the main goal is to automate the process of reasoning on meaning as much as the articulation of meaning representations. Therefore, one of the first questions that have to be addressed is: how can one represent meaning?

Richard Montague stated a hypothesis that links semantics, logic, and language (see [Janssen and Zimmermann, 2021] for a presentation).

*There is in my opinion no important theoretical difference between natural languages and the artificial languages of logicians; indeed I consider it possible to comprehend the syntax and semantics of both kinds of languages with a single natural and mathematically precise theory. ([Montague, 1970], 373)*

This hypothesis raises several practical issues: what are the syntactic and semantic phenomena one can hope to represent with mathematical logic? conversely, what are the fundamental logic principles and concepts that need to be considered for natural language modeling?

The following section gives a broad presentation of several logic and linguistic phenomena that are common to most of the formalisms and frameworks that resulted from Montague's statement. A detailed overview of these formalisms is given in section 1.3, section 1.4 compares them with respect to their ability to handle phenomena showcased in section 1.2 and in a dialogue modeling perspective.

## 1.2 Logic & linguistic phenomena for natural language modeling

As stated by Montague in 1970, logic and linguistics go hand in hand with regard to several phenomena. For a logical language as much as for a natural one, it is necessary to define the *syntax* (rules stating which sequences are well-formed in the language) and *semantics* (stating the meaning behind well-formed sequences) in order to formally interact with the language. Formal languages are most often assigned a *compositional* semantics (by construction), as they are defined in order for the meaning of a sequence to depend on the meaning of its constituents; natural languages can be seen and studied as compositional, up to some extent.

The *syntax-semantics interface* is a systematic device for computing the meaning of a sentence starting from its syntactic structure. *Quantification* and *negation* are central to any logical language and yield non-trivial phenomena in natural languages. Sentences in natural languages usually exist in an environment of other sentences, a *context*. Following this notion, it is necessary to consider *anaphora* (reference to previously introduced objects) and related phenomena such as *dynamicity*: the possibility to modify previously introduced objects and properties of these objects. Because of the subject of this thesis, we need to present phenomena that are specific to *dialogue* in general and *interrogatives* in particular. Defining all the aforementioned phenomena allows us to outline the principal characteristics in light of which we can present existing frameworks and formalisms, see section 1.3.

**Note.** For the sake of simplicity and coherence, in the following, we will give illustrative examples mainly from First Order Logic (FOL) and scholar English. The phenomena presented in this section are in no case limited to these two illustrations; conversely, the list of presented phenomena is far from exhaustive and only aims to act as a general outline.

### 1.2.1 Syntax, semantics, articulations

**Syntax** The syntax is a set of rules stating which sequences of symbols of a given vocabulary constitute well-formed expressions in the given language. For a logical language, first, a vocabulary is defined, then the rules of syntax are given.

#### Definition 1 (Syntax of FOL)

*Vocabulary of FOL:*

*Truth constants :  $\top$ (true),  $\perp$ (false)*

*Infinite set of variables :  $x, y, z \dots$*

*Infinite set of predicate symbols,*

*coming with an arity (amount of arguments) :  $P, Q, R \dots$*

*Infinite set of function symbols,*

*coming with an arity :  $f, g, h \dots$*

*Parentheses :  $(, )$*

*Negation :  $\neg$*

*Equality symbol :  $=$*

*Conjunction :  $\wedge$*

*Disjunction :  $\vee$*

*Implication :  $\rightarrow$*

*Existential quantifier :  $\exists$*

*Universal quantifier :  $\forall$*

*In FOL, the well-formed expressions are called terms and formulae.*

*Formation rules of FOL:*

- *Any variable is a term.*
- *Any expression  $f(t_1, \dots, t_n)$  where  $f$  is a function symbol of arity  $n$  and  $t_1, \dots, t_n$  are  $n$  terms, is a term.*
- *If  $P$  is an  $n$ -ary predicate<sup>12</sup> symbol and  $t_1, \dots, t_n$  are  $n$  terms, then  $P(t_1, \dots, t_n)$  is a formula.*
- *If  $t_1$  and  $t_2$  are terms, then  $t_1 = t_2$  is a formula.*
- *If  $\phi$  is a formula, then  $\neg\phi$  is a formula.*
- *If  $\phi$  and  $\psi$  are formulae, then  $(\phi \wedge \psi)$ ,  $(\phi \vee \psi)$ , and  $(\phi \rightarrow \psi)$  are formulae.*

---

<sup>12</sup>An  $n$ -ary predicate symbol is a predicate symbol that takes  $n$  arguments; another way of saying this is to say that it is a predicate of *arity*  $n$ .

- If  $\phi$  is a formula and  $x$  is a variable, then  $(\exists x.\phi)$  (there exists a variable  $x$  such that  $\phi$  is true) and  $(\forall x.\phi)$  (for all variables  $x$ ,  $\phi$  is true) are formulae.

**Example 5 (A FOL formula)**

$$(\forall x.(\forall y.((P(x, y) \wedge P(y, c)) \rightarrow P(x, c))))$$

is a well-formed formula of FOL.

Example 5 contains a constant  $c$ , two variables  $x, y$ , a binary<sup>13</sup> predicate symbol  $P$ , parentheses, a conjunction, and an implication symbol. According to the formation rules stated in definition 1, the formula is well-formed.

Similarly, it is possible to define a vocabulary and a set of rules to establish the set of well-formed expressions of (a fragment of) English. As natural languages are prone to changes and creative twists, it is impossible to give strict rules that would work in all usage cases. Many different vocabulary and rule sets can be defined for English, we will only present one here, following [Carnie, 2012]. For natural languages, the vocabulary corresponds to the *syntactic categories* while the lexicon and rules considered for syntax constitute the *grammar*. Table 1.1 shows some of the major parts-of-speech (POS), the syntactic categories associated to lexical items. Table 1.2 lists some of the major syntactic constituents, corresponding to complex syntactic units, computed from the lexical items by using the rules of the grammar.

Abbreviation	Name	Example
N	noun	“unicorn”
PN	proper noun	“Charlie”
V	verb	“sleeps”
TV	transitive verb	“likes”
Adj	adjective	“green”
Adv	adverb	“quickly”
D	determiner	“the”
C	complementizer	“that”

Table 1.1: Major parts-of-speech (POS).

Abbreviation	Name	Example
NP	noun phrase	“the unicorn”
VP	verb phrase	“likes tea”
PP	prepositional phrase	“with a teapot”
AdjP	adjective phrase	“very green”
AdvP	adverb phrase	“very quickly”
S	sentence	“Charlie likes tea”

Table 1.2: Major syntactic constituents.

Example 6 gives a minimal grammar. We use this grammar to compute the syntactic parsing of the sentence “Brutus stabbed Caesar”, illustrated by the tree in example 7.

<sup>13</sup> $n$ -ary with  $n = 2$ .

**Example 6 (A minimal grammar)***Lexicon:*

$$\left\{ \begin{array}{l} PN \rightarrow Brutus \\ PN \rightarrow Caesar \\ TV \rightarrow stabbed \end{array} \right.$$

*Grammar:*

$$\left\{ \begin{array}{l} S \rightarrow NP VP \\ VP \rightarrow TV NP \\ NP \rightarrow PN \end{array} \right.$$

The lexicon in example 6 gives us the information that in this grammar, a PN is either Brutus or Caesar, and the only available TV is “stabbed”. The grammar states that a sentence can only be decomposed as an NP and a VP. The VP, in turn, can only be decomposed as a TV and an NP. The only type of NP available to us in this limited grammar is a PN, which brings us back to the lexicon.

Using the syntactic categories and a grammar it is possible to analyze a sentence in terms of syntactic constituents, as in example 7. It can be done by hand or automatically, using for example the Stanford parser<sup>14</sup>, [Klein and Manning, 2003, Klein et al., 2003], currently available for 6 languages: Arabic, Chinese, English, French, German, and Spanish.

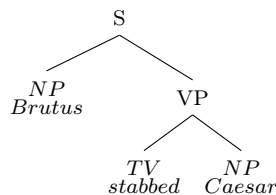
**Example 7 (A simple syntactic tree)**

Figure 1.1: Syntactic tree of the sentence “Brutus stabbed Caesar”.

Example 7 shows a simple syntactic tree corresponding to the sentence “Brutus stabbed Caesar”. In this sentence, “stabbed” is a transitive verb (syntactic label TV). “Brutus” and “Caesar” are proper nouns, and the corresponding syntactic label is NP for noun phrase (as proper nouns are noun phrases). The syntactic constituent obtained after combining the transitive verb with its object is a verb phrase VP. Then, the verb phrase combined with the subject noun phrase yields the full sentence (syntactic label S).

**Semantics** Semantics assigns a meaning to a well-formed (according to syntax) sequence of lexical items and symbols of a language. For logical languages, the semantics of a language can be defined by giving a domain of discourse ( $D$ ) and an interpretation function ( $I$ ) which assigns to each constant an object in  $D$ , to each  $n$ -ary predicate symbol an  $n$ -ary relation on  $D$ , i.e. a subset of  $D^n$ , and to each  $n$ -ary function symbol a function from  $D^n$  to  $D$ .

**Example 8 (A discourse domain and interpretation for FOL)**

Consider a first order language with a binary predicate symbol  $P$ , and a constant  $c$ . Let  $D$  be the set of unicorns. We define  $I$  as follows:

<sup>14</sup><https://stanfordnlp.github.io/CoreNLP/>, visited on 18/06/2021.

- $I(c) \in D$  is defined to be *Charlie*.
- $I(P)$  is the binary relation of being siblings:  $I(P)(x, y)$  is true if and only if  $I(x)$  and  $I(y)$  are siblings.

Then the semantics of the formula presented in example 5 is the following: “For all unicorns  $I(x)$ ,  $I(y)$ , if  $I(x)$  and  $I(y)$  are siblings and  $I(y)$  and  $c$  are siblings, then  $I(x)$  and  $c$  are siblings”, which is true in the discourse domain of unicorns if we suppose that they have sibling relations similar to human and animal ones.

As the formula presented in example 5 is the definition of the property of transitivity, any interpretation of  $P$  as a transitive property would result in this formula being true in this interpretation.

Formal semantics of natural languages, in a way, proceeds conversely to the semantics of logical languages. The goal here is to assign a logical formula to a natural language sentence. We call this formula the *semantic representation* of the sentence. For a natural language expression  $s$ , the semantic representation of  $s$  is denoted  $\llbracket s \rrbracket$ . Deciding upon the right semantic representation of a sentence or a sentence constituent is a fundamentally linguistic question, while determining the course of its computation is the prerogative of computational linguistics.

### Example 9 (Some semantic representations)

$\llbracket Brutus \rrbracket$	=	<i>brutus</i>
$\llbracket stabbed \rrbracket$	=	$\lambda o.s.stab(s, o)$
$\llbracket Caesar \rrbracket$	=	<i>caesar</i>

Table 1.3: Semantic representation of “Brutus”, “stabbed”, “Caesar”.

In example 9, the proper nouns “Brutus” and “Caesar” are represented with the constants *brutus* and *caesar*. The transitive verb “stabbed” is represented as a  $\lambda$ -term<sup>15</sup> that should be read as follows: we are looking for two entities,  $o$  (for “object of the transitive verb”) and  $s$  (for “subject of the transitive verb”), and then we evaluate the truth value of the predicate *stab* applied to the objects  $s$  and  $o$  (that could be variables or other terms or formulae).

Another classical device of formal semantics of natural languages is *type theory*. The idea is that different sorts of objects of natural language have different types and that a type defines the behavior (meaning and authorized operations) of the object. More formally, Montague Semantics [Montague, 1970] uses two atomic types:  $e$ ,  $t$ .  $e$  is the type of individuals/entities, such as constants, and  $t$  is the type of truth values, it is the type of well-formed formulae.

### Definition 2 (Basic types, Montague Semantics)

- $e$  : individual/entity
- $t$  : truth values
- if  $\alpha$  and  $\beta$  are types, then  $\alpha \rightarrow \beta$  is a type

---

<sup>15</sup>For an introduction to  $\lambda$ -calculus, see for example [Alama and Korbmacher, 2021].

**Compositionality** Montague’s ideas follow the modern formulation of the principle of compositionality attributed to Gottlob Frege [Frege, 1892].

*The meaning of a compound expression is a function of the meanings of its parts and of the way they are syntactically combined.* ([Partee, 1984], 281)

At a sentence level, the principle of compositionality implies that one needs to know the meaning of each part of the sentence and the syntactic rules that govern the combination of these parts. Montague semantics [Montague, 1973] is based on Church’s simple theory of types [Church, 1940]. It gives the representation of constituents’ meanings along with rules that dictate how to combine them. The semantic rules are driven by syntax, they enable the computation of the representation of the complex expression’s meaning.

A notable exception to the principle of compositionality in natural languages is constituted by *multi-word expressions* such as “green card”: these two words together do not designate a card that is green, at least in some context such as the United-States residency<sup>16</sup>. Without knowing the context and the fact that these two words produce this specific meaning together, it is not possible to compute the meaning by combining the meanings of “green” and “card”. The study of multi-word expressions is a complete and extremely interesting field of natural language processing, which is not the subject of this thesis; some references can be found here: [Hüning and Schlücker, 2015, Siyanova-Chanturia et al., 2017].

**Syntax-semantics interface** Now that we have defined concepts linked to syntax and semantics, our goal is to be able to build, compositionally, the semantic representation of a sentence, from the syntactic analysis of this sentence. In order to do so, we need to define a set of rules that constitute the interface between syntax and semantics. In the following paragraph, we loosely follow the presentation given in [Roussarie, 2017].

As we have already introduced syntactic trees above, we can now use them in order to introduce a semantic analysis in the syntactic tree. For this, we use types, starting from the leaves of the tree to which we assign known types, and rising through the tree while computing the intermediary types compositionally.

Syntactic category	Type
Proper noun (PN)	$e$
Sentence (S)	$t$
Noun (N), non-transitive verb (V), adjective (Adj)	$e \rightarrow t$
Transitive verb (TV)	$e \rightarrow (e \rightarrow t)$

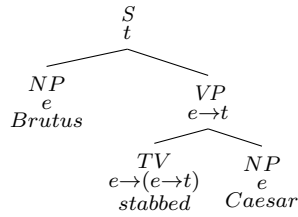
Table 1.4: Syntactic categories and corresponding semantic types, adapted from [Roussarie, 2017].

A term of type  $\alpha \rightarrow \beta$  is such that if it is combined (applied to) with a term of type  $\alpha$ , the result is of type  $\beta$ . The term of type  $\alpha \rightarrow \beta$  is then called a *functor*, the term of type  $\alpha$ , an *argument*. In table 1.4, a noun (N) of type  $e \rightarrow t$  is an object that, if applied to an object of type  $e$ , will yield an object of type  $t$ .

Following the correspondence presented in table 1.4, we can assign types to the tree introduced in example 7. We obtain the tree from example 10.

<sup>16</sup>A *green card* or *permanent resident card* is an identity document proving that a person has permanent residency in the United-States.



**Example 10 (A simple syntactic tree with types)**

The types in the tree from example 10 give us the order in which the semantic representations of the leaves should be combined to compositionally compute the semantic representation of the whole sentence. “stabbed”, which semantic representation is of type  $e \rightarrow (e \rightarrow t)$ , is the functor that should be applied to the argument “Caesar”, which semantic representation is of type  $e$ . This application is computationally evaluated through what is called  $\beta$ -reduction, denoted  $\rightarrow_{\beta}$ . The result of this application is an intermediary semantic representation of a verb phrase, of type  $e \rightarrow t$ .

The semantic representations presented in example 9 can now be written with the corresponding types, traditionally written as superscripts on corresponding objects: see example 11.

**Example 11 (Some semantic representations, with types)**

$\llbracket Brutus \rrbracket$	:	$brutus^e$
$\llbracket stabbed \rrbracket$	:	$\lambda o^e s^e . stab(s, o)^t$
$\llbracket Caesar \rrbracket$	:	$caesar^e$

Table 1.5: Semantic representation of “Brutus stabbed Caesar”.

The semantic representations of example 11 combined with the information given by the syntactic tree of example 10 allow us to compute the semantic representation of the sentence “Brutus stabbed Caesar”.

**Example 12 (Computation of the semantic representation of “Brutus stabbed Caesar”)**

$$\begin{aligned}
 \llbracket Brutus \text{ stabbed } Caesar \rrbracket &= (\llbracket stabbed \rrbracket \llbracket Caesar \rrbracket) \llbracket Brutus \rrbracket \\
 &= (\lambda o s . stab(s, o) \text{ caesar}) \text{ brutus} \\
 &\rightarrow_{\beta} \lambda s . stab(s, caesar) \text{ brutus} \\
 &\rightarrow_{\beta} stab(brutus, caesar)
 \end{aligned}$$

This example concludes our presentation of the linguistic and logic phenomena concerning syntax, semantics, and the articulation and links between the two. The next section introduces quantification and negation, two phenomena deeply rooted in logic whilst ubiquitous in natural language.

### 1.2.2 Quantification, negation

First order logic features two quantification operators:  $\exists$ , the existential quantification, and  $\forall$ , the universal quantification. In classical logic, these two quantifiers are duals and each one can be defined from the other:  $\neg(\exists x.P(x)) \equiv \forall x.\neg P(x)$  and  $\neg(\forall x.P(x)) \equiv \exists x.\neg P(x)$ . In natural languages, quantification-related phenomena are not that straightforward, and choosing the right logic to be able to handle as many different usages as possible is a very important starting point for a language model, see [Gamut, 1991] for an overview.

Some examples are easy to translate to first order logic: consider the sentence “Everyone loves tea”. This sentence can easily be translated in FOL, for example as  $\forall x.\text{love}(x, \text{tea})$ . Yet, quantification in FOL can only be done on variables, not on functions or predicates. In natural language setting, it happens that one quantifies over a verb, and as exposed above, the semantic representation of verbs is often a predicate. Some quantifiers such as “most” are often used in natural languages but do not have direct translation in logic, see [Hackl, 2009] and [Carcassi and Szymanik, 2021] for discussions.

The traditional translation of determiners such as “every” to logic is done with a universal quantifier, of indefinite articles (such as “a”, “some”) with an existential quantifier. Yet, as will be shown below, systematic translation from natural language to logic is not straightforward.

Negation is also a fundamental feature of mathematical logic that yields complicated linguistic phenomena when observed in natural language setting. Consider the sentence “Charlie likes Sasha”. According to what we have presented so far, a semantic representation of this sentence could be  $\exists x.\exists y.\text{likes}(x, y) \wedge (x = \text{charlie}) \wedge (y = \text{sasha})$ . Logically, it could be possible to introduce negation at any point in this formula:

$$\exists x.\exists y.\neg\text{likes}(x, y) \wedge (x = \text{charlie}) \wedge (y = \text{sasha}) \quad (1.1)$$

$$\exists x.\exists y.\text{likes}(x, y) \wedge \neg(x = \text{charlie}) \wedge (y = \text{sasha}) \quad (1.2)$$

$$\exists x.\exists y.\text{likes}(x, y) \wedge (x = \text{charlie}) \wedge \neg(y = \text{sasha}) \quad (1.3)$$

$$\exists x.\exists y.\text{likes}(x, y) \wedge \neg(x = \text{charlie}) \wedge \neg(y = \text{sasha}) \quad (1.4)$$

$$\exists x.\exists y.\neg\text{likes}(x, y) \wedge \neg(x = \text{charlie}) \wedge \neg(y = \text{sasha}) \quad (1.5)$$

Yet, it is not clear which of these sentences actually correspond to valid sentences in English. While it is possible to provide interpretations for (1.1) (“Charlie doesn’t like Sasha”), (1.2) (“It’s not Charlie who likes Sasha”), (1.3) (“Charlie likes someone who is not Sasha”), it becomes difficult to interpret (1.4) and (1.5). For (1.4), is it “Someone who is not Charlie likes someone who is not Sasha”? For (1.5), “Someone who is not Charlie doesn’t like someone who is not Sasha”? While these two interpretations are indeed possible, they do not read as common natural language examples.

### 1.2.3 Inter- and intra-sentential interactions

**Context** Up to now, we have only discussed linguistic phenomena occurring inside a single sentence. In natural languages, sentences are encountered in *context*: collocated with one or more other sentences, linked with other sentences by different relations and common references. A sequence of sentences can form a *discourse*, and relations occurring in discourse are then called *discourse relations*. Discourse is considered in context; several different theories of context can apply, see for example [Lihoreau and Rebuschi, 2014] for an overview. The main contexts that will be considered in this thesis are, on one hand, the immediate context of a sentence in discourse or dialogue (all the information introduced by previous sentences in the discourse), and on the

other hand, the general context in which all our natural language productions exists (we are speaking 2021 international English, we know that France is a country, etc.). In the following, we will not discuss the general context modeling, and when the context will be mentioned, it will be the immediate context, unless explicitly stated otherwise.

**Anaphora** One of the immediate consequences of the existence of a context is that it is possible to “store” objects and information in it, and therefore, to reference previously-stored facts. One of the classical reference phenomena in natural language is the *anaphora*: the use of an expression whose interpretation depends on another expression in context.

**Example 13 (An anaphora)** *Alice has a car. It is red.*

In example 13, the first sentence introduces two entities: Alice, and the car. The second sentence starts with a neuter pronoun, “it”. Pronouns are among the most frequent anaphoric phenomena encountered in natural languages. “it” in the second sentence can only reference the car, assuming Alice’s pronouns are “she/her”. Therefore, we can interpret the information provided by the second sentence in the example as “Alice’s car is red”.

**Example 14 (A problematic anaphora)** *Alice doesn’t have a car. \*It is red.*

Consider example 14. The first sentence introduces the entity Alice, but it is up to discussion whether or not it introduces the entity “car”. Indeed, the second sentence (prefixed by an asterisk, traditionally signifying that the sentence is not valid) couldn’t occur in this context, as the pronoun “it” references an object that does not exist. We can thus conclude that the first sentence does not introduce “Alice” and “car” at the same level of context storage: we can indeed make references to Alice in the following, not to her car, and still, we have information about Alice and a car, namely that she doesn’t have one, which doesn’t amount to no information at all about a car. Yet, if we consider the sequence “Alice doesn’t have a car. She wouldn’t know where to park it.”, the anaphora can be solved (see [Qian, 2014] for detailed discussions). This is one of the many complex phenomena one has to take into account when working with sentences in context.

Another phenomenon linked with anaphora and context considerations lies within the so-called donkey-sentences<sup>17</sup> [Geach, 1962].

**Example 15 (Donkey-sentence)** *Every person who caresses a unicorn loves it.*

This sentence is valid and understandable by a human reader. From formal semantics point of view, it seems possible to translate this sentence to first-order logic, the major difficulty compared to what we discussed above lying in the translation of pronouns. A first attempt, using free (not linked to quantifiers) variables to translate pronouns, would result in the following:

$$\forall x.(person(x) \wedge \exists y.(unicorn(y) \wedge caress(x, y)) \rightarrow love(x, y))$$

$love(x, y)$  uses the variable  $y$ , linked to the existential quantifier; yet, as  $love(x, y)$  occurs on the righthand side of the implication, it is not under the scope of this existential quantifier, the variable  $y$  is free in  $love(x, y)$ . So this translation does not reflect the fact that the unicorn that the person loves is the same unicorn the person caresses.

---

<sup>17</sup>The prototypical sentences showcasing these phenomena featured donkeys and farmers, hence the name.

If we correct the first shot at translating by extending the scope of the existential quantifier, we obtain the following:

$$\forall x.\exists y.((person(x) \wedge unicorn(y) \wedge caress(x,y)) \rightarrow love(x,y))$$

This time, the variable  $y$  is bound to the right quantifier. However, this translation does not reflect the expected meaning of the sentence given in example 15. It is always true if there is at least one entity in the domain of interpretation that is not a unicorn.

$$\forall x.\forall y.((person(x) \wedge unicorn(y) \wedge caress(x,y)) \rightarrow love(x,y))$$

While this translation corresponds to the meaning of the original sentence, it raises questions related to the systematic way of translating natural language sentences to logic formulae. We stated above that the traditional way of translating “a” to logic is with an existential quantifier, yet we have just shown that this doesn’t hold in the donkey-sentence case. Donkey-sentences constitute an important subject in formal semantics of natural language, see in particular [Kanazawa, 1994, Geurts, 2002, Champollion et al., 2019].

**Dynamicity** The previous paragraphs have introduced problematics linked to context in general and anaphora in particular. They also introduced the idea of dynamicity within a sentence, through the discussions revolving around donkey-sentences. For now, we focused on an idea of a context as a storage place, where entities that can be referenced are stored. The next question is to see how can one interact with the context with each new information: can one modify an entity stored in the context? is it possible to update information stored in the context? how to manage contradictory information? Managing the context as the sentences unravel is the central question of dynamical frameworks.

#### 1.2.4 Towards dialogue modeling

**Dialogue** The current section has first focused on intra-sentential phenomena, then (starting from context-related considerations) we have introduced the idea of discourse as a sequence of sentences. Similarly, it is possible to see dialogue as a linear sequence of utterances, following one another. First formal approaches to dialogue can consider that assertive utterances can be modeled using the same mechanisms as the ones for declarative sentences. Of course, not all utterances are sentences, and working with non-sentential utterances is a major part of dialogue studies. Non-Sentential Utterances (NSU) constitute around 10% in English-speaking corpora such as the British National Corpus (BNC) [Fernández et al., 2007], see Chapter 3 for more about BNC. NSUs can appear under various forms and for different types of illocutionary use.

**Interrogatives** If we only know how to model declarative sentences, it is impossible for us to account for a major phenomenon of dialogue: questions. The logical models we presented above are compositional and truth-conditional, they assign a truth value to each declarative sentence by computing it using the sentence’s constituents. For interrogative sentences, truth-conditional approaches do not hold, as it is not clear what assigning a truth value to a question would mean [Groenendijk and Stokhof, 1984]. Is a yes/no-question true or false? What does it mean for a wh-question to be true or false? There is an extensive classic literature on formal semantics of interrogative sentences, such as [Hamblin, 1976], [Hintikka, 1983], [Groenendijk and Stokhof, 1984], [Krifka, 2001]. In the following, we present and discuss several ideas and frameworks that stem from them.

### 1.3 Frameworks & formalisms for sentence & discourse semantics

The previous section has introduced the main logic and linguistic phenomena within which formal semantics of natural language operates. The following presents major formal semantics frameworks and formalisms in perspective of these phenomena. Our goal is to give a short description of each formalism along with a characterization of its ability to handle the aforementioned phenomena. This presentation is in no way universal, it is explicitly directed towards the goal of this thesis: dialogue modeling.

The choice of formalisms for our overview was guided by classical formal semantics literature but also by the logic and linguistic phenomena we focus on in this thesis. In particular, all the formalisms from our overview present a way to model semantics of natural language, and all of them have at least a partial treatment of logical quantification and negation. Our list of formalisms is by no means exhaustive, but it allows us to come up with observations and comparisons between existing well-established approaches, in order to outline the requirements for models yet to come.

In this section, we give short presentations of the main formalisms and frameworks that are most frequently found in literature: Montague Semantics (MS), File Change Semantics (FCS) and Discourse Representation Theory (DRT), Segmented Discourse Representation Theory (SDRT), Continuation Style Dynamic Semantics (CSDS), Neo-Davidsonian Event Semantics (NDES), Inquisitive Semantics (IS), Minimal Recursion Semantics (MRS), Type Theory with Records (TTR), KoS, and Abstract Meaning Representation (AMR).

#### 1.3.1 Montague semantics and its descendants

**Montague Semantics (MS)** Montague semantics [Montague, 1970] opened the way to logical approaches to natural language. Its goal is to model the meaning of natural language sentences using logic. Montague semantics uses syntactic structures of natural language, in order to produce semantic representations of sentences. It doesn't handle inter-sentential linguistic phenomena such as pronoun binding and anaphora across several sentences, as it was thought for isolated sentences.

This model is anchored in logic by design, it handles some quantification and simple negation-related phenomena. It was built in order to be compositional and plants the seeds of first syntax-semantics interfaces. As Montague semantics doesn't take dynamicity into account, it presents no native way of representing the context.

Montague's work contains some preliminary thoughts on models for questions, but it was not designed for interrogatives nor, furthermore, dialogue. The last chapter of Dowty's *Introduction to Montague semantics* [Dowty et al., 2012] lists twelve developments of MS designed in perspective of the treatment of questions, among which [Hamblin, 1976]. Some of the frameworks and formalisms presented in this section stem from these references, see in particular Robin Cooper's TTR.

**File Change Semantics (FCS) & Discourse Representation Theory (DRT)** Among the most well known formal models of natural language following Montague's ideas come File Change Semantics (FCS), introduced by Irene Heim in [Heim, 1982] and Discourse Representation Theory (DRT), introduced by Hans Kamp in [Kamp, 1981]. Though those were developed independently, they share one main idea – consider the meaning of segments composing a discourse as a relation between meanings of previous and yet-to-come sentences. The sentences are then seen as objects that can entail modifications (*changes*, according to FCS) in previously conducted analysis of

discourse (stored in *files*, according to FCS). The core notion underlying the DRT framework is the one of *discourse representation structures* (DRS) – a representation of discourse that is being enriched as new sentences are analyzed, as a human hearer would proceed with a mental representation of a discourse [Kamp and Reyle, 2013].

FCS focuses on the notions surrounding discourse referents<sup>18</sup> [Karttunen, 1976], identifying them with *file cards*, stored in the files [Heim, 2012]. Then, the introduction of an indefinite reference starts a new file card; the use of a definite reference updates an existing file card. DRSs use sets of reference markers that should be understood both as existential quantifiers (necessary for the representation’s sake) and free variables (needed to be able to enrich representations as the discourse continues). Combining two DRSs in order to create one containing all the information provided by the previous two is a complex task that fails if variables assignments from both DRSs accidentally collide. As for File Change Semantics, it relies on the existence of a translation between syntactic structures and logical forms but doesn’t formally settle for a specific translation. DRT is compositional in that representations in DRT are computed through combinations of DRSs. It integrates some dynamical aspects but often mishandles long-scope quantification.

FCS assigns a *file change potential* to logical forms representing sentences, defining functions from files to files. This idea paves the way for many modern-day context models, see in particular [Schlangen, 2020]. DRT and Montague semantics have led to several developments such as  $\lambda$ -DRT [Bos et al., 1994], which adds compositionality to DRT by integrating it with  $\lambda$ -calculus, and [Muskins, 1996], that presents a combination of DRT with classical type theory. Neither DRT nor FCS were thought for dialogue.

**Segmented Discourse Representation Theory (SDRT)** Since the ’90s, Nicolas Asher and Alex Lascarides started developing a theory that would extend dynamic semantics by exploiting additional information contained in discourse, in particular discourse coherence. On top of the order of sentences in a discourse, their inherent syntax, and the way they can be composed, the idea underlying Segmented Discourse Representation Theory (SDRT) is to take into consideration relations between segments of discourse, called Elementary Discourse Units (EDUs, [Asher and Lascarides, 2003]). SDRT is then defined using extensions of DRSs, which are enriched with discourse relations such as **Explanation**, combined with semantics that is an extension of DRT’s semantics enriched with the handling of discourse relations.

This multi-layered structure allows to handle a wider collection of discourse-related phenomena. In particular, it copes with non-chronological accounts of events (ex: “John smiled. Mary gave him flowers the day before.”), anaphora and temporal anaphora in particular, see [Asher and Lascarides, 2003] for details. SDRT is concerned with the link between semantics and pragmatics. It is based on two logic structures: one logic to represent the information content and another for packaging the information and updating discourse. The *logic of information content* is a non-monotonic logic, tailored for inferring discourse relations and allowing at least a basic handling of quantification and negation. The formal language underlying information packaging and updating discourse is called *glue logic*, see [Asher and Lascarides, 2003] for a complete presentation.

SDRT models discourse relations (such as rhetorical ones) and is dynamic. As the SDRT models the relations between the Elementary Discourse Units (EDUs) and those relations can be used to build Complex Discourse Units (CDUs) and compute the relations between CDUs,

---

<sup>18</sup>Following [Qian, 2014], we use an effect-based definition of discourse referents: “if an indefinite reference can serve as antecedent in an anaphoric relation, it introduces a discourse referent; otherwise, it does not.”



SDRT is compositional. Yet, it is not compositional in a syntax-semantics interface perspective as one cannot compute the relations between the EDUs using syntactic information.

[Asher and Pogodalla, 2010] present a version of SDRT in continuation semantics, using the CSDS framework (see below). This version of SDRT uses syntax and a syntax-semantics interface to build semantic representations of discourse. [Asher and Lascarides, 2003] provides a presentation of SDRT for dialogue, adding an analysis of imperatives and interrogatives, a way to keep track of the speaker, along with some epistemic (concerned with belief and knowledge) considerations. The set of relations is extended with additions such as **Indirect Question Answer Pair** (IQAP). See Chapter 2, section 2.2.4 for a more detailed presentation of discourse relations in SDRT.

**Continuation Style Dynamic Semantics (CSDS)**<sup>19</sup> Continuation Style Dynamic Semantics is a formal semantics framework in the tradition of Montague Semantics (MS) and Discourse Representation Theory (DRT). The main idea of CSDS is to extend MS's type system and syntax-semantics interface in order to account for context in discourse: the sentences that were in the discourse before the one being represented and the sentences yet to come.

CSDS follows MS and uses syntactic structures of natural language for computing the semantic representations of sentences in discourse. It encodes DRT's ideas of dynamicity in order to be able to account for anaphoric phenomena. The original version of CSDS doesn't provide a way to model interrogatives; the work presented in this thesis gives a way to extend the framework to questions. CSDS uses Church's simple theory of types and  $\lambda$ -calculus to compute the semantic representations of sentences. In particular, the original article presents a thorough account of donkey sentences with precise handling of quantifiers. CSDS is compositional by design. It is fully operationalized through the Abstract Categorical Grammars framework [de Groote, 2001, Pogodalla, 2016]. The context of discourse is viewed as two-fold: the left-hand context, containing the information introduced by the sentences preceding the one we want to represent, and the right-hand context, the *continuation*, containing the information from the sentences yet to come. CSDS wasn't originally designed for dialogue, but as mentioned above, dialogue can be seen as a linear aggregation of utterances, similarly to the way discourse is a linear aggregation of sentences. Following this idea, [Boritchev and de Groote, 2020] outline an adaptation of CSDS for dialogue.

**Type Theory with Records (TTR) & KoS**<sup>20</sup> Type Theory with Records (TTR) presented by Robin Cooper in [Cooper, 2005a], [Cooper, 2005b], [Cooper, 2008] is a framework that represents situations, using entities typed with *record types* [Betarte and Tasistro, 1998], complex types that contain multiple fields and can depend on one another. Its goal is to enrich Montague Semantics by using a more complex, therefore expressive, type-set.

KoS, which is a toponym<sup>21</sup> and not an acronym, but loosely corresponds to Conversation Oriented Semantics, is a framework for dialogue context and interaction [Ginzburg, 2012, Maraev et al., 2018]. KoS is based on the TTR formalism. In KoS, dialogue is viewed as a game taking place on a *dialogue game board* (DGB), and each utterance induces a modification of the DGB following a set of rules.

---

<sup>19</sup>This paragraph is a short presentation of Continuation Style Dynamic Semantics (CSDS, [de Groote, 2006]). For the extensive presentation, see 1.5.

<sup>20</sup>This paragraph gives a short presentation of KoS [Ginzburg, 2012]. For a more detailed presentation, see Chapter 2, section 2.2.3.

<sup>21</sup><https://en.wikipedia.org/wiki/KoS>, visited on 29/06/21.

In TTR, linguistic objects (ex: “unicorn”, “tea”) can be viewed as records, using insights from linguistic frameworks such as head-driven phrase structure grammar (HPSG, [Pollard and Sag, 1994]). KoS features a TTR and HPSG-based grammar that allows for syntax to be taken into account, with sentence-level compositionality.

In general, TTR and KoS are well fit to represent concepts and situations, working on a higher semantics representation level than syntax-semantics interface-driven approaches. [Ginzburg, 2012] is a book which main subject is the formalization of questions in dialogue, using KoS.

[Ginzburg and Purver, 2012] present an account of quantification in TTR while [Cooper and Ginzburg, 2012] formalizes negation. From a dynamic point of view, DGBs can be considered as instant representations of the dialogue at a given point: we know what the current utterance is, but we also have access to the results of all that has previously been said. [Maraev et al., 2018] presents an implementation of dialogue management in KoS/TTR.

KoS allows keeping track of the dialogue structure. Using the DGBs grants a visual way of following the dialogical moves of the participants. DGBs are used both to represent private and public information. Public information comes from utterances that all the participants of the dialogue hear, whereas private information is personal to a participant. DGBs are used to store, at a given point, information about the *questions under discussion* (QUD), facts (propositions that have been agreed upon), the utterance that is being analyzed along with its information: speaker, addressee, content. TTR and KoS are closely linked with development of *topoi* [Breitholtz, 2020] theory, stating that reasoning used in natural language interactions is often *enthymematic*: we use arguments (*enthymemes*) that can only be understood using implicit premises from the general context (*topoi*). Chapter 3 of [Breitholtz, 2020] develops an account of topoi theory in TTR.

### 1.3.2 FOL inspired approaches, anchored in linguistics

**Neo-Davidsonian Event Semantics (NDES)**<sup>22</sup> NDES is a formalism in which every sentence is considered in terms of occurring *events* or *states* conjuncted with ways sentence constituents relate to this event, through thematic roles, such as **Agent** or **Theme**, see table 1.7 for examples. The presentation of Neo-Davidsonian Event Semantics given in [Parsons, 1990] provides extensive linguistic justification for this formalism. The goal of NDES is to produce semantic representations of sentences. Parsons’ original presentation is not anchored in syntax, but later versions of NDES such as [Champollion, 2017] produce the same semantic representations as Parsons’ starting from syntactic parses of sentences. In Champollion’s presentation, semantic representations of elementary constituents of the sentences are typed and several closure operators allow to handle dynamicity issues such as anaphora resolution when left-hand context appears. Neither version of NDES provides a way to model interrogatives; the work presented in this thesis proposes such extensions.

Events and states are bounded through existential quantifiers in NDES. Consequently, there is an interaction between implicit quantification of events and explicit quantification arising from quantified phrases. NDES representations provide easy ways to prove logical entailment between sentences (ex: “Charlie kissed Sasha in the garden” entails “Charlie kissed Sasha”). [Bernard and Champollion, 2018] show an advanced discussion of negation in Champollion’s NDES, giving an account of negative events.

Recent updates such as [Champollion, 2017] have transformed NDES into a compositional and dynamic framework. Representations of declarative sentences can then be built in a way that

---

<sup>22</sup>This paragraph is a short presentation of Neo-Davidsonian Event Semantics (NDES, [Parsons, 1990]). For the extensive presentation, see 1.6.



flattens down the sentence’s syntactic structure and makes thematic roles accessible for further work [Amblard, 2007]. There also are some implemented versions, in particular in formalisms such as Robust Minimal Recursion Semantics (RMRS, [Copestake, 2007], see below for a presentation of MRS) and in the Lambda Calculator<sup>23</sup>, [Champollion et al., 2007]. A first outline of NDES for questions and answers in dialogue can be found in [Boritchev and Amblard, 2020].

**Minimal Recursion Semantics (MRS)** Minimal Recursion Semantics (MRS) is a framework with complex descriptions for representation and reasoning [Copestake et al., 2005], designed to handle scope ambiguities. It contains numerous rules with elaborate articulations, and its authors position it as a “meta-language for describing semantic structures in some underlying object language” [Copestake et al., 2005], for example, FOL with generalized quantifiers (see [Westerståhl, 2019] for a presentation of generalized quantifiers). The elementary unit of MRS is called an *elementary predication* (EP): a single relation with its associated arguments (ex: `beyond(x,y)`). MRS structures are then defined using bags of EPs, *handles*, and handles constraints. Handles control the underspecification of MRS in order for it not to overgenerate meanings.

MRS’s design aims for *expressive adequacy*, namely correct expression of linguistic meaning, and *grammatical compatibility*, linking the semantic representations to grammatical information, among which, syntax. The flatness of MRS structures is motivated by issues related to variable binding operator’s scopes. In particular, MRS is concerned with (generalized) quantifiers, which results in a generally very good handling of quantification and negation in natural language, but might be problematic for cases such as donkey sentences.

MRS is compositional in terms of EPs. Lexical items are represented as EPs, and bags of EPs are used to build MRS structures according to rules of semantic composition. [Copestake, 2007] shows that Robust Minimal Recursion Semantics (RMRS) allow for a fully compositional implementation. Studies such as [Egg, 1998, Yao and Zhang, 2010] show that Underspecified Minimal Recursion Semantics (UMRS) and MRS can be used for question representation. Difficulties arise in accounting for the partial aspects of answers, that are prevalent in dialogue. [Hough et al., 2015] give a presentation of Robust Minimal Recursion Semantics with incremental processing (RMRS-IP, [Peldszus et al., 2012]) in regards to situated dialogue. They conclude that RMRS-IP needs enriching in order to be properly used for dialogue semantics.

**Inquisitive Semantics (IS)**<sup>24</sup> Inquisitive Semantics was specifically built to provide a way to model interrogative and declarative sentences in an equivalent way. In IS, all propositions are considered to have both an informative and an inquisitive content, which allows for a uniform treatment of declarative and interrogative sentences. The logic features two projection operators, ! and ?. The first one transforms any proposition into a purely informative one by canceling its inquisitive content. Conversely, the second one transforms any proposition into a purely inquisitive one by canceling its informative content.

IS is a logical language with its syntax and semantics; it doesn’t account for syntactic structures or handle linguistic phenomena such as anaphora. Extensions such as Dynamic Inquisitive Semantics (DIS, [Dotlačil and Roelofsen, 2019]) account for these aspects. As IS is a logic for interrogatives, it provides an extensive model for questions. IS uses FOL quantification and negation. The interesting part comes from the interaction of the logical operators with the pro-

---

<sup>23</sup><http://lambdacalculator.com/>, visited on 18/06/2021.

<sup>24</sup>This paragraph is a short presentation of Inquisitive Semantics (IS, [Ciardelli et al., 2018]). For the extensive presentation, see 1.7.

jection operators: for example, existential quantification introduces an inquisitive dimension in a formula, that can be canceled using the ! projection operator if one wants a purely informative existential quantification.

[Ciardelli et al., 2017] introduce typed Inquisitive Semantics, a compositional take on Inquisitive Semantics that addresses quantification, negation, some wh-questions, notions of entailment and coordination. [Champollion et al., 2015] and [Champollion et al., 2017] use typed Inquisitive Semantics to model questions and presuppositions. DIS is a dynamic version of IS; [Dotlačil and Roelofsen, 2019] shows how to use it for anaphora, questions, and interactions between anaphora and questions. [Groenendijk, 2008] presents thoughts on modeling pragmatics of dialogue with IS. The work presented in this thesis is part of this perspective.

**Abstract Meaning Representation (AMR)** Among formal, computational linguistics approaches, one can find Abstract Meaning Representation (AMR, [Banarescu et al., 2013]). It is a formalism that originally represented sentences through relations between predicates and their arguments. Extensions of AMR now allow to augment these representations with thematic roles analysis. Each sentence is represented with a rooted, directed, acyclic graph with labeled edges (each label represents a relation or a thematic role) and labeled leaves that represent concepts. A single AMR can represent several natural language sentences as it is not an annotation of individual words in the sentence.

AMR has been developed mainly for English. The goal of AMR representations is to abstract from syntactical constraints: sentences that have the same meaning but different formulations are represented with the same AMR. The representations use thematic roles (see table 1.7) and frames from PropBank frameset [Kingsbury and Palmer, 2002]. Questions are marked with the label `amr-unknown`. Polar questions are represented as declarative sentences with an `amr-unknown` in the `:polarity` field, corresponding to the idea that one doesn't know the truthfulness of the sentence. Choice questions' representation use `amr-choice`, usually in place of `or`, receiving any number of numbered `:op` (option) arguments<sup>25</sup>.

AMR doesn't have a universal quantifier, it doesn't handle any quantification-related phenomenon; it is not logic-based. Negation is represented with a - in the `:polarity` field of the corresponding structure. AMR does not develop an algorithm for compositionally translating sentences to representations. However, the AMR guidelines develop a thorough specification that guides the annotation process: AMR results in a semantic annotation of a sentence more than in a semantic representation of one. There are some robust semantic parsers for AMR, such as <https://github.com/IBM/transition-amr-parser> and <https://bollin.inf.ed.ac.uk/amreager.html> [Damonte et al., 2017]. [Bonial et al., 2019, Bonial et al., 2020] present extensions of AMR for dialogue, adding annotations for speech acts and resulting in a full annotation schema for dialogue.

## 1.4 Discussion and comparison in a dialogue modeling perspective

We introduce major phenomena from linguistics and logic that will be discussed in this thesis. Then, we present formalisms and frameworks that manage these phenomena to various degrees. This current section draws a comparison between these formalisms in the perspective of formal semantics modeling of dialogue.

---

<sup>25</sup>Source: <https://github.com/amrisi/amr-guidelines/blob/master/amr.md#questions> at 06/05/20, and

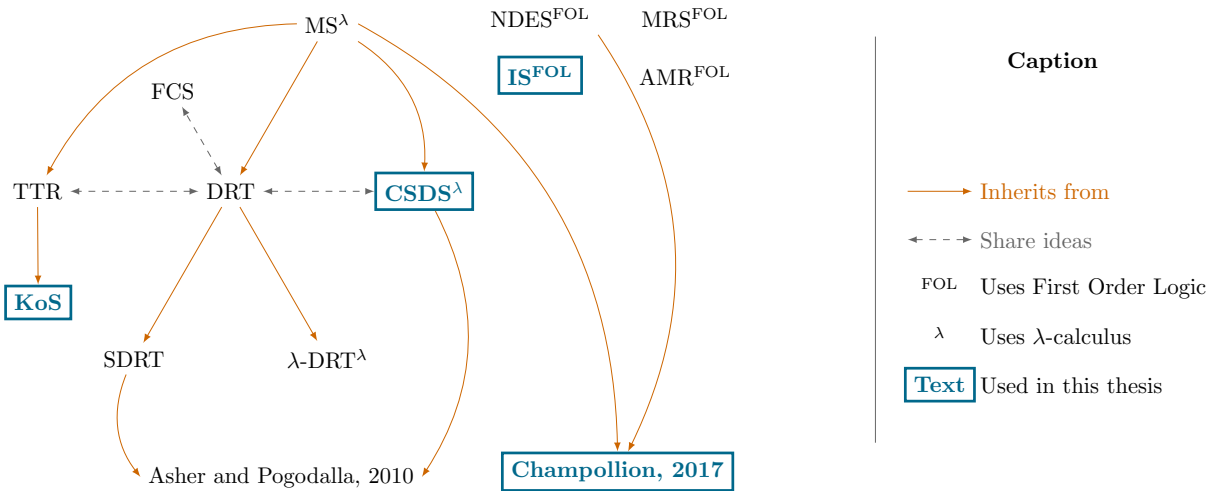


Figure 1.2: Genealogical graph of the formal semantics frameworks and formalisms presented in this thesis.

Figure 1.2 shows a visual summary of the links between the formalisms presented in the previous section. Montague Semantics (MS) was the starting point for formal semantics frameworks such as Discourse Representation Theory (DRT), Type Theory with Records (TTR), and Continuation Style Dynamic Semantics (CSDS). File Change Semantics (FCS) and DRT share foundational ideas about context modeling. TTR and CSDS include insights from DRT. DRT spawned approaches such as  $\lambda$ -DRT and Segmented Discourse Representation Theory (SDRT). TTR resulted in KoS.

First order logic (FOL) is the main formal inspiration of frameworks such as Neo-Davidsonian Event Semantics (NDES), Minimal Recursion Semantics (MRS), Inquisitive Semantics (IS), Abstract Meaning Representation (AMR). MS,  $\lambda$ -DRT, and CSDS make use of  $\lambda$ -calculus. Mixed approaches were undertaken for CSDS and SDRT, resulting in [Asher and Pogodalla, 2010], and for MS and NDES, resulting in [Champollion, 2017]. This thesis undertakes the mingling of CSDS, quantified NDES, and IS, adding insights from KoS.

CSDS offers a compositional and dynamic way to compute the semantic representations of declarative sentences. As it is anchored in  $\lambda$ -calculus, it is computationally and logically robust. Quantified NDES presents two major advantages: NDES gives easy access to the different semantic constituents of a sentence, and the quantified version of NDES takes care of the logical difficulties that can arise with quantified sentences in more classical versions of NDES. The combination of CSDS and quantified NDES should allow us to keep the advantages of both approaches and result in a compositional, dynamic, and logically robust framework.

As dialogue is the main subject of our study, we need to be able to account for dialogue structures and for different types of utterances, in particular, for questions. IS takes into account declarative and interrogative propositions without distinction by definition, thus giving us a way to account for different types of utterances in dialogue in a uniform way. KoS outlines the interactions occurring throughout a conversation, therefore giving us guidelines for the semantic modeling of dialogue. The combination of CSDS, quantified NDES, and IS should result in a compositional, dynamic, and logically robust framework with a uniform treatment of declarative

---

[Banarescu et al., 2013].

and interrogative sentences in dialogue, while the addition of insights from KoS should provide us with strong guidelines for taking dialogue structure and dialogue interactions into account.

	MS	FCS	DRT	SDRT	CSDS	TTR	NDES	MRS	IS	AMR
<b>Syntax</b>	✓	Part.	✗	With CSDS	✓	With HPSG	Champollion	✓	✗	✗
<b>Semantics</b>	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
<b>Compositionality</b>	✓	✗	Part. originally, then $\lambda$ -DRT	Part.	✓	Part.	Champollion	✓	Typed IS	✗
<b>Syntax-semantics interface</b>	✓	✗	✗	With CSDS	✓	✗	Champollion	✓	✗	✗
<b>Quantification &amp; negation</b>	Sentence-level	Part.	Part.	✓	✓	✓	✗ Bernard & Champollion	✓	✓	Part.
<b>Anaphora</b>	✗	✓	✓	✓	✓	✗	Champollion	✗	DIS	✗
<b>Dynamicity</b>	✗	✓	✓	✓	✓	✓	Part.	✗	DIS	✗
<b>Context</b>	✗	✓	✓	✓	✓	✓	Champollion	✗	✗	✗
<b>Discourse relations</b>	✗	✗	✗	✓	✗	KoS	✗	✗	✗	Part.
<b>Interrogatives</b>	✗	✗	✗	✓	✗	KoS	Boritchev & Amblard	UMRS	✓	✓
<b>Dialogue</b>	✗	✗	✗	✓	Boritchev & de Groote	KoS	✗	Part. with RMRS-IP	✗	Dialogue AMR

Table 1.6: Semantic formalisms and their features, the ✓ stand for “addresses the phenomenon”, ✗ for “doesn’t address it”.

Table 1.6 sums up the characteristics of the frameworks and formalisms presented in section 1.3 with respect to the linguistic and logic phenomena introduced in section 1.2. All the formalisms from our overview present a way to model semantics of natural language, and all of them have at least a partial treatment of logical quantification and negation. SDRT (sometimes partly, sometimes with variations) treats all the phenomena considered in the table. SDRT is originally concerned with coherence of discourse and dialogue, and semantics-pragmatics interface. Intermingling with CSDS gives SDRT the missing tools for the syntax-semantics interface.

CSDS, on its own, rises up to all the phenomena but dialogue-related ones; Champollion’s version of NDES does the same, though it doesn’t treat dynamicity as thoroughly as CSDS does.

The compositional formalisms mostly have a syntax-semantics interface, although it is not the case of DRT and IS, as they do not handle syntax at all. MRS, CSDS, and NDES all present different but possible ways of modeling syntax and semantics, compositionally, with a syntax-semantics interface. On the contrary, TTR and AMR come from a completely different perspective on semantic modeling.

Context, anaphora, and dynamicity are related, as we showed in 1.2. FCS, DRT, SDRT, CSDS, and NDES are frameworks within which all three phenomena are represented. It is interesting to note that DIS (the dynamic version of IS) achieves dynamicity without taking care of a specific context modeling and anaphora resolution; similarly, TTR is context-centered and

very concerned with dynamicity issues, but as TTR is not linguistic-based, it doesn't take care of anaphora.

SDRT, TTR (through KoS), and AMR are the three formalisms in our overview that are able to represent dialogue-related phenomena, interrogatives, and discourse relations. AMR lacks syntactical and compositional aspects, achieved in SDRT through combination with CSDS. MRS checks on all aspects of the syntax and semantics as well as for interrogatives and dialogue, with modifications. Yet, MRS has no approach for context-related phenomena.

We saw that SDRT fulfills all the criteria that we considered in this section. In a way, this thesis could end here, as the requirements we outlined for a formal semantics model of dialogue in natural language have been satisfied. The main argument for investigating alternative paths towards our goal comes from the fact that SDRT is, in design, more of a pragmatic model of discourse and dialogue. The models it produces are of higher level than the ones, syntax-semantics interface based, that we aim to obtain. Ideally, we would like for our model to be integrated in a complete syntax-semantics and semantics-pragmatics pipeline. Our question is then the following: if what we want to aim for is the first half of a complete pipeline model that starts from a transcription of a real-life dialogue and outputs a formal-semantics, logical, compositional, and dynamic representation, what is the best way to achieve this?

The models presented in Chapter 5 of this thesis constitute our attempt to answer this question. We use NDES and CSDS to produce a syntax-semantics interface with a good handle of quantification, negation, and context-related phenomena. IS gives us a robust way to represent both interrogative and declarative sentences. Insights from KoS modify the dynamic interactions with the context that are given by CSDS.

Therefore, we aim to fill three cells of table 1.6:

- Provide CSDS with a complete account of interrogative sentences.
- Provide NDES (Champollion's version) with a way to represent complete dialogues.
- Augment IS in view of dialogue modeling.

First, we give detailed presentations of CSDS, NDES and IS in the three following sections.

## 1.5 Continuation Style Dynamic Semantics

Montague semantics offers numerous expressivity advantages due to its compositionality, but it lacks dynamic notions. The work presented in [de Groote, 2006] and extended in [Lebedeva, 2012] shows a way to add dynamicity to Montague semantics by using simply-typed  $\lambda$ -calculus to represent meaning and  $\beta$ -reduction to introduce dynamicity. To do so, a parallel is drawn between the computation of the representation of a piece of discourse and the execution of a program by adapting the notion of *continuation* – a programming environment/context within which some code is executed – to natural language. Thus, the continuation of a piece of discourse corresponds to the bits coming right after that same piece of discourse and relying on it (though it can be empty) in order to be understood. The following presents Continuation Style Dynamic Semantics (CSDS)<sup>26</sup> as introduced in [de Groote, 2006].

---

<sup>26</sup>The name Type Theoretical Dynamic Logic (TTDL) can also be found in literature, see in particular [Qian, 2014].

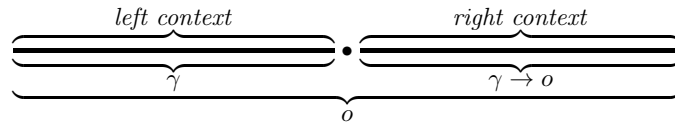
### 1.5.1 Formal set-up

Consider three atomic types:  $\iota, o, \gamma$ . ( $\iota, o$ ) are directly inherited from Church's simple theory of types [Church, 1940] and the third ( $\gamma$ ) is specifically introduced for the purpose of CSDS, to account for the notions of context. ( $\iota, o$ ) are equivalent to ( $e, t$ ) in Montague's tradition.

#### Definition 3 (Basic types, CSDS)

- $\iota$  : individual/entity
- $o$  : proposition
- $\gamma$  : left context

Then, a sentence is represented as a combination of its left and right contexts. As a sentence's semantic representation is a proposition, it is of type  $o$ . Therefore, the type of the right context is  $\gamma \rightarrow o$ , which is consistent with the denotational semantics idea of continuation introduced above.



The sentence under consideration is represented by the central dot, it is surrounded by its left context, of type  $\gamma$ , and its right context, of type  $\gamma \rightarrow o$ . The total computation of the sentence's representation is of type  $o$ , the type of propositions.

Figure 1.3: Schematic view of a sentence in CSDS perspective.

A discourse is built dynamically by combining its left context (the previous discourse) with its right context (the discourse yet to come). Thus, discourse is constructed in the same way as a single sentence, resulting in definition 4.

**Definition 4** Let  $s$  be the syntactic category of sentences,  $d$  be the syntactic category of discourses. Then, the corresponding semantic interpretations are:

$$\llbracket s \rrbracket = \gamma \rightarrow (\gamma \rightarrow o) \rightarrow o$$

$$\llbracket d \rrbracket = \gamma \rightarrow (\gamma \rightarrow o) \rightarrow o$$

This definition does not give a constructive way to compute the representation of the meaning of discourse by composing meanings of its constituent sentences. To address that issue, the linear aggregation operator  $.$  is defined as follows:

#### Definition 5 (Sentence composition)

Let  $D$  be a piece of discourse,  $S$  be a sentence. Then, the semantics of the discourse composed of  $D$  followed by  $S$  is given by:

$$\llbracket D.S \rrbracket := \lambda e \phi. \llbracket D \rrbracket e (\lambda e' \phi. \llbracket S \rrbracket e' \phi)$$

The term takes two arguments:  $e$  of type  $\gamma$ , left context of  $(D.S)$ , and  $\phi$  of type  $\gamma \rightarrow o$ , right context of  $(D.S)$ . The result is then composed of the representation of  $D$  combined with  $e$  (left context of  $(D.S)$ ) and a third term. This third term corresponds to the newly computed continuation of  $D$ , as it takes as argument  $e'$ , the left context of  $S$  (combination of  $e$  and the content of  $D$ ) and returns a result composed of the representation of  $S$  combined with  $\phi$ , the right context of  $(D.S)$  and thus of  $S$ .

Yet, sentence composition is not sufficient to address the issues of handling operators in a dynamic context and anaphora solving. To handle these phenomena, it is necessary to define objects that dynamically update and access the context:

### Definition 6 (Context aggregation operator and oracles)

Let  $c$  be a context (of type  $\gamma$ ). Hereafter we consider that  $c$  can be viewed as a finite set of individuals. The context aggregation operator, necessary in order to add new individuals to the context set, is typed as follows:

$$\_ :: \_ : \iota \rightarrow \gamma \rightarrow \gamma$$

$\_ :: \_$  takes two arguments: a new individual of type  $\iota$  and the context  $c$ . It outputs the updated context.

Conversely, if  $x$  is a pronoun, we type the corresponding oracle (a choice operator retrieving the right individual for the anaphora solving from the context) as follows:

$$sel_x : \gamma \rightarrow \iota$$

$sel_x$  takes the context  $c$  as input and retrieves from it an individual corresponding to the pronoun  $x$ . For example, if  $c = (Mary :: John)$ ,  $sel_{her}(c) = Mary$ <sup>27</sup>.

The following work is presented under the hypothesis of the existence and good behavior of these oracles. We do not discuss here the technical issues of anaphora resolution in a real-life setting.

CSDS's semantic representation displays all the advantages of Montague's approach combined with the newly possible handling of dynamic aspects. As this section shows, CSDS allows anaphora solving by considering discourse in a dynamic way. Similarly, it is possible to handle dynamic quantification.

## 1.5.2 CSDS – an example

We illustrate CSDS on the following toy example:

**Example 16** *Charlie hears Sasha. She answers him.*

For the sake of simplicity, we consider that the left context is empty when the example occurs. We start by decomposing this example in a discourse part  $D$ , corresponding to the first sentence, and a sentence part  $S$ , corresponding to the second one.

$$D = \text{Charlie hears Sasha.}$$

$$S = \text{She answers him.}$$

Then, we compute  $[[D.S]]$  following the previously given definition of the linear aggregation operator (see Definition 5). To properly conduct this computation, we follow the syntactic structure of the sentences.

<sup>27</sup>Note: the introductory article was written in 2001, pronoun situation in English was different then.



## Constituent analysis

First, we need to give the syntactic categories of the lexical constituents of  $D$  and  $S$ , see Section 1.2.1 for background. After that, we define the types of corresponding  $\lambda$ -terms, which helps us to give the proper semantic representations of the constituents of  $D$  and  $S$ .

## Syntactic categories

$$\begin{aligned} \text{Charlie, Sasha, she, him} & : NP \\ \text{hears, answers} & : TV = NP \rightarrow NP \rightarrow S \end{aligned}$$

Where  $NP$  stands for the syntactic category of noun phrases,  $TV$  the syntactic category of transitive verbs, and  $S$  the syntactic category of sentences.

## Typing

Then, we define the types corresponding to the syntactic categories.

$$\begin{aligned} \llbracket s \rrbracket & = \gamma \rightarrow (\gamma \rightarrow o) \rightarrow o \\ \llbracket np \rrbracket & = (\iota \rightarrow \llbracket s \rrbracket) \rightarrow \llbracket s \rrbracket \\ \llbracket tv \rrbracket & = \llbracket np \rightarrow np \rightarrow s \rrbracket = \llbracket np \rrbracket \rightarrow \llbracket np \rrbracket \rightarrow \llbracket s \rrbracket \end{aligned}$$

Where  $\llbracket s \rrbracket$  is the type of sentences,  $\llbracket np \rrbracket$  the type of noun phrases, and  $\llbracket tv \rrbracket$  the type of transitive verbs.

## Semantic representation

$$\begin{aligned} \llbracket \text{Charlie} \rrbracket & = \lambda\psi e. \psi \mathbf{c}(\mathbf{c} :: e) \\ \llbracket \text{Sasha} \rrbracket & = \lambda\psi e. \psi \mathbf{s}(\mathbf{s} :: e) \\ \llbracket \text{she} \rrbracket & = \lambda\psi e \phi. \psi(\text{sel}_{\text{she}} e) e \phi \\ \llbracket \text{him} \rrbracket & = \llbracket \text{he} \rrbracket = \lambda\psi e \phi. \psi(\text{sel}_{\text{he}} e) e \phi \end{aligned}$$

**Remark.** The semantic representation of “him” and “he” coincide in this case because of the chosen example.

$$\begin{aligned} \llbracket \text{hears} \rrbracket & = \lambda o s e \phi. s(\lambda x e. o(\lambda y e. \mathbf{hear} \ x \ y \wedge \phi e) e) e \\ \llbracket \text{answers} \rrbracket & = \lambda o s e \phi. s(\lambda x e. o(\lambda y e. \mathbf{answer} \ x \ y \wedge \phi e) e) e \end{aligned}$$

These  $\lambda$ -terms are read as follows: each is a term that awaits for four entities,  $o, s, e, \phi$ .  $o$  and  $s$  refer, respectively, to the *object* and the *subject* of the transitive verb (“hear” or “answer”).  $e$  and  $\phi$  refer, respectively, to the left and right contexts of the sentence under consideration.



## Computation

First, we compute the semantic representation of  $D$ , according to the syntactic structure of the sentence.

Semantic representations of the intermediary nodes of the tree are computed bottom-up. According to the types of  $\llbracket tv \rrbracket$  and  $\llbracket np \rrbracket$ ,  $\llbracket hears \rrbracket$  is the functor applied to the argument  $\llbracket Sasha \rrbracket$ . Similarly, typing gives us that  $\llbracket hears \rrbracket \llbracket Sasha \rrbracket$  is the functor applied to the argument  $\llbracket Charlie \rrbracket$ .

Unfolding the term by using  $\beta$ -reduction, we obtain the following computation:

$$\begin{aligned}
 \llbracket hears \rrbracket \llbracket Sasha \rrbracket \llbracket Charlie \rrbracket &= \left( \lambda o s e \phi . s (\lambda x e . o (\lambda y e . \mathbf{hear} \ xy \wedge \phi e) e) e \right) \llbracket Sasha \rrbracket \llbracket Charlie \rrbracket \\
 &\rightarrow_{\beta} \left( \lambda s e \phi . s (\lambda x e . \llbracket Sasha \rrbracket (\lambda y e . \mathbf{hear} \ xy \wedge \phi e) e) e \right) \llbracket Charlie \rrbracket \\
 &\rightarrow_{\beta} \left( \lambda e \phi . \llbracket Charlie \rrbracket (\lambda x e . \llbracket Sasha \rrbracket (\lambda y e . \mathbf{hear} \ xy \wedge \phi e) e) e \right) \\
 &= \left( \lambda e \phi . (\lambda \psi e . \psi \mathbf{c} (\mathbf{c} :: e)) (\lambda x e . \llbracket Sasha \rrbracket (\lambda y e . \mathbf{hear} \ xy \wedge \phi e) e) e \right) \\
 &\rightarrow_{\beta} \left( \lambda e \phi . (\lambda e . (\lambda x e . \llbracket Sasha \rrbracket (\lambda y e . \mathbf{hear} \ xy \wedge \phi e) e) \mathbf{c} (\mathbf{c} :: e)) e \right) \\
 &\rightarrow_{\beta} \left( \lambda e \phi . (\lambda x e . \llbracket Sasha \rrbracket (\lambda y e . \mathbf{hear} \ xy \wedge \phi e) e) \mathbf{c} (\mathbf{c} :: e) \right) \\
 &\rightarrow_{\beta} \left( \lambda e \phi . (\lambda e . \llbracket Sasha \rrbracket (\lambda y e . \mathbf{hear} \ cy \wedge \phi e) e) (\mathbf{c} :: e) \right) \\
 &\rightarrow_{\beta} \left( \lambda e \phi . \llbracket Sasha \rrbracket (\lambda y e . \mathbf{hear} \ cy \wedge \phi e) (\mathbf{c} :: e) \right) \\
 &= \left( \lambda e \phi . (\lambda \psi e . \psi \mathbf{s} (\mathbf{s} :: e)) (\lambda y e . \mathbf{hear} \ cy \wedge \phi e) (\mathbf{c} :: e) \right) \\
 &\rightarrow_{\beta} \left( \lambda e \phi . (\lambda e . (\lambda y e . \mathbf{hear} \ cy \wedge \phi e) \mathbf{s} (\mathbf{s} :: e)) (\mathbf{c} :: e) \right) \\
 &\rightarrow_{\beta} \left( \lambda e \phi . (\lambda y e . \mathbf{hear} \ cy \wedge \phi e) \mathbf{s} (\mathbf{s} :: \mathbf{c} :: e) \right) \\
 &\rightarrow_{\beta} \left( \lambda e \phi . (\lambda e . \mathbf{hear} \ \mathbf{c} \ \mathbf{s} \wedge \phi e) (\mathbf{s} :: \mathbf{c} :: e) \right) \\
 &\rightarrow_{\beta} \left( \lambda e \phi . \mathbf{hear} \ \mathbf{c} \ \mathbf{s} \wedge \phi (\mathbf{s} :: \mathbf{c} :: e) \right)
 \end{aligned}$$

Thus,

$$\llbracket D \rrbracket = \lambda e \phi . \mathbf{hear} \ \mathbf{c} \ \mathbf{s} \wedge \phi (\mathbf{s} :: \mathbf{c} :: e)$$

Similarly, the computation of  $\llbracket S \rrbracket$  gives us:

$$\llbracket S \rrbracket = \lambda e \phi . \mathbf{answer} \ (sel_{she} e) (sel_{he} e) \wedge \phi e$$

Now, we can compute the semantics of  $\llbracket D.S \rrbracket$  :

$$\begin{aligned}
 \llbracket D.S \rrbracket &= \lambda e \phi. \llbracket D \rrbracket e \left( \lambda e'. \llbracket S \rrbracket e' \phi \right) \\
 &= \lambda e \phi. \left( \lambda e \phi. \mathbf{hear} \mathbf{c} \mathbf{s} \wedge \phi(\mathbf{s} :: \mathbf{c} :: e) \right) e \left( \lambda e'. \llbracket S \rrbracket e' \phi \right) \\
 &\rightarrow_{\beta} \lambda e \phi. \left( \lambda \phi. \mathbf{hear} \mathbf{c} \mathbf{s} \wedge \phi(\mathbf{s} :: \mathbf{c} :: e) \right) \left( \lambda e'. \llbracket S \rrbracket e' \phi \right) \\
 &\rightarrow_{\beta} \lambda e \phi. \left( \mathbf{hear} \mathbf{c} \mathbf{s} \wedge (\lambda e'. \llbracket S \rrbracket e' \phi)(\mathbf{s} :: \mathbf{c} :: e) \right) \\
 &\rightarrow_{\beta} \lambda e \phi. \left( \mathbf{hear} \mathbf{c} \mathbf{s} \wedge \llbracket S \rrbracket(\mathbf{s} :: \mathbf{c} :: e) \phi \right) \\
 &= \lambda e \phi. \left( \mathbf{hear} \mathbf{c} \mathbf{s} \wedge (\lambda e \phi. \mathbf{answer} (sel_{she} e)(sel_{he} e) \wedge \phi)(\mathbf{s} :: \mathbf{c} :: e) \phi \right) \\
 &\rightarrow_{\beta} \lambda e \phi. \left( \mathbf{hear} \mathbf{c} \mathbf{s} \wedge \left( \lambda \phi. \mathbf{answer} (sel_{she}(\mathbf{s} :: \mathbf{c} :: e)) (sel_{he}(\mathbf{s} :: \mathbf{c} :: e)) \wedge \phi(\mathbf{s} :: \mathbf{c} :: e) \right) \phi \right) \\
 &\rightarrow_{\beta} \lambda e \phi. \left( \mathbf{hear} \mathbf{c} \mathbf{s} \wedge \mathbf{answer} (sel_{she}(\mathbf{s} :: \mathbf{c} :: e)) (sel_{he}(\mathbf{s} :: \mathbf{c} :: e)) \wedge \phi(\mathbf{s} :: \mathbf{c} :: e) \right)
 \end{aligned}$$

Unfolding the selection operator, we have:

$$\begin{aligned}
 (sel_{she}(\mathbf{s} :: \mathbf{c} :: e)) &= \mathbf{c} \\
 (sel_{he}(\mathbf{s} :: \mathbf{c} :: e)) &= \mathbf{s}
 \end{aligned}$$

Which gives us the desired final representation of D.S:

$$\llbracket D.S \rrbracket = \lambda e \phi. \mathbf{hear} \mathbf{c} \mathbf{s} \wedge \mathbf{answer} \mathbf{c} \mathbf{s} \wedge \phi(\mathbf{s} :: \mathbf{c} :: e)$$

The computation we unraveled in this section is quite long and prone to mistakes. The next section presents the Abstract Categorical Grammars, a grammatical formalism that, once implemented, allows for an automatic computation of CSDS-based semantic representations.

### 1.5.3 Abstract Categorical Grammar

The Abstract Categorical Grammars (ACGs) are a type-theoretic grammatical formalism for natural languages, where the semantic representations are expressed as typed  $\lambda$ -terms. The main principle of ACGs is to use the same type theory for both syntax and semantics [de Groote, 2001]. The grammars are then defined using *abstract signatures* linked by a *lexicon*.

The following presentation loosely follows the ones from [de Groote, 2001] and [Maršík, 2016]. Formally, an ACG is defined on a set of atomic types  $A$ . Then, the set  $\mathcal{T}(A)$  of *linear implicative types*, i.e types corresponding to implicative fragments of linear logic [Girard, 1987], built upon  $A$  is inductively defined as follows:

1. if  $a \in A$ , then  $a \in \mathcal{T}(A)$ ;
2. if  $\alpha, \beta \in A$ , then  $(\alpha \rightarrow \beta) \in \mathcal{T}(A)$ .

Then, we can define *higher order linear signatures*.

**Definition 7 (Higher order linear signature)**

A higher order linear signature is defined as a triple  $\Sigma = \langle A, C, \tau \rangle$  where:

1.  $A$  is a finite set of atomic types;
2.  $C$  a finite set of constants;
3.  $\tau : C \rightarrow \mathcal{T}(A)$  a function that assigns to each constant in  $C$  a linear implicative type in  $\mathcal{T}(A)$ .

Then, we define a lexicon, build upon higher order linear signatures.

**Definition 8 (Lexicon between two higher order linear signatures)**

Let  $\Sigma_1 = \langle A_1, C_1, \tau_1 \rangle$  and  $\Sigma_2 = \langle A_2, C_2, \tau_2 \rangle$  be two higher order linear signatures, called vocabularies. Then, a lexicon  $\mathcal{L}$  from  $\Sigma_1$  to  $\Sigma_2$  is defined as a pair  $\mathcal{L} = \langle F, G \rangle$  such that:

1.  $F : A_1 \rightarrow \mathcal{T}(A_2)$  is a function that interprets the atomic types of  $\Sigma_1$  as linear implicative types built upon  $A_2$ ;
2.  $G : C_1 \rightarrow \Lambda(\Sigma_2)$  is a function that interprets the constants of  $\Sigma_1$  as linear  $\lambda$ -terms built upon  $\Sigma_2$ .
3. The interpretation functions are compatible with the typing relation.

Once we have defined the signatures and the lexicon, we can give the formal definition of an abstract categorial grammar.

**Definition 9 (ACG)**

An abstract categorial grammar (ACG) is a quadruple  $\mathcal{G} = \langle \Sigma_1, \Sigma_2, \mathcal{L}, s \rangle$  where:

1.  $\Sigma_1 = \langle A_1, C_1, \tau_1 \rangle$  and  $\Sigma_2 = \langle A_2, C_2, \tau_2 \rangle$  are two higher order linear signatures called the abstract vocabulary and the object vocabulary, respectively;
2.  $\mathcal{L} : \Sigma_1 \rightarrow \Sigma_2$  is a lexicon from the abstract vocabulary to the object vocabulary;
3.  $s \in \mathcal{T}(A_1)$  is a type of the abstract vocabulary called the distinguished type of the grammar.

Figure 1.4 illustrates the structure of an ACG, with  $\Sigma_1$  and  $\Sigma_2$  two signatures and  $\mathcal{L}$  a lexicon.

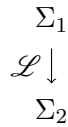


Figure 1.4: Schematic view of an ACG.

Any ACG generates two languages, an *abstract language*, and an *object language*. In the context of this thesis, the considered ACGs will constitute structures of the form presented in figure 1.5: one ACG linking an abstract syntactic lexicon with a lexicon of surface forms (roughly corresponding to natural language formulations); a second ACG linking an abstract syntactic lexicon with a semantic interpretation lexicon.

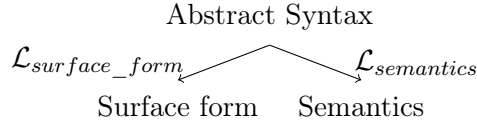


Figure 1.5: Schematic view of ACGs used for Natural Language Processing, in this thesis.

The ACGs framework is encoded in the ACG toolkit (ACGtk, [Pogodalla, 2016]). Documentation, installation instructions, and examples can be found here: <https://gitlab.inria.fr/ACG/dev/ACGtk>. CSDS is encoded in ACGtk, which allows us to compute the examples using CSDS automatically, through ACGtk.

CSDS is quite agnostic in terms of the underlying logical formation [de Groote, 2006]: it can be used to provide a dynamic extension to standard MS but also NDES or IS. The presentation in section 1.5 was made using MS. The following section presents in detail another semantic framework for natural language statements: NDES as presented by Lucas Champollion in [Champollion, 2017].

## 1.6 Quantificational event semantics

We begin this section by presenting the ideas of event semantics, [Davidson, 1967]. Then, we introduce the Neo-Davidsonian take on event semantics, [Parsons, 1990]. Finally, we show the compositional presentation of NDES given by Lucas Champollion in [Champollion, 2017], which is one of the theoretical pillars of the model presented in this thesis.

### 1.6.1 Event semantics – First order logic representations

A classical way to represent the meaning of natural language sentences is through the usage of First-Order Logic (FOL). [Montague and Thomason, 1975] uses FOL to represent the meaning of a verb with  $n$  syntactic arguments as an  $n$ -ary predicate.

#### Example 17 (Semantic representation of a sentence, MS)

$$\llbracket \textit{Brutus stabbed Caesar} \rrbracket = \textit{stab}(b, c)$$

In example 17, the verb “stabbed” is represented by the binary predicate **stab**. The subject (“Brutus”) and the object (“Caesar”) are represented respectively by the letter **b** and the letter **c**, arguments of the predicate **stab**.

One can see for example Chapter 15 of [Jurafsky and Martin, 2015] for a complete explanation of FOL meaning representation for natural language. In the following, we will focus on the application of FOL to the representation of events. First, we outline the notions of *state* and *event*. [Jurafsky and Martin, 2015] describe states and events as follows:

*States are conditions or properties, that remain unchanged over an extended period of time. Events denote changes in some states. [Jurafsky and Martin, 2015]*

[Davidson, 1967] presents a view of verbs as links between events and their arguments. FOL is then used to represent the meaning of a verb with  $n$  syntactic arguments as a predicate of arity  $n + 1$ , the new argument being the event, quantified existentially.

**Example 18 (Semantic representation of a sentence, Davidsonian event semantics)**

$$\llbracket \textit{Brutus stabbed Caesar} \rrbracket = \exists e.stab(e, b, c)$$

In example 18, the verb “stabbed” is represented by, first, the introduction of an event  $e$ , that has scope over the ternary predicate  $stab$ . The event, the subject ( $b$ ) and the object ( $c$ ) are the arguments of the predicate  $stab$ .

Both Montague’s and Davidsonian views are rigid with respect to the number of syntactic arguments of the verb. Yet, it is easy to find examples where it shouldn’t be the case. In the sentence “Charlie drank”, the predicate representing the verb “drink” takes one argument, “Charlie”, which gives us the montagovian representation  $drink(C)$  and davidsonian representation  $\exists e.drink(e, C)$ . In the sentence “Charlie drank tea”, the predicate representing the verb “drink” takes two arguments, “Charlie” and “tea”, which gives us the montagovian representation  $drink(C, t)$  and davidsonian representation  $\exists e.drink(e, C, t)$ . Thus, in montagovian and davidsonian representations, several predicates with different arities should be used in order to represent the semantics of the same verb, depending on the amount of optional arguments. A way to bypass this problem is to adopt the Neo-Davidsonian Event Semantics.

### 1.6.2 Neo-Davidsonian Event Semantics

In Neo Davidsonian Event Semantics, each event predicate has an existentially quantified variable as unique argument; all the following predicates can be related to this event predicate by taking the same variable as argument. This type of semantic representation is called Neo-Davidsonian Event representation, [Davidson, 1967, Parsons, 1990].

**Example 19 (Semantic representation of a sentence, NDES)**

$$\llbracket \textit{Brutus stabbed Caesar} \rrbracket = \exists e.stab(e) \wedge agent(e, b) \wedge theme(e, c)$$

This type of representation uses semantic/thematic<sup>28</sup> roles to link the event with its arguments and modifiers. [Jurafsky and Martin, 2015] present a list of commonly used thematic roles, see table 1.7. Another list of semantic arguments is given by the DIT++ [Bunt, 2009] annotation schema for dialogue, mostly used for English; part of DIT++ is composed of an extensive list of semantic arguments, see table 1.8 (at the end of this chapter).

As stated in [Bender and Lascarides, 2019], there is no fixed list of semantic roles, no crosslinguistic one, and even no English one. In the following, we will always specify the list that is used.

### 1.6.3 Compositional Neo-Davidsonian Event Semantics

NDES is combinatorially simple and human-readable, but traditional vision often lacks compositionality. [Champollion, 2017] presents a compositional version of NDES, produced by binding the event variables in a specific way and using closure operators to prevent scope ambiguities. Champollion’s vision considers that verbs denote generalized quantifiers over events; this framework is then called *quantificational event semantics*. More formally, consider three atomic types:  $e, t, v$ . ( $e, t$ ) are directly inherited from Montague Semantics [Montague, 1970]. The third ( $v$ ) is introduced in order to account for events in a separate way.

---

<sup>28</sup>Both names can be found in literature.

Thematic Role	Definition
Agent	The volitional causer of an event
Experiencer	The experiencer of an event
Force	The non-volitional causer of an event
Theme	The participant most directly affected by an event
Result	The end product of an event
Content	The proposition or content of a propositional event
Instrument	An instrument used in an event
Beneficiary	The beneficiary of an event
Source	The origin of the object of a transfer event
Goal	The destination of an object of a transfer event

Table 1.7: Some commonly used thematic roles with their definitions, [Jurafsky and Martin, 2015], Chapter 19.

#### Definition 10 (Basic types, NDES)

- $e$  : individual/entity
- $t$  : truth values
- $v$  : event
- if  $\alpha$  and  $\beta$  are types, then  $\alpha \rightarrow \beta$  is a type

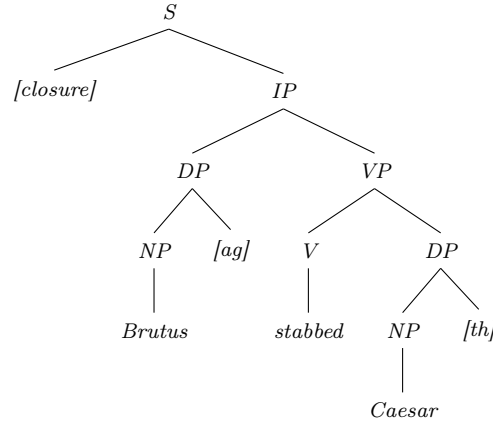
The semantic representation of a sentence  $S$  (of type  $t$ ) is then built through the following process:

1. syntactic parsing of  $S$  ;
2. building of the deep syntactic tree of  $S$  (also called *logical form* of  $S$ ) ;
3. translation of the elements of the tree to their semantic representations through the syntax-semantics interface ;
4. composition of the elements of the tree following the types constraints of the semantic representations and syntactic tree structure.

We will not engage in the syntactic discussion of the more suitable way to parse a sentence here. In the following, we will consider that the syntactic parsing is given to us (*e.g.* by the Stanford parser<sup>29</sup>, [Klein and Manning, 2003, Klein et al., 2003]). Example 21 shows a simple tree resulting from the syntactic parsing of the sentence “Brutus stabbed Caesar”.

<sup>29</sup><https://stanfordnlp.github.io/CoreNLP/>, visited on 18/06/2021.

**Example 20** (Deep syntactic tree for the sentence “Brutus stabbed Caesar”)



Where  $[ag]$  stands for the thematic role **agent** and  $[th]$  for the thematic role **theme**. Then, we define the semantic types corresponding to these syntactic categories.

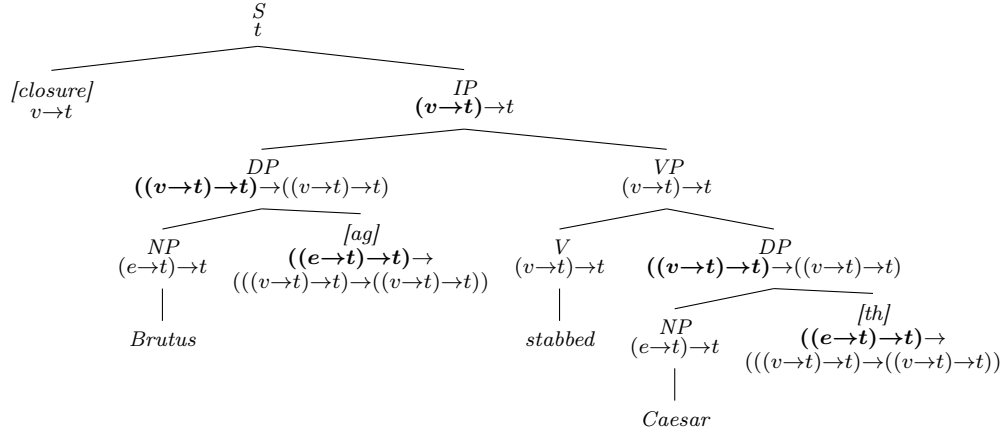
**Definition 11** (Basic syntax-semantics interface)

$$\begin{aligned}
 S &: t \\
 [closure] &: v \rightarrow t \\
 IP &: (v \rightarrow t) \rightarrow t \\
 DP &: ((v \rightarrow t) \rightarrow t) \rightarrow ((v \rightarrow t) \rightarrow t) \\
 VP &: (v \rightarrow t) \rightarrow t \\
 NP &: (e \rightarrow t) \rightarrow t \\
 N &: e \rightarrow t \\
 V &: (v \rightarrow t) \rightarrow t \\
 [agent] &: ((e \rightarrow t) \rightarrow t) \rightarrow (((v \rightarrow t) \rightarrow t) \rightarrow ((v \rightarrow t) \rightarrow t)) \\
 [theme] &: ((e \rightarrow t) \rightarrow t) \rightarrow (((v \rightarrow t) \rightarrow t) \rightarrow ((v \rightarrow t) \rightarrow t))
 \end{aligned}$$

$$\begin{aligned}
 \llbracket [closure] \rrbracket &= \lambda e. \top \\
 \llbracket Brutus \rrbracket &= \lambda P. P(brutus) \\
 \llbracket stabbed \rrbracket &= \lambda f. \exists e [stab(e) \wedge f(e)] \\
 \llbracket Caesar \rrbracket &= \lambda P. P(caesar) \\
 \llbracket [ag] \rrbracket &= \lambda Q. \lambda V. \lambda f. Q(\lambda x. V(\lambda e. [f(e) \wedge ag(e) = x])) \\
 \llbracket [th] \rrbracket &= \lambda Q. \lambda V. \lambda f. Q(\lambda x. V(\lambda e. [f(e) \wedge th(e) = x]))
 \end{aligned}$$

The elements of the tree are then composed following the types constraints introduced by the syntax-semantics interface and the tree structure.

**Example 21** (Deep syntactic tree for the sentence “Brutus stabbed Caesar”, with types)



The types give us the order in which to compute the semantic representation of the sentence “Brutus stabbed Caesar”: the  $\lambda$ -term representing **theme** ( $[th]$ ) is the functor and the  $\lambda$ -term representing **Caesar** is the argument, because of their respective types. This gives us the following outline for the computation:

$$\llbracket \text{Brutus stabbed Caesar} \rrbracket = \left( \left( \llbracket [ag] \rrbracket \llbracket Brutus \rrbracket \right) \left( \left( \llbracket [th] \rrbracket \llbracket Caesar \rrbracket \right) \llbracket stabbed \rrbracket \right) \right) \llbracket [closure] \rrbracket$$

**Example 22** (Computation of the semantic representation of “Brutus stabbed Caesar”)

$$\begin{aligned} & \llbracket [ag] \rrbracket \llbracket Brutus \rrbracket \\ &= \left[ \lambda Q. \lambda V. \lambda f. Q(\lambda x. V(\lambda e. [f(e) \wedge ag(e) = x])) \right] \left[ \lambda P. P(brutus) \right] \\ &\rightarrow_{\beta} \lambda V. \lambda f. \left[ \lambda P. P(brutus) \right] \left[ (\lambda x. V(\lambda e. [f(e) \wedge ag(e) = x])) \right] \\ &\rightarrow_{\beta} \lambda V. \lambda f. \left[ \lambda x. V(\lambda e. [f(e) \wedge ag(e) = x]) \right] \left[ brutus \right] \\ &\rightarrow_{\beta} \lambda V. \lambda f. V(\lambda e. [f(e) \wedge ag(e) = brutus]) = \textcircled{1} \end{aligned}$$

Similarly,

$$\llbracket [th] \rrbracket \llbracket Caesar \rrbracket = \lambda V. \lambda f. V(\lambda e. [f(e) \wedge th(e) = caesar]) = \textcircled{2}$$

$$\begin{aligned} & \textcircled{2} \llbracket stabbed \rrbracket \\ &= \left[ \lambda V. \lambda f. V(\lambda e. [f(e) \wedge th(e) = caesar]) \right] \left[ \lambda f. \exists e [stab(e) \wedge f(e)] \right] \\ &\rightarrow_{\beta} \lambda f. \left[ \lambda f. \exists e [stab(e) \wedge f(e)] \right] \left[ (\lambda e. [f(e) \wedge th(e) = caesar]) \right] \\ &\rightarrow_{\beta} \lambda f. \exists e [stab(e) \wedge \left[ (\lambda e. [f(e) \wedge th(e) = caesar]) \right] \left[ e \right]] \\ &\rightarrow_{\beta} \lambda f. \exists e [stab(e) \wedge [f(e) \wedge th(e) = caesar]] = \textcircled{3} \end{aligned}$$



$$\begin{aligned}
 \textcircled{1} \textcircled{3} &= \left[ \lambda V. \lambda f. V(\lambda e. [f(e) \wedge ag(e) = brutus]) \right] \left[ \lambda f. \exists e [stab(e) \wedge [f(e) \wedge th(e) = caesar]] \right] \\
 &\rightarrow_{\beta} \lambda f. \left[ \lambda f. \exists e [stab(e) \wedge [f(e) \wedge th(e) = caesar]] \right] \left[ (\lambda e. [f(e) \wedge ag(e) = brutus]) \right] \\
 &\rightarrow_{\beta} \lambda f. \exists e [stab(e) \wedge \left[ \lambda e. [f(e) \wedge ag(e) = brutus] \right] \left[ e \right] \wedge th(e) = caesar] \\
 &\rightarrow_{\beta} \lambda f. \exists e [stab(e) \wedge [f(e) \wedge ag(e) = brutus] \wedge th(e) = caesar] = \textcircled{4}
 \end{aligned}$$

$$\begin{aligned}
 \textcircled{4} \llbracket [closure] \rrbracket &= \left[ \lambda f. \exists e [stab(e) \wedge [f(e) \wedge ag(e) = brutus] \wedge th(e) = caesar] \right] \left[ \lambda e. \top \right] \\
 &\rightarrow_{\beta} \exists e [stab(e) \wedge \left[ \lambda e. \top \right] \left[ e \right] \wedge ag(e) = brutus] \wedge th(e) = caesar] \\
 &\rightarrow_{\beta} \exists e [stab(e) \wedge [\top \wedge ag(e) = brutus] \wedge th(e) = caesar] \\
 &= \exists e [stab(e) \wedge ag(e) = brutus \wedge th(e) = caesar]
 \end{aligned}$$

Which concludes the computation.

The computation presented in example 22 produces a semantic representation that is equivalent to the one from example 19 up to a rewriting:  $ag(e) = brutus$  is equivalent to  $agent(e, b)$ . Quantificational event semantics results in semantic representations that are equivalent to the ones of traditional Neo-Davidsonian Event Semantics, but it achieves the result through a compositional process with a syntax-semantics interface.

Examples from papers such as [Champollion, 2011] and [Champollion, 2015] are implemented on the Lambda Calculator<sup>30</sup>, [Champollion et al., 2007]. The Lambda Calculator is an interactive graphical application for semantic representation derivation computation following syntactic trees.

This presentation of NDES is compositional by construction. It keeps NDES's advantages such as combinatorial simplicity, human readability, and, more formally, ease of access to semantic roles-related variables. In addition, the compositionality of Champollion's version offers a direct road to the computational operationalization of methods for semantic representation.

## 1.7 Inquisitive Semantics

We have seen that in FOL, a proposition is interpreted as a truth value. In traditional modal logic, a proposition is interpreted as a set of possible worlds. Inquisitive Semantics interprets a proposition as a set of sets of possible worlds. Intuitively, an inquisitive proposition can be seen as a set of classical (modal) propositions. This allows questions to be assigned semantics akin to Hamblin's alternative semantics [Hamblin, 1976].

Formally, in inquisitive logic, a *proposition* is defined to be a non-empty set of sets of possible worlds that is downward-closed with respect to set inclusion. As a consequence, conjunction, disjunction, and entailment can be defined in a standard way, i.e., as intersection, union, and inclusion, respectively. Let us illustrate this by an example.

Consider a person called Cheryl. Suppose it is known that Cheryl's birthday is either May 15th, June 17th, or July 14th. Accordingly, we define a set of possible worlds,  $W =$

---

<sup>30</sup><http://lambdacalculator.com/>, visited on 18/06/2021.

$\{w_{5.15}, w_{6.17}, w_{7.14}\}$ , where each possible world corresponds respectively to one of Cheryl's possible birthdates. Then, the proposition  $\varphi_1$  that *Cheryl is born on May 15th* is interpreted as follows:

$$\llbracket \varphi_1 \rrbracket = \{\{w_{5.15}\}, \emptyset\} \quad (1.6)$$

The proposition  $\varphi_2$  that *she is born on June 17th* is interpreted in a similar way:

$$\llbracket \varphi_2 \rrbracket = \{\{w_{6.17}\}, \emptyset\} \quad (1.7)$$

Then, the inquisitive disjunction of  $\varphi_1$  and  $\varphi_2$  is interpreted as the union of their interpretations:

$$\llbracket \varphi_1 \vee_i \varphi_2 \rrbracket = \{\{w_{5.15}\}, \{w_{6.17}\}, \emptyset\} \quad (1.8)$$

This disjunction does not correspond to a proposition asserting that Cheryl's birthday is either May 15th or June 17th, it rather corresponds to the question *whether Cheryl's birthday is May 15th or June 17th*, assuming that she is born on one of these two dates. The mere assertion that *her birthday is either May 15 or June 17* is interpreted in a different way:

$$\{\{w_{5.15}, w_{6.17}\}, \{w_{5.15}\}, \{w_{6.17}\}, \emptyset\} \quad (1.9)$$

It is known that intensional logic can be defined by embedding it in Gallin's Ty2 [Gallin, 1975]. We provide here below a similar embedding for first-order inquisitive logic<sup>31</sup>:

$$\begin{aligned} \mathbf{R}_i t_1 \dots t_n &:= \mathcal{P}(\mathbf{R} t_1 \dots t_n) \\ &\text{where } \mathcal{P} a = \lambda b^{s \rightarrow t}. \forall w^s. (b w) \rightarrow (a w) \\ \varphi \wedge_i \psi &:= \lambda a^{s \rightarrow t}. (\varphi a) \wedge (\psi a) \\ \varphi \vee_i \psi &:= \lambda a^{s \rightarrow t}. (\varphi a) \vee (\psi a) \\ \varphi \rightarrow_i \psi &:= \lambda a^{s \rightarrow t}. \forall b^{s \rightarrow t}. (\forall w^s. (b w) \rightarrow (a w)) \rightarrow ((\varphi b) \rightarrow (\psi b)) \\ \neg_i \varphi &:= \lambda a^{s \rightarrow t}. \forall w^s. (a w) \rightarrow \neg(\exists b^{s \rightarrow t}. (\varphi b) \wedge (b w)) \\ \forall_i x^e. \varphi x &:= \lambda a^{s \rightarrow t}. \forall x^e. \varphi x a \\ \exists_i x^e. \varphi x &:= \lambda a^{s \rightarrow t}. \exists x^e. \varphi x a \end{aligned}$$

Inquisitive logic also features two *projection operators*, ! and ?. The first one transforms any proposition into a purely informative one by canceling its inquisitive content. Conversely, the second one transforms any proposition into a purely inquisitive one by canceling its informative content. These projection operators may be defined as follows:

$$\begin{aligned} !\varphi &:= \neg_i \neg_i \varphi \\ ?\varphi &:= \varphi \vee_i \neg_i \varphi \end{aligned}$$

Figure 1.6 shows the decomposition of a proposition in its non-informative and non-inquisitive projections.

These two operators are useful to turn a question into an assertion, and vice versa. For instance, by applying ! to (1.8), one obtains the proposition asserting that *Cheryl's birthday is either May 15 or June 17*:

$$\llbracket !( \varphi_1 \vee_i \varphi_2 ) \rrbracket = \{\{w_{5.15}, w_{6.17}\}, \{w_{5.15}\}, \{w_{6.17}\}, \emptyset\} \quad (1.10)$$

<sup>31</sup>Where  $\mathbf{s}$  is the type of possible worlds,  $\mathbf{t}$  is the type of truth values,  $\mathbf{e}$  is the type of individuals, following [Montague, 1970]

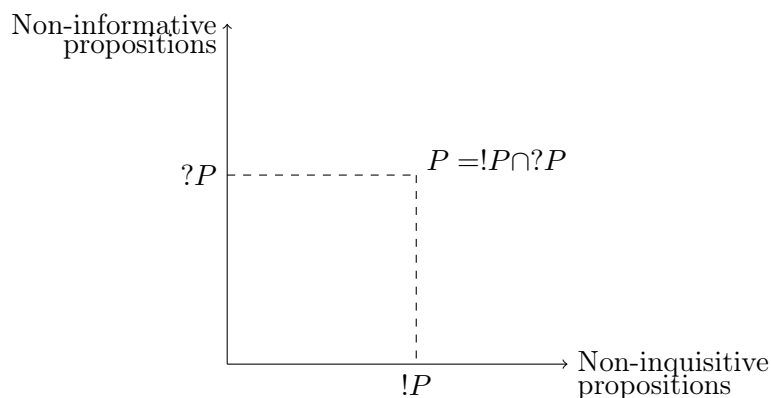


Figure 1.6: Propositions with informative and inquisitive content, projection operators, [Ciardelli et al., 2018].

On the other hand, by applying  $?$  to (1.6), one obtains an inquisitive proposition that corresponds to the issue *whether Cheryl is born on May 15 or not*:

$$\llbracket ?\varphi_1 \rrbracket = \{\{w_{6.17}, w_{7.14}\}, \{w_{5.15}\}, \{w_{6.17}\}, \{w_{7.14}\}, \emptyset\} \quad (1.11)$$

Inquisitive propositions being downward-closed sets, they are completely characterized by their maximal elements. In the sequel of this thesis, we will use the notation  $[a, b, c, \dots]$  to denote the downward-closure of the set  $\{a, b, c, \dots\}$ . With this convention, Example (1.11) can be rewritten as follows:

$$\llbracket ?\varphi_1 \rrbracket = [\{w_{6.17}, w_{7.14}\}, \{w_{5.15}\}] \quad (1.12)$$

Example (1.8) illustrates that an inquisitive proposition has both an informative and an inquisitive content. It is even the fact that every inquisitive proposition  $\phi$  may be defined as the conjunction of a purely informative proposition with a purely inquisitive one:

$$\varphi = !\varphi \wedge ?\varphi$$

Inquisitive Semantics offers a logically strong way to model declarative and interrogative sentences uniformly. Accordingly, a Montague-like semantic framework based on IS would also allow for this uniform treatment. In order to build such a compositional semantic framework for natural language, based on IS, we want to integrate it in a framework such as CSDS, following an approach similar to what [Asher and Pogodalla, 2010] did for SDRT.

## 1.8 Conclusion

This chapter presents the foundational ideas, phenomena, and formalisms that we need for the sequel of this thesis. We outlined logic and linguistic phenomena that we used to introduce logic-based approaches to natural language and highlight the need for a formal language model to be compositional and dynamic to account for frequent linguistic phenomena. We presented an overview of the existing formalisms and frameworks and compared them in a dialogue modeling perspective. This helped us to start to outline the desiderata for the model that is the object of this thesis: logic-based, compositional, and dynamic.

We want to follow the tradition set up by Montague-style semantics by linking syntax and semantics in a compositional way through the syntax-semantics interface by placing our approach within the Continuation Style Dynamic Semantics formalism, [de Groot, 2006]. This approach is insufficient to model interrogatives and in general, dialogue, so we follow the footsteps of [Groenendijk et al., 1997]. In order to do so, we integrate Inquisitive Semantics [Ciardelli et al., 2018] in our model. We try to characterize better the pragmatics of dialogue by making links with discourse. Then we go from discourse to dialogue, anchoring our approach in Quantificational event semantics, [Champollion, 2017].

As the subject of this thesis is dialogue modeling, the next chapter focuses more specifically on dialogue-related phenomena and formalisms, presenting an overview of the existing theories along with some typologies and language-related considerations.

Role	Definition
Agent	Participant in an event who intentionally or consciously initiates an event, and who exists independently of the event.
Beneficiary	Participant in an eventuality that is advantaged or disadvantaged by the eventuality, and that exists independently of the event.
Cause	Participant in an event that initiates the event, but does not act with any intentionality or consciousness; the participant exists independently of the event.
Goal	Participant in an event that is the (non-locative, non-temporal) end point of an action; the participant exists independently of the event.
Instrument	Participant in an event that is manipulated by an agent, and with which an intentional act is performed; it exists independently of the event.
Partner	Participant in an event that is intentionally or consciously involved in carrying out the event. Participant is not the principal agent of the event, and exists independently of the event.
Patient	Participant in an event that undergoes a change of state, location or condition, is causally involved or directly affected by other participants, and exists independently of the event.
Pivot	Participant in a state that is characterised as being in a certain position or condition throughout that state, and has a major or central role or effect in that state. A pivot is more central to the state than a participant in a theme role, and exists independently of the state.
Purpose	Set of facts or circumstances that an agent wishes or intends to accomplish by performing some intentional action.
Reason	Set of facts or circumstances explaining why a state exists or an event occurs.
Result	Participant in an event that comes into existence through the event; it indicates a terminal point for the event: when that is reached, the event does not continue.
Setting	Set of (non-locative and non-temporal) facts or circumstances of the occurrence of an event or a state.
Source	Non-locative, non-temporal starting point of an event. The source exists independently of the event.
Theme	Participant in a state or an event that (i) in the case of an event, is essential to the event taking place, but does not have control over the way the event occurs and is not structurally changed by the event, and (ii) in the case of a state, is characterised as being in a certain position or condition throughout the state, and is essential to the state being in effect but not as central to the state as a participant in a pivot role. The theme of a state or event exists independently of the state or event.
Manner	The way or style of performing an action or the degree/strength of a cognitive or emotional state.
Medium	The physical setting, device or channel that allows an event to take place.
Means	Procedure for performing an action in terms of component steps, or a methodology by which an intentional act is performed by an agent. A means does not necessarily exist independently of the event.
Location	Place where an event occurs, or a state is true, or a thing exists.
Initial Location	Participant in an event that indicates the location where an event begins or a state becomes true; initial-location exists independently of the event.
Final Location	Location where an event ends or a state becomes false; final-location exists independently of the event.
Path	Intermediate location or trajectory between two locations, or in a designated space, where an event occurs.
Distance	Length or extent of space that plays a role in an eventuality.
Time	Participant that indicates an instant or a time interval during which a state exists or an event takes place.
Duration	Length or extent of time during which an event occurs or a state is true.
Initial Time	Indication of the point in time when an event begins or a state becomes true.
Final Time	Indication of a point in time when an event ends or a state ceases to be true.
Amount	Quantity of something other than time or space, or number of objects of a certain kind, which plays a role in an event or a state.
Attribute	Property that an event or state associates with one of the other participants.

Table 1.8: Semantic roles defined in ISO 24617-4 – DIT++ [Bunt, 2009].

# Dialogue and formal semantics

## Contents

---

<b>2.1</b>	<b>Gricean maxims</b> . . . . .	<b>52</b>
<b>2.2</b>	<b>Studies of dialogue</b> . . . . .	<b>53</b>
2.2.1	Speech Act Theory and annotation schema . . . . .	53
2.2.2	Dialogue context & the common ground . . . . .	55
2.2.3	KoS . . . . .	56
2.2.4	SDRT . . . . .	58
<b>2.3</b>	<b>Questions &amp; Answers in dialogue</b> . . . . .	<b>59</b>
2.3.1	Typologies of questions and answers . . . . .	60
2.3.2	Clarification Requests . . . . .	65
<b>2.4</b>	<b>Question semantics</b> . . . . .	<b>65</b>
2.4.1	Questions in inquisitive semantics . . . . .	68
<b>2.5</b>	<b>Semantics of <i>wh</i>-questions</b> . . . . .	<b>69</b>
2.5.1	Semantic representations of <i>wh</i> -questions . . . . .	70
<b>2.6</b>	<b><i>Wh</i>-words – thematic roles correspondence</b> . . . . .	<b>74</b>
2.6.1	French Interrogatives . . . . .	77
<b>2.7</b>	<b>Conclusion of the chapter</b> . . . . .	<b>78</b>

---

The previous chapter introduces concepts and frameworks of formal semantics of sentences and of sentences in discourse. Now, we want to focus on formal semantics of dialogue, in particular on formal semantics of utterances in context. To do so, we will first present studies of dialogue in general through Speech Act Theory, see section 2.2.1. Then we focus on dialogue context through notions such as the *common ground* [Stalnaker, 2002] in section 2.2.2 and the KoS framework in section 2.2.3, before addressing the notions of discourse relations through the SDRT framework in section 2.2.4.

The work presented in this thesis stems from formal semantics of discourse while aiming for dialogue modeling. Discourse and dialogue share many similarities, in particular, both can be seen as incrementally built, sentence by sentence for discourse and utterance by utterance for dialogue. Dynamic phenomena such as anaphora take place both in discourse and dialogue, and both can be modeled in a compositional way. These similarities motivate our approach.

It is important to note that while dialogue can be considered as a particular case of discourse in literature, it is not the case in this thesis. In the view we take on dialogue and discourse,

the two present huge differences. As a dialogue is built between at least two participants, it is necessary to consider at least two points of view when modeling dialogue. In discourse semantics, the classical approach assigns truth conditions to sentences; in dialogue, a statement can be seen as true by one participant and false by the other, leading to a multiplication of ways the truth conditions can be evaluated. Alongside these truth-condition-related issues lie the ones related to the question-answer relationship. Should one assign truth values to utterances? Would it make sense to assign a truth condition to a question? Questions can be characterized by their set of possible answers. As for answers, though some can be made using complete declarative sentences, one can often observe specific *elliptic* phenomena, see example 23.

### Example 23 (Ellipsis in dialogue)

**A**<sub>1</sub> Who's coming tonight?

**B**<sub>2</sub> Charlie

In this example, **B**<sub>2</sub> corresponds to the declarative sentence “Charlie is coming”. This declarative sentence can be built by combining the question **A**<sub>1</sub> with its answer **B**<sub>2</sub>; it cannot be computed independently, without taking the *context* of the dialogue into account.

The sections of the second part of this chapter are motivated by the following inquests: What is a question? What is an answer? How do they articulate: what makes one accept that an utterance answers a question? What makes an answer satisfactory, i.e. resolving for the question it is supposed to answer? How to define question/answer relationship in a precise, computational way? These issues are at the core of the specificities of dialogue that need to be addressed in order to produce a strong model.

**Note.** Examples in this chapter mostly come from the *Dialogues in Games* corpus (DiG), see Chapter 3 for a full presentation and discussion. This corpus is composed of transcriptions of recordings of people playing the *Catan* board game, in French. The participants are designated by the colour of their tokens: **Red (R)**, **White (W)**, **Yellow (Y)**, **Blue (B)**.

## 2.1 Gricean maxims

Among the first observations that were made on discourse and dialogue interactions, considered to be foundational for pragmatics, one can find Paul Grice's contributions [Grice, 1975]. He states the cooperative principle:

*Make your contribution such as is required, at the stage at which it occurs, by the accepted purpose or direction of the talk exchange in which you are engaged.*

The cooperative principle is then divided in *Grice's four maxims of conversation*, also called *Gricean maxims*.

**Definition 12 (Gricean maxims)**

**The maxim of quantity** *An utterance needs to be as informative as possible, give as much information as is needed, and no more.*

**The maxim of quality** *An utterance needs to be truthful, and not give information that is false or that is not supported by evidence.*

**The maxim of relation** *An utterance needs to be relevant, and say things that are pertinent to the discussion.*

**The maxim of manner** *An utterance needs to be as clear, as brief, and as orderly as possible and needs to avoid obscurity and ambiguity.*

A large part of the dialogue theories presented below is inspired by the Gricean maxims and takes them as a starting point. Throughout our studies of real-life data in dialogue (see in particular Chapter 4), we encountered several occasions on which the Gricean maxims were broken. In particular, this is the case for DinG and STAC<sup>32</sup> corpora – as these two corpora are composed of multilogues between people playing a non-cooperative game, at least the maxim of quality tends to be left aside.

## 2.2 Studies of dialogue

The view this thesis takes on dialogue is not the most conventional. Indeed, dialogue is a subject of interest for several fields: linguistics, as dialogue uses language; computation, through natural language processing, but also sociology and psychology as dialogue takes place between human beings. In most of this thesis, we focus on a formal semantics approach to dialogue; yet, we want to contextualize our study by presenting classical approaches to dialogue such as Speech Act Theory and classical ideas for question and answer annotation such as the ones of [Austin, 1975] and [Freed, 1994].

### 2.2.1 Speech Act Theory and annotation schema

Conversations are built through sequences of utterances. In speech act theory, a subfield of conversation studies, utterances are studied as parts of what is called a *speech act*, described as follows:

*A speech act is an act of uttering that can be performed by a speaker meaning that one is doing so.* [Green, 2020]

This description corresponds to the idea that a speech act is an utterance that conveys the meaning the speaker gives to it; in particular, according to this definition, promising, asserting, asking are speech acts, while insulting and whispering are not. This definition also allows for speech acts to be performed without uttering any words (for example, with a nod). Close to the notion of speech act lies the *illocution* [Austin, 1975], the two terms sometimes being used synonymously. Austin uses the terms *illocutionary force* and *locutionary act*: a locutionary act is the act of uttering, while the illocutionary force gives this act a meaningful dimension. This draws a distinction between what a speaker says and the force/flavor/meaning of her utterance.

<sup>32</sup>See Chapter 3, section 3.2.1 for a presentation of STAC.



The decomposition of dialogue in utterances and the classification of these utterances according to different possible dimensions results in the production of various annotation schema for dialogue. Among the most well-known is the Dialog Act Markup in Several Layers (DAMSL) schema [Core and Allen, 1997]. DAMSL has three layers: Forward Communicative Functions, Backward Communicative Functions, and Utterance Features, see table 2.1. The Forward Communicative Functions assign to an utterance one of the possible action-labels, such as **Assert** or **Offer**. These labels are similar to ones from Speech Act Theory. Backward Communicative Functions indicate how the utterance under consideration relates to the previous utterances in the dialogue (for example, by answering a question). Utterance Features include information about an utterance’s form and content, following ideas similar to Austin’s locutionary acts and illocutionary forces. The core idea of DAMSL is to allow an utterance to be annotated as performing several actions at once: answering a question, but also informing, promising an action, *etc.*

DAMSL was originally focused on *task-oriented dialogues*: dialogues that are built because the participants want to achieve a specific goal, finish a specific task. Among the most classical task-oriented dialogues are the ones where participants need to communicate to each other the location of or the path to something using a shared map; dialogues aiming at booking a train/plane ticket. Task-oriented dialogues can be opposed to *real-life non-constrained data*, which can be obtained by leaving a recording device to willing participants without setting any constraint on a specific topic of conversation. For more about real-life data, see Chapter 3.

Forward Communicative Functions	Backward Communicative Functions	Utterance Features
Statement <ul style="list-style-type: none"> <li>• Assert</li> <li>• Reassert</li> <li>• Other-statement</li> </ul>	Agreement <ul style="list-style-type: none"> <li>• Accept</li> <li>• Accept-Part</li> <li>• Maybe</li> </ul>	Information Level <ul style="list-style-type: none"> <li>• Task</li> <li>• Task Management</li> <li>• Communication Management</li> </ul>
Influencing Addressee Future Action <ul style="list-style-type: none"> <li>• Open-option</li> <li>• Directive: Info-request or Action-Directive</li> </ul>	<ul style="list-style-type: none"> <li>• Reject-Part</li> </ul>	Other
Committing Speaker Future Action <ul style="list-style-type: none"> <li>• Offer</li> <li>• Commit</li> </ul>	Understanding <ul style="list-style-type: none"> <li>• Signal-Non-Understanding <ul style="list-style-type: none"> <li>• Signal-Understanding: Acknowledge, Repeat-Rephrase, Completion</li> </ul> </li> <li>• Correct-Misspeaking</li> </ul>	Communicative Status <ul style="list-style-type: none"> <li>• Abandoned</li> </ul>
Performative Other Forward Function	Answer Information-Relation	<ul style="list-style-type: none"> <li>• Uninterpretable</li> </ul>
		Syntactic Features <ul style="list-style-type: none"> <li>• Conventional Form</li> <li>• Exclamatory Form</li> </ul>

Table 2.1: DAMSL annotation schema, adapted from [Core and Allen, 1997].

One of the most detailed annotation schemas for dialogue nowadays is given by DIT++. DIT++ is a semantics-based framework for dialogue analysis and annotation. It is an extension of the Dynamic Interpretation Theory (DIT) [Bunt, 1994], originally developed for information dialogues, mingled with dialogue act types from DAMSL and other annotation schemas and

semantic frameworks such as SDRT. It contains a Taxonomy of Dialogue Acts, an annotation scheme, and the definition of the Dialogue Markup Language (DiAML)<sup>33</sup>. As opposed to DAMSL, DIT++ is specifically developed in view of an application-independent dialogue act annotation. It defines the ISO 24617-2 norm for written, oral, and multimodal dialogue [Bunt et al., 2012]. Dialogues annotated according to this ISO standard can be found in the DialogBank<sup>34</sup>. The ISO 24617-2 norm states the following about dialogue acts:

*Dialogue acts can be interpreted, formally, as update operations that are applied to the information states of the interacting participants. A dialog act is defined by its communicative function and its semantic content. [Bunt et al., 2012]*

The semantic content defines the objects of the dialogue act. The communicative function establishes the way in which the dialogue act updates the information state of the interacting participants. Furthermore, the rhetoric relations between dialogue acts can be annotated with an indication of a way one dialogue act triggers the execution of another dialogue act, allowing for a dialogue act composition.

DIT++ is extremely detailed. Table 1.8 shows the full list of thematic roles used in the annotations, and the two reference webpages show the extensive annotation schema that contains several levels of analysis. One of the goals of our future work is to integrate insights from DIT++ in our dialogue model and annotations.

### 2.2.2 Dialogue context & the common ground

Now, if we consider dialogue as the combination of a context (previous utterances) and a continuation (utterances that yet have to come), it seems that it should be possible to obtain a dynamic semantic representation of a dialogical interaction by following dynamic semantic representations of discourse. However, it appears that dialogue cannot be simply considered in that way as a sub-type of general discourse, as dialogue involves more than one participant and, therefore, discourse phenomena appear in an altered way in dialogue. The multiple participant nature of dialogue triggers issues linked to the notion of *common ground* as it is stated in [Stalnaker, 2002]:

*It is common that  $\phi$  in a group if all members accept (for the purpose of the conversation) that  $\phi$ , and all believe that all accept that  $\phi$ , and all believe that all believe that all accept that  $\phi$ , etc.*

It follows that it is absolutely not granted that a proposition that is uttered by one of the participants of the conversation is common to all the participants. It is usual for an utterance to first be discussed (in a broad sense of discussion) before being accepted. Consider the following dialogue between **A** and **B**:

#### Example 24 (A simple dialogue)

**A**<sub>1</sub> You turn left here.

**B**<sub>2</sub> There?

**A**<sub>3</sub> No, here.

<sup>33</sup><https://dit.uvt.nl/>, visited on 06/07/21.

<sup>34</sup><https://dialogbank.uvt.nl/>, visited on 06/07/21.

First, **A** utters a declarative sentence. Then, **B** asks a question about one of the constituents of **A**'s utterance; thus, **B** does not directly accept **A**'s proposition. **A** then answers **B**'s question negatively, therefore switching its content to the opposite one: the utterance that might be further discussed is at this point “*You turn left here, not there.*”. Though this example is quite simple, it illustrates that modeling dialogical interaction involves taking into consideration several complex phenomena that must be added to the ones encountered in discourse modeling. The following section introduces KoS, a formalism designed for dialogue context modeling.

### 2.2.3 KoS

Early research on computational dialogue modeling was done in the late 1990s [Traum et al., 1999, Cooper and Larsson, 1999, Ginzburg, 1996]. In these approaches, dialogues are conceptualized as a kind of game in which each of the participants has, at any given point in time a certain set of *goals* (e.g. find out about a particular piece of information) and a specific set of possible *dialogue moves* that can be used (e.g. ask a question, assert something) to further these goals. A crucial concept is the one of *information state*: a representation of the information that is present in the dialogue at a particular point in time and is updated each time a new utterance is produced. Importantly, information states constrain the set of available moves at every point (e.g. an assertion can be rejected or accepted, and a question needs to be answered).

KoS [Ginzburg, 2012] is a formal semantic framework based on Type Theory with Records (TTR), see Chapter 1 for an introductory presentation. The core concept of KoS is the *Dialogue Game Board* (DGB): a representation of the dialogue context, separate for each dialogue participant, dynamically changing as the dialogue unfolds. A DGB represents the information state of the participants, separated in a private part (that contains information specific to the participant) and a public one, resulting from the publically shared interactions.

#### Definition 13 (Dialogue gameboard (DGB), [Ginzburg, 2012])

Formally, a DGB is defined as a record of the following type:

$$DGBT_{\text{type}} =_{\text{def}} \left[ \begin{array}{ll} \textit{spkr} : & IND \\ \textit{addr} : & IND \\ \textit{utt-time} : & Time \\ \textit{c-utt} : & \textit{addressing}(\textit{spkr}, \textit{addr}, \textit{utt-time}) \\ \textit{Pending} : & \textit{list}(\textit{LocProp}) \\ \textit{Moves} : & \textit{list}(\textit{LocProp}) \\ \textit{Facts} : & \textit{set}(\textit{Prop}) \\ \textit{QUD} : & \textit{poset}(\textit{Question}) \\ \textit{non-resolve-cond} : & \forall q \in \textit{QUD} [\neg \textit{Resolve}(\textit{FACTS}, q)] \end{array} \right]$$

The fields of this type have the following intended meanings:

- *spkr* stores the entity of type IND (individual) that is the speaker of the current utterance;
- *addr* stores the entity of type IND (individual) that is the addressee of the current utterance;
- *utt-time* stores the time code of the utterance;
- *c-utt* presents the context of the utterance: who is the speaker, to whom the utterance is addressed, at which time was it uttered;

- *Pending* is a list of locutionary propositions (in the sense of [Austin, 1975]) that keeps track of ungrounded utterances: utterances that have not yet been interpreted to produce a *Fact* or a *QUD* update;
- *Moves* is a list of locutionary propositions that contains the last dialogue move that was made (often represented apart by a *LatestMove* field), the dialogue moves that happened before, and, depending on the considered model, the *Moves to come*: usually, in this case, private to the person who's *DGB* we are considering.
- *Facts* is a set<sup>35</sup> of locutionary propositions, containing the propositions that have been accepted by all the participants of the conversation (the shared knowledge);
- *QUD* is a partially ordered set that specifies the currently discussed questions, called questions under discussion: it contains issues that have been raised during the conversation and not yet resolved; *non-resolve-cond* is the non resolving condition that specifies that the *QUD* cannot contain issues that are already resolved by locutionary propositions stored in *Facts*.

Definition 13 gives a very general presentation of a *DGB*. In practice, some of the fields tend to change or be left out depending on the phenomena the *DGB* is used to model. In particular, see our take on a *DGB*-based model in Chapter 5, section 5.4. Once the *DGBs* are defined, they can be modified by conversationally related actions. [Ginzburg, 2012] gives the following definition for *conversation rules*:

*We call a mapping that indicates how one DGB can be modified by conversationally related action a conversational rule, and the types specifying its domain and its range, respectively, the preconditions and the effects. [Ginzburg, 2012]*

[Ginzburg, 2012] defines over 30 different conversational rules, applicable in various dialogical settings. Some focus on initialization and termination of conversations; it is the case of the rules called *Greeting*, *Countergreeting*, *Parting*, *Counterparting*, *Disengaging*. Another group of rules is defined for turn-holding purposes only: *No-Turn-Change*, *Turn-Change*, *Turnholder-Underpecified*. Several rules are designed to account for the effects of reactive moves, made by the participants of the conversation in reaction to a previous move. We focus in particular on a group of conversational rules that is especially relevant for our view of dialogue through the questions and answers articulations with the context. These conversational rules focus on updating the *QUD* and *FACT* fields of the *DGB*: *Assert QUD-incrementation*, *Ask wh-question QUD-incrementation*, *p? QUD-incrementation*, *Fact Update/QUD downdate*, *NonResolve*, *QCoord*.

*Assert QUD-incrementation*, *Ask wh-question QUD-incrementation*, and *p? QUD-incrementation* all stem from the same update mechanism for the *DGB*: when a new utterance occurs, a new question under discussion is likely to appear. What is interesting here is that assertions are as likely as questions to produce a new question under discussion and thus trigger a *QUD* update; asserting a statement does not automatically imply that it is accepted by all the participants of the conversation. Until it is accepted, the statement is in fact a question under discussion and should be stored in the *QUD*.

Once a question under discussion is resolved, the rule *Fact Update/QUD downdate* is triggered: the questioned content is removed from the *QUD* field; simultaneously, the *FACT* field is updated with the part of the content the participants of the discussion have agreed upon. The

<sup>35</sup>The important difference between a list and a set here is that a list is ordered while a set is not.

conversational rule `NonResolve` takes care of removing from the QUD field all elements whose question component `q` is such that an element of `FACTS` resolves `q` or the lack of desire to discuss `q`. Finally, `QCoord` ensures that when a new utterance introduces an update on a previous entry in QUD, the two are coordinated so that the QUD can be updated in consequence.

The view of dialogue developed in KoS is very detailed. The use of TTR gives considerable expressivity because of the multiplication of types. KoS is partly implemented, see for example [Maraev et al., 2018]. In the work developed in this thesis, we are deeply interested in the insights from KoS, especially from the point of view of the interactions between utterances and the dialogue context. However, we cannot directly use KoS to achieve the dynamic model of dialogue we are aiming for, as KoS operates on a higher level of dialogue modeling: we would like to model the dialogue starting from the syntactic structures of the utterances composing the conversation, and then use a syntax-semantics interface to compute the semantic representation of these utterances and then of the whole conversation. KoS provides us with tools we can use for computing the semantic representation of the conversation based on the semantic representation of the utterances – see Chapter 5, section 5.4 for our use of KoS for dialogue modeling in a dynamic setting.

KoS is a framework designed to model dialogue and in particular the phenomena surrounding the dialogue context. The next section gives the presentation of SDRT, a formalism that focuses on discourse and dialogue relations.

#### 2.2.4 SDRT

Segmented Discourse Representation Theory (SDRT) is a model of discourse and dialogue, positioned by its authors on the semantics-pragmatics interface. It follows that SDRT’s modeling occurs at a level different from most of the formalisms and frameworks presented in Chapter 1, section 1.3: the goal of SDRT is to compute the additional information provided by the interpretation of sentences (or utterances) in context. Central to SDRT is the notion of *coherence*:

*A discourse is defined to be coherent only if each non-initial utterance makes an illocutionary contribution which connects it to some other contribution in the discourse context.* [Asher and Lascarides, 2003]

The connections between the utterances are seen as the underlying rhetorical structure of discourse. [Asher and Lascarides, 2003] gives a glossary of discourse relations. In the following, we only focus on the ones given in the discourse structure annotation schema provided by [Asher et al., 2016] for the STAC corpus, see Chapter 3, section 3.2.1 for a complete presentation. This discourse annotation schema was designed for a certain type of multilogue which is interesting for our dialogue modeling purposes. It contains sixteen categories for discourse relations. The first are defined at content-level; their definition varies according to the type of utterances they involve: indicative (declarative), interrogative, imperative. These first categories are `alternation`, `background`, `continuation`, `elaboration`, `explanation`, `narration`, `question-answer pair` (QAP), and `result`.

Then, some discourse relations are there to structure the discourse/dialogue: `contrast` and `parallel` are among those. This category of discourse relations is close to the common definition of rhetorical relations one can encounter in literature studies for example. Finally, as the goal of SDRT is to model the semantics-pragmatics interface, the annotation schema also contains some cognitive-level discourse relations: `acknowledgment`, `question elaboration` (Q-Elab), and `correction`.

SDRT gives a complete formal model for discourse and dialogue pragmatics. We hope that the compositional models of dialogue we propose in this thesis can benefit from the pragmatic tools and insights from SDRT, so to result in a complete, formal, syntax-semantics-pragmatics pipeline. The coherence considerations developed in SDRT are particularly interesting for us as we focus on the relationship between questions and answers in dialogue: what makes an answer coherent with the question it is supposed to answer? In order to begin to outline answers to this research question, we focus, in the following section, on questions and answers in dialogue.

## 2.3 Questions & Answers in dialogue

A first approach to defining *questions*, as dialogue acts, comes from syntax. We can define *questions* as utterances marked with a question mark/a rising intonation/a question particle/a question word. In English, French, Dutch and Italian, direct questions are marked with a question mark, as opposed to indirect questions and other sentences, that are not. In Spanish, the beginning of the span of a question is visually given through the use of an inverted question mark ‘¿’. This is usually true in official written texts, but is less the case in informal writing such as texting and written chat by younger Spanish-speakers [Taulé et al., 2015]. In particular, transcriptions of oral data in Spanish often lack the inverted question mark, perhaps because adding it would be an interpretation of the actual question span.

In Armenian, the question mark (hartsakan nshan) is used in order to highlight a question word in an interrogative sentence and is placed over the last vowel of the question word. Interrogative sentences then finish with the same sign as declarative and exclamative ones: ‘.’. In Greek, the question mark ‘;’ is used even for indirect questions. In Japanese, the syllable ‘ka’ marks the interrogative sentence, often in conjunction with the fullwidth interrogative sign, used in Chinese, Japanese and Korean.

Formal semantics literature on questions such as [Groenendijk et al., 1997] usually draws a difference between interrogative sentences, interrogatives, interrogative acts, and questions. For the sake of simplicity, we will use the word “*question*” to designate the interrogative sentences and the interrogative utterances alike. Semantically, a question is a sentence or phrase used to find out information<sup>36</sup>. Yet, a question can actually be used without the purpose of finding out information; it is typically the case for rhetorical questions.

### Example 25 (Rhetorical question, DinG6)

**Yellow**<sub>430</sub> *mais attends (0.1s) tu as vu les cartes que tu as et tu veux nous faire du un contre un ?*

**Yellow**<sub>430</sub> *“but wait (0.1s) you saw the cards you have and you want us to go for a one-for-one [exchange]?”*

In example 25, **Yellow** isn’t actually expecting an answer from the addressee. Yellow’s utterance is a mere expression of her incredulity due to the boldness of the player proposing the one-for-one exchange. Thus, not all questions are used to seek information. Therefore, even though it is possible to give a general definition of questions, it is easy to see that an actual study of questions in a given language requires some fine-grained research. Section 2.3.1 presents a detailed typology of questions in English and French.

Defining what constitutes an *answer* and how it fulfills its answering role is even more complex. Sometimes, the answer to a question is an entity, one element of information. Sometimes,

<sup>36</sup>Definition based on <https://dictionary.cambridge.org/dictionary/english/question>, visited on 08/05/2020.



the answer defines or shapes a set of entities or elements of information, but is not sufficient to choose an element of this set. Even defining the span of an answer can be tedious: where does the answer start and where does it end? What should be considered as part of the answer?

**Example 26 (Answer’s span, SCoSE [Norricks, 2017])**

**A**<sub>1</sub> *When will you guys get off?*

**B**<sub>2</sub> *My last exam is like...I don’t know, maybe on Monday or on Tuesday...*

In example 26, **B**<sub>2</sub> appears to be an answer to **A**’s question, as it is an utterance that directly follows the question. Yet, it is in fact difficult to identify where the actual answering part of **B**’s utterance is: where does the answer to **A**<sub>1</sub> begin? Where does it end?

- It cannot be “*I don’t know*”, as **B**<sub>2</sub> actually brings some information to **A**.
- It cannot be “*maybe on Monday or on Tuesday...*” as these two days actually correspond to the day on which **B**’s exam is, not the day on which “*the guys get off*”.

Therefore, it is necessary to consider **B**’s utterance as a whole: “*My last exam is like...I don’t know, maybe on Monday or on Tuesday...*” constitutes the answer to **A**’s question, even though this utterance doesn’t actually resolve the issue that has been raised. More about these phenomena in section 2.3.1, where we discuss the *aboutness* phenomenon. Another possible interpretation for example 26, that was given to us by Jonathan Ginzburg, is that “*I don’t know*” has a role similar to one that “*I mean*” would have in the same context: it could be an editing expression, opening the way to the following part of the utterance. In this case, “*I don’t know*” can even less be considered as an answer to **A**<sub>1</sub>.

First, while classifying answers by types is a difficult problem, finding what segment of data is an answer is a whole other problem of its own.

**Example 27 (Simple answer, DinG6)**

**Yellow**<sub>1</sub> *Tu as fait combien ?*

**Yellow**<sub>1</sub> *“How many did you make?”*

**Red**<sub>2</sub> *oh oui 11 d’accord*

**Red**<sub>2</sub> *“Oh yes 11 OK”*

In example 27, the simple answer is “11”, but it is surrounded by phatic elements (“*Oh yes*” and “*OK*”). If **Yellow**<sub>1</sub> was a closed question, these phatic elements could have been simple answers.

Studying questions and answers from a formal perspective in a real-life data perspective requests a delicate balance between formal, precise definitions that name phenomena and actual speaker’s productions classification and interpretation. The following section introduces a typology of questions and answers phenomena in dialogue. This typology is based on formal studies of dialogue and illustrated on real-life corpora data.

### 2.3.1 Typologies of questions and answers

In order to study questions and answers and their articulation, we need to specify our subjects of study. In English, in first approach, one can classify questions as either *closed questions*, *wh-questions*, or *alternative questions*.

**Definition 14 (Closed question, [Aarts et al., 2014])**

A closed question (also called *yes/no-question*) is used to ask a question with “yes” or “no” as an answer. A closed question can take the form of a declarative clause.

**Example 28 (Closed question, DinG6)**

**Yellow**<sub>330</sub> *tu veux me donner du blé ?*

**Yellow**<sub>330</sub> “you want to give me some wheat?”

**Blue**<sub>331</sub> *ouais*

**Blue**<sub>331</sub> “yeah”

In example 28, **Yellow** asks a closed question. As this example is extracted from a corpus of transcriptions of real-life oral dialogues, **Yellow**’s question is formulated through a declarative clause: **Yellow**’s utterance is a declarative sentence with an interrogative prosody. **Yellow**’s question introduces an assertion: “Blue wants to give Yellow some wheat”, that is confirmed by **Blue**’s answer, an informal equivalent of “yes”.

**Definition 15 (Wh-question, [Ginzburg and Sag, 2000])**

*Wh-questions* are defined as questions, in English, that give rise to answers whose semantic types match those of the *wh*-phrase contained in the interrogative. A *wh*-phrase is introduced by a *wh*-word; see table 2.2 for a complete list [Aarts et al., 2014].

WH-WORD	QUERY
What	Entity, object
When	Time, moment
Where	Position, place
Who	Person
Whom	Person
Which	Choice, alternative
Whose	Person, entity
Why	Reason
How	Way, manner, characteristic

Table 2.2: *Wh*-words and corresponding queries.

**Example 29 (Wh-question, DinG6)**

**Yellow**<sub>789</sub> *alors qu’est-ce que je peux faire avec ça ?*

**Yellow**<sub>789</sub> “so what can I do with this?”

In example 29, **Yellow** asks a question introduced by the *wh*-word “what”. It is interesting to see that while the English translation fits in the definition of a *wh*-question quite nicely through the usage of the *wh*-word, the French original is lexically more complex, as the part corresponding to the *wh*-word “what” in English is « qu’est-ce que », in French. More about this at the end of the current section, in 2.6.1.

**Definition 16 (Alternative question, [Aarts et al., 2014])** An alternative question is a subtype of closed question which names possible answers but does not leave the matter open.



**Example 30 (Alternative question, DinG3)**

**Blue**<sub>713</sub> *est-ce que c'est une carte ressource ou est-ce que c'est n'importe quelle carte ?*  
**Blue**<sub>713</sub> *"is it a resource card or is it any card?"*

**White**<sub>714</sub> *une carte ressource*

**White**<sub>714</sub> *"a resource card"*

In example 30, **Blue** asks a question that has characteristics of a closed question: it bears the linguistic cue « est-ce que » in French, translated through the subject-verb inversion “is it” in English. Then, **Blue**<sub>713</sub> continues by naming two possible answers: “a resource card”, “any card”. **White**'s answer closes the matter by choosing one of these possibilities.

[Ginzburg and Sag, 2000] gives a definition of answers as “phrasal utterances used to respond to queries”. As this definition in turn raises the issue of defining what it means to respond to a query, we dive into a more detailed view of answers. [Ginzburg, 2012] distinguishes 5 types of notions for the answers: (1) *simple answerhood*, (2) *aboutness*, (3) *strong exhaustive answer*, (4) *partially resolving answer*, and (5) *question/question relations* (answer to a question by a question), see table 2.3.

Category	Description
Simple Answer	Single-item utterance, concerning an instantiation of a question or its negation
Aboutness	Conditionals, modalities, disjunctions, quantifiers; gives additionnal info, doesn't resolve the issue
Strongly exhaustive answer	Entails all the (true) simple answers
Potentially resolving answer	Either brings a simple answer or shows that there is no answer
Question	Answer that is an interrogation

Table 2.3: Taxonomy of answers, adapted from [Ginzburg, 2012].

**Definition 17 (Simple answer, [Ginzburg, 2012])** *A simple answer is a single-item utterance in a dialogue concerning either an instantiation of a question or the negation of such an instantiation.*

**Example 31 (Simple answer, DinG3)** <sup>37</sup>

**Blue**<sub>1</sub> *C'est quoi qui rapporte le plus de thune ?*

**Blue**<sub>1</sub> *"What brings in the most cash?"*

**Yellow**<sub>2</sub> *Le 6*

**Yellow**<sub>2</sub> *"The 6"*

The answer is unique (6). It is an instantiation of the problem “Among the possible numbers on the dice, what is the one that brings in the most cash?”.

<sup>37</sup>In Catan, profit depends on the number on which the die falls.

When answers are not simple ones, different cases emerge. For closed questions, sometimes the answer is not a straight confirmation or a straight negation of the informational content. It is the case for conditional answers (“*if condition then yes*”) and answers under modalities (“*maybe*”, “*probably*”, “*possibly not*”).

**Definition 18 (Aboutness, [Ginzburg, 2012])** *An answer A is about a question if and only if A entails a disjunction of simple answers.*

**Example 32 (Aboutness, DinG3)**

**Red**<sub>1</sub>      *Tu fais quelque chose ?*  
**Red**<sub>1</sub>      “*Are you doing something?*”

**Yellow**<sub>2</sub>    *Ah ben j’ai construit 2 chemins si tu veux j’ai pas l’habitude de faire des trucs de fou comme ça*

**Yellow**<sub>2</sub>    “*Well I built 2 roads you know I’m not used to doing crazy stuff like that*”

In example 32, **Red** asks a closed question. **Yellow** answers with a long sentence that amounts to a negative answer, but a human observer needs to make an inferential step to be able to deduce this: as **Yellow** is not used to ‘*doing crazy stuff like that*’, she will not be doing a lot of that, so she will not be doing something at the present moment. For discussions of these types of enthymematic reasonings, see [Breitholtz, 2014].

*Wh*-questions can be answered with disjunctive answers, presenting several possible short answers (in the context of example 64, “*The 3 or the 6, I don’t remember*”) or with quantified answers (including generalized quantifiers such as “*at most*”, “*a few*”, *etc.*).

**Example 33 (Aboutness, DinG8)**

**Blue**<sub>1</sub>      *Qui se fait souvent de la pierre ?*  
**Blue**<sub>1</sub>      “*Who often makes stone?*”

**White**<sub>2</sub>    *Euh, je m’en suis faite une seule depuis tout à l’heure*

**White**<sub>2</sub>    “*Uh, I’ve only made one since earlier*”

In this example, **White**’s answer allows us to rule her out of the set of players among whom we are looking for the ones that often make stone. Thus, this answer gives additional information but doesn’t completely solve the issue raised by **Blue**.

**Definition 19 (Strongly exhaustive answer, [Ginzburg, 2012])** *An answer A is strongly exhaustive for a question Q if and only if A is true and entails all the A<sub>i</sub> that are simple answers to Q.*

Intuitively, an answer is considered to be strongly exhaustive if independently from what comes next in the course of the dialogue, this answer cannot become more precise. It isn’t necessarily a simple answer, but the pragmatic environment of the dialogue allows us to approximate it as such. As pointed out by Jonathan Ginzburg, “yes” and “no” constitute strongly exhaustive answers to polar questions. As far as *wh*-questions are concerned, we have not found examples of strongly exhaustive answers in DinG yet. Example 34 presents a constructed example of a strongly exhaustive answer in DinG’s context.

**Example 34 (Strongly exhaustive answer, fabricated)**

**Red**<sub>1</sub>      *Tu veux quoi contre du blé ?*  
**Red**<sub>1</sub>      “What do you want to trade for wheat?”

**Yellow**<sub>2</sub>    *De la pierre ou du bois*  
**Yellow**<sub>2</sub>    “Stone or wood”

**Yellow**<sub>2</sub> is true and entails both “stone” and “wood”, that are simple answers to **Red**<sub>1</sub>.

**Definition 20 (Potentially resolving answer, [Ginzburg, 2012])**

A potentially resolving answer is an answer that either brings a simple answer, either shows that there is no answer.

This characterization of an answer is interesting as it seems to come directly from pragmatics. It finds an impact directly in our model. This definition supposes that the question is solved but doesn't specify the origin of the information that solves the question. It may come from the first component of the answering couple (the information itself), so from an effective answer. It may also come from the second component of the couple, the polarity. It is actually also possible that there is no answer.

**Example 35 (Potentially resolving answer (self), DinG5)**

*Yellow is talking to herself, pausing between the question and the answer.*

**Yellow**<sub>1</sub>    *Alors qu'est-ce que je peux faire avec ça ?*

**Yellow**<sub>1</sub>    “So what can I do with this?”

**Yellow**<sub>2</sub>    *Euh ben rien en gros rien du tout*

**Yellow**<sub>2</sub>    “Uh, well, nothing, basically nothing at all”

In example 35, **Yellow** both asks the question and gives the answer. Here, the resolution of the question comes directly from the information contained in the answer. The question/answer mechanism is covertly used both to support **Yellow**'s thinking process and to convey information to the other players.

Knowing that it's possible to answer a question with a question forces us to consider the interactive aspect of dialogues, drawing us closer to the discursive modeling in the tradition of [Asher and Lascarides, 2003]. This reference gives us the definition of a **Question elaboration** that stresses out the relation between two questions linked in a way such that any answer to the second one gives a strategy that leads to the solving of the query of the first one.

**Example 36 (Question/question, DinG2)**

**White**<sub>1</sub>    *Tu as combien de moutons ?*

**White**<sub>1</sub>    “How many sheep do you have?”

**Blue**<sub>2</sub>      *Tu voudrais combien de moutons ?*

**Blue**<sub>2</sub>      “How many sheep would you like?”

In example 36, **White** asks **Blue** a *wh*-question, the answer to which would be expected to be an amount. **Blue** replies with another question, possibly because she doesn't want to reveal the content of her hand but is willing to trade if the trade is worth it for her. In this context, **Blue**'s

question allows her to get information from White while giving as little information as possible back. Solving **Blue**'s question will allow **White** to get an answer to her original question, though it might not be the most honest or exhaustive one.

The current section presents a typology of questions and answers in dialogue with illustration from the real-life data corpus of oral multilogues, DinG. The beginning of this section shows the way questions can be classified according to their formulations, but also to the expected answers. Then, we present different answer types, separated by the resolving power. Next section focuses on a specific type of (sometimes non-sentential) questions called Clarification Requests.

### 2.3.2 Clarification Requests

Linguistic studies of dialogue based on real-life data show evidence of a frequently appearing type of utterance: *clarification request* (CR). They are designed to help the dialogue continue even when there are lapses in the process of exchanging conversational content. [Purver, 2004] defines clarification requests as “device[s] allowing a user to ask about some feature (e.g. the meaning or form) of an utterance, or part thereof”. Clarification requests can be used for different meta-communicational reasons [Ginzburg, 2012]:

- To ask for a repetition or to request clausal confirmation: a confirmation question concerning the semantic contribution of a particular constituent within the entire clausal content of the utterance;
- To confirm intended content: query on a content a speaker intends to associate with a given (sub-)utterance, independently of the remaining content of the clause being resolved or not;
- To correct the other speaker.

We would like our models to account for clarification request phenomena in dialogue in a computational (logic, compositional, dynamic) way, following implementations such as the CLARIE system [Purver, 2006] and formalizations such as the one using TTR in [Ginzburg, 2012]. In particular, we are interested in the action of clarification requests on the previous utterance. Corpus observations conducted on the British National Corpus (BNC, see [Burnard, 2000]) list several types of clarification requests [Purver et al., 2003], see table 2.4. Computationally speaking, how different is the action of those different types of clarification requests?

Questioning the action of different types of questions, in particular clarification requests, in dialogue, leads us to consider questions from a formal semantics point of view, closer to what we developed throughout Chapter 1. The next section focuses on formal semantics of questions.

## 2.4 Question semantics

We started this chapter with general considerations on dialogue as a whole. Then, we presented discussions that are relative to the questions and answers phenomena in dialogue; in particular, we gave definitions and classifications of different types of questions and answers. This gave us the capacity to consider our subject of study with more precision. The current section focuses on the semantics of questions, following the [Groenendijk and Stokhof, 1984] foundational article. The authors start by a presentation of the methodological principles that govern their studies, which are compositionality and links between compositionality, syntax, and semantics, close to the idea of a syntax-semantics interface. For a detailed presentation of these notions, see

Category	Description	Example
Wot	words used to request repetition	Eh? / What?
Explicit (exp)	context-independent CRs	What did you say? / Did you say 'Bo'?
Literal Reprise (lit)	verbatim repetitions of the utterance (& changing of the pronouns)	Did BO leave? / Did Bo LEAVE?
<i>Wh</i> -substituted Reprise (sub)	lit where a constituent is replaced by a <i>wh</i> -phrase	Did WHO leave? / Did Bo WHAT?
Reprise sluice (slu)	bare <i>wh</i> -phrase used as CR	Who? / What? / Where?
Reprise Fragments (RF)	bare phrase used as CR	Bo? / Leave?
Gaps	omitted targeted constituent	Did Bo ...?
Fillers	guess for the utterance intended in the unfinished antecedent	A: Did Bo... B: Win?

Table 2.4: Categories of clarification requests, [Purver et al., 2003].

Chapter 1, section 1.2. Then, Groenendijk and Stokhof state three essential attributes of a logically adequate formal semantics model of questions:

1. The model has to include a notion of *equivalence* between the interrogatives.
2. The model has to include a relation of *entailment* between interrogatives.
3. The model has to include a notion of *answerhood*.

The need for a notion of equivalence between the interrogatives comes from the will for the model to be logically adequate: if two interrogatives can be used in the same way for logical inferences, the two must be logically equivalent, even if the linguistic formulations are different. Example 37 introduces two interrogatives that are equivalent in meaning, as answering one would be equivalent to solving the other. It is important to note that Groenendijk and Stokhof, following Bulnap's work on the logic of questions and answers [Belnap, 1982], specify that treating the semantics of two equivalent questions equivalently does not necessarily imply that they should be completely the same, but rather that they should be assigned meanings which can be related to one another in a systematic way.

### Example 37 (Equivalent interrogatives, fabricated)

**A**<sub>1</sub> *Is Charlie older than Sasha?*

**A'**<sub>1</sub> *Is Sasha older than Charlie?*

The need for a notion of entailment comes from the will to draw inferences between the logical representations of utterances. If we want to be able to build logical argumentations using the semantic representation of utterances as arguments, we need to be able to draw relations between questions that are less than equivalent but more than unrelated. This is particularly true for questions that entail one another: if one expression entails another, they should be assigned meanings so that the meaning of one is included in the meaning of the other. Example 38 shows two questions, (1a) and (1b), that are entailed: an answer to (1a) would be part of the answer to (1b).

---

**Example 38 (Entailed questions, adapted from [Groenendijk and Stokhof, 1984])**

- (1) a. Which unicorns drink tea and talk?  
 b. Which unicorns drink tea?

In contrast, example 39 shows two questions that are not entailed, unless the complete set of unicorns we are talking about is defined beforehand.

**Example 39 (Not entailed questions, adapted from [Groenendijk and Stokhof, 1984])**

- (2) a. Which unicorns are drinking tea?  
 b. Which unicorns are not drinking tea?

Finally, a model of questions needs to include a notion of answerhood, even if the main focus stays on questions. The authors present three principal approaches to the theory of questions and answers, each of which constitutes an argument towards the need for a notion of answerhood. First, if questions are considered from a *categorial* point of view, they can be seen as incomplete objects, lacking the answer.

**Definition 21 (Categorial principle, [Groenendijk and Stokhof, 1984])** *The syntactic category and the semantic type of an interrogative are determined by the category and type of its characteristic constituent analysis.*

In the categorial approach, the notion of answerhood is necessary in order to be able to tell whether an answer is a right match to complete the “incomplete” question. The next theory is called the *propositional* approach: questions and answers are both seen as propositions.

**Definition 22 (Propositional principle, [Groenendijk and Stokhof, 1984])** *The semantic interpretation of an interrogative should give its answerhood conditions, i.e. it should determine which propositions count as its semantic answers.*

The need for a notion of answerhood is written verbatim in the definition of the propositional principle given in [Groenendijk and Stokhof, 1984], see definition 22.

Finally, the imperative epistemic approach shifts the perspective on questions and answers completely by taking a more external point of view. The questions are seen as imperatives, demanding for some information to be given.

**Definition 23 (Imperative epistemic principle, [Groenendijk and Stokhof, 1984])** *The semantic interpretation of an interrogative is a request for information (knowledge).*

In the imperative epistemic view, the only notion of answerhood that is needed is the one that would give us the capacity to evaluate whether the information given by the answer fulfills the request handed by the question.

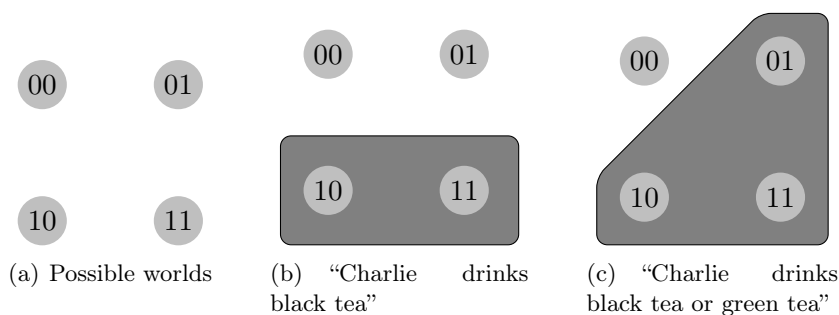
Groenendijk and Stokhof’s arguments towards the need for a notion of answerhood in semantic modeling of questions connect with the typological considerations for answers we presented in section 2.3. We attempt to assemble all these conceptions in the model we give in Chapter 5, section 5.2.

The next section showcases the treatment of questions in Inquisitive Semantics. Inquisitive Semantics’s core ideas stem from the discussions presented in [Groenendijk and Stokhof, 1984], but also in [Groenendijk et al., 1997], in particular in its positioning with regards to Charles Leonard Hamblin’s Alternative Semantics, [Hamblin, 1970, Hamblin, 1976], and the works of Lauri Karttunen [Karttunen, 1976, Karttunen, 1977].

### 2.4.1 Questions in inquisitive semantics

We have already started to give a presentation of Inquisitive Semantics (IS) in Chapter 1, sections 1.3 and 1.7. In the following section, we focus on IS’s approach to questions. As introduced before, inquisitive logic features two projection operators, ! and ?. The first one transforms any proposition into a purely informative one by canceling its inquisitive content, which corresponds to a declarative sentence. Conversely, the second one transforms any proposition (but the always true one) into a purely inquisitive one by canceling its informative content.

In order to illustrate the next notions with examples, let us consider the unicorn Charlie. Charlie either drinks green tea, or black tea, both, or no tea at all. Figure 2.1 shows the four possible worlds defined by Charlie’s preferences in hot beverages.



The first digit stands for Charlie drinking black tea (0 for ‘no’, 1 for ‘yes’), the second digit for Charlie drinking green tea. For the sake of simplicity, we only represent the maximal elements of the downward closed sets here, see Chapter 1, section 1.7 for the presentation of inquisitive propositions.

Figure 2.1: Visualization of possible worlds.

In fact, figure 2.1 shows the visualization of different propositions, standard for IS, in terms of possible worlds. The second and third illustrations in this figure correspond to propositions that are informative only, without an inquisitive component. This fact is reflected by the illustration, as the gray areas contain only one alternative for each of these pictures. Figure 2.1(a) illustrates the empty proposition, which is not inquisitive but is informative.

Following classical typologies of questions such as the ones presented in this chapter, section 2.3.1, IS provides ways to capture the semantics of different types of questions in English such as closed, alternative, open disjunctive, and *wh*- (mention-all and mention-some) questions. Consider the questions presented in example 40, where the hyphens, up arrow ( $\uparrow$ ), and down arrow ( $\downarrow$ ) are used to mark the prosody patterns for each sentence.



**Example 40 (What tea is Charlie drinking?)**

- (3) a. Does Charlie drink black-or-green tea<sup>↑</sup>?  
 b. Does Charlie drink black tea<sup>↑</sup> or is it green tea<sup>↑</sup>?  
 c. Does Charlie drink black tea<sup>↑</sup> or is it green tea<sup>↓</sup>?

(3a) is a single clause closed question, (3b) is a multi-clause open disjunctive question – it doesn’t presuppose exclusivity between the two alternatives “drinking black tea” and “drinking green tea”. (3c) is a multi-clause close alternative question because of the prosody pattern that gives an indication that the two alternatives are supposed to be exclusive.

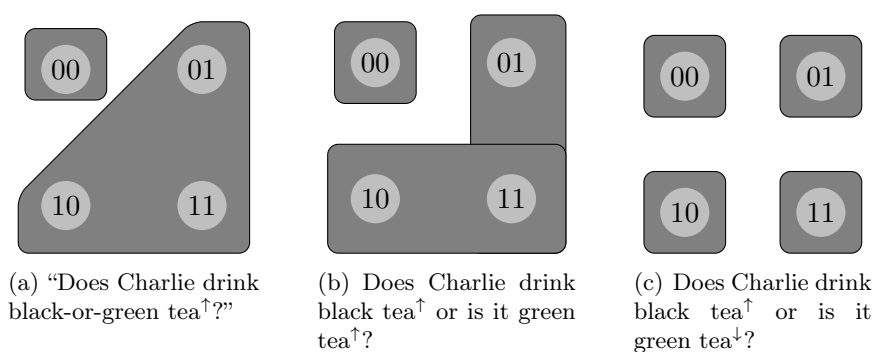


Figure 2.2: Visualization of the interrogative sentences from example 40.

Figure 2.2 shows the visualization corresponding to the questions presented in example 40. All of the propositions illustrated in this figure are inquisitive, which is illustrated by the fact that all three pictures feature several alternatives. As the three propositions considered in figure 2.2 are questions, them yielding semantic representations that are purely inquisitive corresponds to our intuition.

The next section presents an overview of the treatment of the semantics of *wh*-question in different formalisms and frameworks, starting with IS and continuing with SDRT and KoS.

## 2.5 Semantics of *wh*-questions

It is fairly usual, when studying questions, to separate *wh*-questions from the other types of questions. This is particularly true for works based on English, as *wh*-questions, in English, are quite easily identifiable by a fixed list of linguistic tokens (see section 2.3.1, table 2.2). For other languages, the linguistic cues can get more subtle, see in particular section 2.6.1 for our take on *wh*-words in French.

*Wh*-questions are subject to specific linguistic phenomena, in particular from the semantics and pragmatics point of view. We define *mention-all*, *mention-some*, and *single-match wh*-questions, following the presentation given in [Ciardelli et al., 2018].

### Definition 24 (Mention-all *wh*-questions, [Ciardelli et al., 2018])

Mention-all *wh*-questions ask for a complete specification of the individuals that have a certain property, i.e., for a specification of the complete extension of the property in the relevant domain of discourse.



Consider the questions presented in example 41. (4a) is an example of mention-all *wh*-question: in order to answer the question, we need to give the complete list of teas that Charlie possesses.

**Example 41 (Examples of *wh*-questions, adapted from [Ciardelli et al., 2018])**

- (4) a. *What tea does Charlie have?*  
b. *What is a drink that Charlie likes?*  
c. *Which tea is Charlie drinking right now?*

**Definition 25 (Mention-some *wh*-questions, [Ciardelli et al., 2018])**

Mention-some *wh*-questions ask for an instance of a certain property.

(4b) is an example of mention-some *wh*-question: the person asking the question doesn't need an extensive list of every drink that Charlie likes but only one instance of such a drink.

**Definition 26 (Single-match *wh*-questions, [Ciardelli et al., 2018])**

Single-match *wh*-questions ask for the unique individual having a certain property.

(4c) is an example of a single-match *wh*-question: the person asking the question wants to know the content of the cup Charlie is drinking right now, and only that, which is supposedly unique. Now that we have defined and exemplified subtypes of *wh*-questions, we can dive into details of different semantic representations proposed for *wh*-question throughout the literature.

### 2.5.1 Semantic representations of *wh*-questions

The following section presents an overview of semantic representation of *wh*-questions through three formalisms and frameworks:

- Inquisitive Semantics (IS), that we have extensively discussed through Chapter 1 and this current chapter, section 2.4.1;
- Segmented Discourse Representation Theory (SDRT), first introduced in Chapter 1, section 1.3 and presented more extensively in this current chapter in section 2.2.4;
- KoS, first introduced in Chapter 1, section 1.3 and presented more extensively in this current chapter in section 2.2.3.

We sketch the approach each of these frameworks takes to the semantics of *wh*-questions and put them in perspective of the aim of this thesis: producing a semantic model of dialogue.

**Inquisitive Semantics** We start by presenting the semantic representation of *wh*-questions in IS. Consider again the questions presented in example 41.

**Example 42 (Semantic representations of (4a))**

*The issue expressed by (4a) is solved in an information state (set of possible worlds)  $s$  if the information available in  $s$  determines exactly which individuals in the domain are teas in Charlie's pantry.*

$$\llbracket \text{What tea does Charlie have?} \rrbracket_{IS} = [\forall x? \text{ has Charlie } x]$$

The semantic representation of (4a) corresponds to IS’s notation for semantic representations of mention-all *wh*-questions. It uses the ‘?’ operator, which projects the following proposition (has Charlie  $x$ ) to a fully inquisitive one. Then, we are looking for all the  $x$  satisfying this inquisitive proposition.

Next, let’s consider the mention-some *wh*-question presented in example 41, (4b).

**Example 43 (Semantic representations of (4b))**

$$\llbracket \textit{What is a drink that Charlie likes?} \rrbracket = [\exists x \textit{ likes Charlie } x]$$

For “What is a drink that Charlie likes?”, the person asking the question only needs one instance of a drink that Charlie likes, which makes this question a mention-some *wh*-question. IS gives to this question a semantic representation that uses an existential operator. The existential operator introduces inquisitiveness to the proposition in its scope while specifying that we are looking for only one individual.

**Example 44 (Semantic representations of (4c))**

*In order to simplify the representation, we leave the time modality “right now” aside.*

$$\llbracket \textit{Which tea is Charlie drinking (right now)?} \rrbracket = [\exists x \textit{ drinks Charlie } x]$$

Example 44 shows the semantic representation of question (4c) from example 41. IS models single-match *wh*-questions as special cases of mention-some *wh*-questions. Mention-some *wh*-questions can be solved by giving at least one individual, while single-match *wh*-questions await one individual and one only. Yet, if we consider single-match *wh*-questions from the point of view of the partitioning of the logical space of possible worlds as the one we presented in section 2.4.1, they are quite different from mention-some *wh*-questions. The alternatives for the first always form a partition of a subset of the logical space, while this is not necessarily the case for the alternatives for the latter.

The presentation of IS we give here is narrowed down to the tools and intuitions that we need in the following chapters of this thesis. We strongly advise any interested reader to read [Ciardelli et al., 2018] so that they can form their own picture and understanding of this formalism. The next paragraph presents the SDRT approach to semantics of *wh*-questions.

**SDRT** As we mentioned in section 2.2.4, Segmented Discourse Representation Theory’s scope lies primarily within the semantics-pragmatics interface. Therefore, the treatment of questions in SDRT is twofold: first, [Asher and Lascarides, 2003] gives a general definition of the semantic value of a question. Then, the pragmatics of questions is addressed through the use of discourse relations such as the ones we list in section 2.2.4.

The formal definition of the semantic value of a question demands fluency in several logical notions that we do not use in the following of this thesis, therefore we will only give the intuition behind this definition, and advise any curious reader to refer to [Asher and Lascarides, 2003] for a complete formal account. Following [Groenendijk and Stokhof, 1984], the authors use  $\lambda$ -abstracts to represent interrogatives, where the  $\lambda$  binds the variable corresponding to the *wh*-phrase used in the *wh*-question. The semantics of a question is then defined in terms of its answers, with conditions on exhaustiveness different from the ones presented in [Groenendijk and Stokhof, 1984].

**Definition 27 (Informal definition of the semantic value of a question)**

Adapted from [Asher and Lascarides, 2003].

Consider the question  $Q = ?\lambda x_1 \dots x_n P(x_1, \dots, x_n)$ . The semantic value of  $Q$  is the set of propositions  $p$  such that:

- $p$  replaces each  $\lambda$ -variable  $x_i$  with a value  $\alpha_i$ ;
- $p$  is true;
- $p$  either entails that some particular objects occur in the extension of  $P(x_1, \dots, x_n)$  or it entails that no object occurs in the extension of  $P(x_1, \dots, x_n)$ .

To better understand the last point of definition 27, consider the following question:

(5) Who came to the party?

According to the first part of the last point of definition 27, “someone” does not constitute an answer to this question, even though it is a partial answer. On the contrary, “no one” is considered to be an answer to the question (if it is true), according to the second part of the last point of definition 27.

On a pragmatic level, *wh*-questions partake in the same discourse relations as the other types of questions. Among them, one can find in particular:

- **Question Answer Pair (QAP)**, that holds between a question and its true, direct answer;
- **Indirect Question Answer Pair (IQAP)**, that holds between a question and an utterance which content is sufficient for a participant of the conversation to compute a direct answer using their knowledge and beliefs;
- **Partial Question Answer Pair (PQAP)**, that holds between a question and an utterance which content gives a participant of the conversation the necessary information, using their knowledge and beliefs, to eliminate at least one direct answer (but not enough information to compute a true direct answer).

The next paragraph focuses on the treatment of *wh*-questions in KoS, which is a formalism that models dialogue on a pragmatic level, from a context-centered point of view.

**KoS** In section 2.2.3, we defined the dialogue gameboard (DGB) model of dialogue. Similarly to SDRT, KoS models questions in two ways: on one hand, through their semantic content, represented with a record<sup>38</sup>; on the other hand, through their effect on the dialogue gameboard, through conversational rules. We define the semantic content of questions following [Ginzburg, 2012, Maraev et al., 2018].

**Definition 28 (Questions in KoS, [Maraev et al., 2018])** Let  $A$  be the expected type of an answer,  $Q$  a family of propositions such that for any answer  $a$ ,  $Q(a)$  is the meaning of an answer  $a$  as a proposition. Then, the types of polar questions<sup>39</sup> and *wh*-questions are defined as follows:

$$\text{PolarQuestion} =_{def} \left[ \begin{array}{ll} A = Bool : & \text{Type} \\ Q : & A \rightarrow Prop \end{array} \right]$$

<sup>38</sup>See Chapter 1, section 1.3 for pointers on Type Theory with Records (TTR).

<sup>39</sup>[Maraev et al., 2018] uses the term “polar” to designate what we have called “closed” questions in this thesis, see definition 14.

$$\text{UnaryWhQuestion} =_{\text{def}} \left[ \begin{array}{l} A = \text{Ind} : \quad \text{Type} \\ Q : \quad A \rightarrow \text{Prop} \end{array} \right]$$

Example 45 illustrates this definition on a closed and a *wh*-question.

**Example 45 (Questions in KoS, [Maraev et al., 2018])**

$$\llbracket \text{“Do you live in Paris?”} \rrbracket = \left[ \begin{array}{l} A = \quad \quad \quad \text{Bool} \\ Q = \quad \lambda a. \text{IF } a \text{ THEN } \text{live}(\text{Paris}) \text{ ELSE } \neg \text{live}(\text{Paris}) \end{array} \right]$$

$$\llbracket \text{“Where do you live?”} \rrbracket = \left[ \begin{array}{l} A = \quad \quad \text{City} \\ Q = \quad \lambda a. \text{live}(a) \end{array} \right]$$

This treatment of questions is very similar for closed questions and *wh*-questions. In general, it is quite common in the literature to view all types of questions as variations on *wh*-questions. This comes from the fact that some of the *wh*-questions have relatively simple semantics: *where*-questions ask for a location, *when*-questions ask for a time. We develop these considerations further in section 2.6.

The main conversational rule triggered by a *wh*-question is **Ask *wh*-question QUD-incrementation**. This rule updates the DGB by adding a new issue to the QUD; while its effects are fairly standard, [Ginzburg, 2012] gives a detailed list of pre-conditions for the application of this rule. The reason behind all the precautions comes from a will to include Non-Sentential Utterances (NSUs) in the model and in particular, the ones that need a specification on part of the preceding utterance in order to be understood. This type of utterances is called *focus-establishing constituent* (FEC) in [Ginzburg, 2012]. Consider the fragment of conversation given in example 46.

**Example 46 (Fragment of conversation with an FEC, [Ginzburg, 2012])**

**A**<sub>1</sub> A student complained about one of our teachers.

**B**<sub>2</sub> Who?

The *wh*-word “who” in **B**’s utterance establishes the focus, in the previous utterance, on either of the two individuals (student or teacher). In order for **B**<sub>2</sub> to be an acceptable dialogue move, **A**<sub>1</sub> has to contain at least one entity of type individual.

Both KoS and SDRT apprehend dialogue modeling from a meta point of view: they establish rules that oversee dialogue coherence and account for the pragmatics of utterances. KoS uses relatively simple semantic representations, that are integrated into the overall TTR-based representation of the DGB. SDRT leaves the door open for compositional computation of the semantic representation of utterances through a chosen syntax-semantics interface, which results in quite heavy but very complete formulae. The semantic representations of utterances are then linked through discourse relations, that implement the coherence constraints.

The formalisms presented in the current section all present a high-level model of *wh*-questions, slightly distanced from linguistic cues considerations we developed in section 2.3. The next section takes up the complementary approach, presenting links between *wh*-word in English, then in French, and thematic roles.

## 2.6 *Wh*-words – thematic roles correspondence

We have given an introduction to the notion of thematic roles in section 1.6.2, along with several possible lists. Thematic roles, in general, are determined by both the grammatical function of the analyzed semantic constituent of the sentence (i.e subject or object of the verb) and the central predicate of the sentence (usually the verb). Switching a sentence to the passive voice changes the grammatical functions of the constituents (as subjects become objects and vice-versa) but does not change their thematic roles, because the predicate stays unchanged.

The goal of this section is to outline the possibility of a correspondence between *wh*-words and thematic roles, in view of an automation of the computation of semantic representations of utterances involving *wh*-words. It is difficult to come up with a set of thematic roles which would both be exhaustive, allowing to represent any possible sentence, and computationally realistic – without overlapping and/or redundant scopes of thematic roles (one sentence constituent should ideally correspond to one and only one thematic role). There is a great diversity in thematic role sets one can find in the literature. An example of commonly used thematic roles is presented in [Jurafsky and Martin, 2015], see table 2.5.

THEMATIC ROLE	DEFINITION
Agent	The volitional causer of an event. ex: <i>The waiter</i> spilled the soup
Experiencer	The experiencer of an event. ex: <i>John</i> has a headache
Force	The non-volitional causer of an event. ex: <i>The wind</i> wrecks the house
Theme	The participant most directly affected by an event. ex: Brutus stabbed <i>Caesar</i>
Result	The end product of an event. ex: The city built <i>a house</i>
Content	The proposition or content of a propositional event. ex: Mary said “ <i>I met John in the supermarket</i> ”
Instrument	An instrument used in an event. ex: Brutus stabbed Caesar with <i>a knife</i>
Beneficiary	The beneficiary of an event. ex: Mary bought a gift for <i>John</i>
Source	The origin of the object of a transfer event. ex: John flew in from <i>Paris</i>
Goal	The destination of an object of a transfer event. ex: Mary drove to <i>Warsaw</i>

Table 2.5: Commonly used thematic roles.

A first glance on this list of roles throws us back to the main issue presented before – the non-exhaustivity of such a list. In the following, we chose to build our own set starting from the previous one by pairing, when possible, one *wh*-word to one thematic role. Note that we do not try to give a universal approach here but merely aim for an outline that can help for implementation considerations. We will consider the three following groups of *wh*-words, partitioned according to similarity and increasing semantic complexity considerations.

**Group 1** Who, Whom, Whose

**Group 2** Where, Why, When

**Group 3** What, Which, How

### Group 1

**Who** According to table 2.5, three thematic roles correspond to “Who” – **Agent**, **Experiencer**, and **Theme**. When “Who” is in an object position, the corresponding thematic role is **Theme**; whenever it is in a subject position, it can be either **Agent** or **Experiencer**. For the sake of simplicity, we unite these two latter roles under the thematic role **Agent**.

**Whom** According to table 2.5, two thematic roles correspond to “Whom” – **Theme** and **Beneficiary**. For the sake of simplicity and following our pairing rule, we unite these roles under the thematic role **Theme**.

**Whose** As no thematic role can clearly be identified as corresponding to “Whose”, we create the role **Possessor**.

### Group 2

**Where** As discussed above, the **position**, **place** awaited as an answer to a question beginning with “Where” can be viewed as objects corresponding to the thematic roles **Source** or **Goal**. As our main aim is to ease computability and reduce redundancy, we combine those two thematic roles in one, that we will call **Location**.

**Why** The word “causer” that can be found both in the definitions of **Agent** and **Force** can be an indication of the fact that those are the thematic roles that should be paired with “Why”. As those roles are very unspecific, we chose here to create a new thematic role, identified by the fact that the corresponding sentence constituent should be introduced by “because”. We call this role **Reason**.

**When** It is interesting to notice that Jurafsky and Martin’s list does not contain a single role able to represent temporal data. The simplest solution is to create a thematic role called **Temporality**. The difficulty lies in the multiplicity of temporal representations that language can come up with: first, durations can be relative (ex: *since 2015*) or absolute (ex: *for 5 years*). Then, one has to distinguish punctual durations (ex: *on October the 5th, 2015*) from time intervals (ex: *in November 2015*). On top of that (and this is where the conjunction of semantic representations can fail), it is important to notice that a set of punctual durations is not equivalent to a time interval: “every Thursday” cannot be considered as a time interval. Yet, it is possible to find punctual durations inside all time intervals (ex: *every Thursday of November 2015*). Therefore, though we ignore the difficulties here by pairing “When” with **Temporality**, they do not disappear.

### Group 3

**What** The problem of “What” lays in the non-specificity of its usage. Picking among the roles in table 2.5, an answer to a question starting by “What” could be a **Force**, a **Result**, a **Content** or an **Instrument**, depending on the main predicate. Such a multiplicity of roles is unmanageable regarding our pairing task. Yet, these thematic roles are so different that uniting them under one designation would badly hurt the expressivity of our model.

Therefore, it is necessary, when considering questions introduced by “What”, to look at the whole *wh*-phrase contained in the interrogative. Then, the thematic role corresponding to the answer will be the one semantically typing the focus phrase of the interrogative (ex: “*What time is it?*” corresponds to the thematic role **Temporality**).

**Which** This case should be treated in the same way as the previous one. Therefore, “Which + focus phrase” corresponds to the thematic role of the focus phrase. Ex: “*Which city do you prefer, Paris or London?*” call for an answer of semantic type **Location**.

**How** The case of “How” is a delicate one, as, as in the two previous cases, the answers expected for a question starting by “How” vary greatly depending on the phrase that follows the interrogative word (one can think about the difference between “How much” and “How good”). Yet, unlike for the answers to questions starting by “What” and “Which”, it is possible to come up with a unique designation for all the answers to questions starting by “How”: we call this thematic role **Characteristic**. Still, as for **Temporality**, this unique notation only hides difficulties without dismissing them.

To sum up the discussion above, we present the thematic roles we chose with definitions and examples in table 2.6.

Thematic role	Definition
Agent	The volitional causer of an event. ex: <i>The waiter</i> spilled the soup
Theme	The participant most directly affected by an event. ex: Brutus stabbed <i>Caesar</i>
Location	The place where an event happens. ex: I live in <i>Paris</i>
Reason	Explanation given for an event. ex: I am happy because <i>she's getting married</i>
Temporality	Time when an event happens. ex: I go to the pool <i>every Thursday</i>
Possessor	The owner or reference person for an object or a characteristic. ex: I forgot <i>Cheryl's</i> birthday
Characteristic	The characterisation of an event. ex: I dropped the book <i>loudly</i>

Table 2.6: New set of thematic roles.

Finally, keeping in mind the previous discussion on one-to-one *wh*-word to thematic role pairing, we obtain table 2.7. This table allows us to systematically construct semantic representations of lexical terms.

Though we changed the original list of thematic roles presented in table 2.5, these changes are not significant enough for us not to be able to use the computability results given in [Jurafsky and Martin, 2015].



<b>Wh-word</b>	<b>Thematic roles</b>
What + focus phrase	<b>Thematic role</b> (focus phrase)
When	<b>Temporality</b> (Temp.)
Where	<b>Location</b> (Loc.)
Who	<b>Agent, Theme</b> (Ag., Th.)
Whom	<b>Theme</b> (Th.)
Which + focus phrase	<b>thematic roles</b> (focus phrase)
Whose	<b>Possessor</b> (Pos.)
Why	<b>Reason</b> (Re.)
How	<b>Characteristic</b> (Ch.)

Table 2.7: *Wh-words and thematic roles pairing.*

### 2.6.1 French Interrogatives

The typology that is presented in the previous section stands first and foremost for English. Chapters 3 and 4 introduce and develop the studies we conducted on DinG, a corpus of French oral data. In order to operationalize our approach to DinG, it is important to see how and to what extent can studies conducted on English be transposed to French. The following section shows how the classification of interrogatives can be more tricky in French than in English and draws some correspondences between the two languages.

One can distinguish two types of French interrogatives: total ones, corresponding to English closed questions, and partial ones, corresponding to English *wh*-ones [Riegel et al., 1994]. Unlike in English, in French, partial questions can be driven by multiple morphological variations of interrogative pronouns and adverbs, which are not linguistically identified as easily as *wh*-words. Table 2.8 presents correspondences that can be drawn. It was constructed according to the following process:

1. Retrieval of a list of English *wh*-words from [Aarts et al., 2014].
2. Retrieval of a list of French interrogative pronouns and adverbs from [Riegel et al., 1994].
3. Translation of the English set of words into French, using **Reverso**.
4. Translation of the English set of words into French, using **Linguee**.
5. Translation of the French set of words into English, using **Reverso**.
6. Translation of the French set of words into English, using **Linguee**.
7. Compilation of the previously obtained information.
8. Verification, using **Systran**.

Now, the theories developed in the previous sections can be applied, giving the French interrogative words to thematic roles correspondence (see table 2.8).

It is interesting to notice that though French separates “How much” (« combien ») from the other questions beginning by “How”, the thematic role-interrogative word pairing stays relevant, as our definition of the thematic role set scopes is inclusive enough. The considerations developed in the current section are extremely useful for us both in the practical parts of our work, in Chapters 3 and 4, and for the more theoretical models presented in Chapter 5.



<b>Wh-word</b>	<b>French equivalents</b>	<b>Thematic roles</b>
What	que, qu' quoi, de quoi quel, quelle, quels, quelles	Thematic role (focus phrase)
When	quand	Temporality (Temp.)
Where	où	Location (Loc.)
Who	qui quel, quelle, quels, quelles lequel, laquelle, lesquels, lesquelles	Agent, Theme (Ag., Th.)
Whom	qui lequel, laquelle, lesquels, lesquelles	Theme (Th.)
Which	qui lequel, laquelle, lesquels, lesquelles auquel, à laquelle, auxquels, auxquelles	Thematic role (focus phrase)
Whose	à qui	Possessor (Pos.)
Why	pourquoi	Reason (Re.)
How	en quoi comment	Characteristic (Ch.)
How (much)	combien	

Table 2.8: English to French correspondence.

## 2.7 Conclusion of the chapter

Chapter 1 introduces the linguistic and logic phenomena that are at the core of the work we develop in this thesis. Then, we present an overview of the formal semantics formalisms and frameworks and put them in perspective of the aforementioned phenomena. As we aim for a model of dialogue, the current chapter operates a shift in perspective from discourse to dialogue.

We give a general presentation of approaches to dialogue, starting with the Gricean maxims and a first incursion in annotation schemas. We will come back in details to annotation schemas in Chapter 4, as annotation has constituted an important part of our dialogue studies. Then, we present one of the central notions for our work: the dialogue context, which stores the information exchanged throughout the dialogue and is used to feed the utterances while they are produced. Finally, we go back to two formalisms that we already introduced in Chapter 1: KoS and SDRT. In the current chapter, we focus on details of these two formalisms that are of particular interest for us in this thesis, in the approach we take to dialogue modeling. While SDRT constitutes an important reference point for us, we directly use ideas inspired by KoS to build some of the models presented in Chapter 5.

Next, we narrow down our presentation and focus on utterances in dialogue, in particular on questions and answers and mechanisms underlying the relationship between the two. We give a typology of questions and answers following classical references, along with a presentation of clarification requests. We exemplify using the DinG corpus, see Chapter 3, section 3.3.2 for a complete description.

From questions and answers in dialogue, we shift our focus again and concentrate on question semantics, specifically from a formal semantics point of view, making a link with the considerations developed in Chapter 1. In particular, we give the intuitions from question modeling in Inquisitive Semantics that are central to the models we present in Chapter 5, sections 5.3 and 5.4.

The final narrowing we do in this chapter leads us to an extensive presentation of the way *wh*-questions are considered across different formalisms. This then leads us to develop our own considerations on the correspondence between *wh*-words and thematic roles, in view of an operationalization of approaches using thematic roles for the semantic representation of utterances (such as, for instance, NDES). We end this chapter with a section on French interrogatives, that presents a high practical interest for us, in particular in Chapter 4, because we aim to model real-life data and ours happens to be in French.



Part II

Real-life data



# 3

## Real-life corpora

### Contents

---

<b>3.1 Introduction</b> . . . . .	<b>83</b>
<b>3.2 A corpora herbarium</b> . . . . .	<b>85</b>
3.2.1 An overview of corpora of English . . . . .	86
3.2.2 An overview of corpora of French . . . . .	88
<b>3.3 Dialogues in Games (DinG)</b> . . . . .	<b>97</b>
3.3.1 A game of <i>Catan</i> . . . . .	99
3.3.2 Corpus description . . . . .	100
3.3.3 Collection process . . . . .	101
3.3.4 Data processing . . . . .	102
<b>3.4 DinG in the corpora landscape</b> . . . . .	<b>109</b>
<b>3.5 Conclusion of the chapter</b> . . . . .	<b>111</b>

---

### 3.1 Introduction

The work presented in this thesis mostly comes from a computational linguistics point of view – our goal is to study natural languages and create models that would be logically sound yet would give a good enough approximation of the real phenomena. As natural languages are our objects of study, it seems that creating data to model would be easy, as all of us are “natural language speakers”. Indeed, the first outlines of a computational linguistics model are drawn based on handmade examples, designed to highlight the particular phenomena that are being studied. In our case, the first corpus we used to design our models is the Unicorn Corpus (UniC). UniC is a toy handmade corpus of sentences in English and French, 18 in each language: 9 questions (1 polar + 8, one per *wh*-word) and 9 corresponding declarative sentences. See example 47 for one interrogative/declarative sentence pair and tables 3.1 and 3.2 for the complete corpus.

**Example 47** *Example of an interrogative/declarative sentence pair in UniC, English and French.*

<b>Where-question</b>	<i>Where is the unicorn?</i>
	<i>Où est la licorne ?</i>

**Where-answer**      *The unicorn is at home.*  
                           *La licorne est à la maison.*

Example 47 is designed to model the question/answer relationship between a *wh*-question starting with the word “where” and a complete (non-elliptic) answer. As we show in Chapter 1, formal semantics models of natural language mainly focus on fully formed declarative sentences such as the *where-answer* in example 47. Therefore, it is important for the development of a new model to check that it behaves similarly to the previous ones on similar sentences. Then, our object of study is dialogue and more specifically the question/answer relationship. As shown in Chapter 2, a first classification of questions separates the *wh*-questions from the closed questions. We created the examples in UniC to acknowledge this first typology. We also show in Chapter 2 that the *wh*-words in English to French equivalents correspondence introduces a multiplicity of formulations and phenomena. We produce the French examples in UniC to account for this aspect.

WH-WORD	QUESTION (ENGLISH)	ASSERTION (ENGLISH)
∅	Is Charlie a unicorn?	Charlie is a unicorn.
What	What colour is the unicorn?	The unicorn is blue.
When	When will the unicorn grow?	The unicorn will grow soon.
Where	Where is the unicorn?	The unicorn is at home.
Who	Who is Charlie?	Charlie is a unicorn.
Whom	Whom is the unicorn talking to?	The unicorn is talking to Charlie.
Which	Which type of unicorn is Charlie?	Charlie is a blue unicorn.
Whose	Whose unicorn is Charlie?	Charlie is a free unicorn.
Why	Why is Charlie a unicorn?	Because unicorns are great.
How	How big is the unicorn?	The unicorn is small.

Table 3.1: UniC – English.

WH-WORD (ENGLISH)	QUESTION (FRENCH)	ASSERTION (FRENCH)
∅	Est-ce que Charlie est une licorne ?	Charlie est une licorne.
What	De quelle couleur est la licorne ?	La licorne est bleue.
When	Quand la licorne grandira-t-elle ?	La licorne grandira bientôt.
Where	Où est la licorne ?	La licorne est à la maison.
Who	Qui est Charlie ?	Charlie est une licorne.
Whom	À qui parle la licorne ?	La licorne parle à Charlie.
Which	Quel type de licorne est Charlie ?	Charlie est une licorne bleue.
Whose	À qui est cette licorne ?	Charlie est une licorne libre.
Why	Pourquoi Charlie est une licorne ?	Parce que les licornes sont géniales.
How	De quelle taille est la licorne ?	La licorne est petite.

Table 3.2: UniC – French.

Working on UniC has allowed us to start developing our model in a very controlled environment. Sentences from UniC contain no anaphora resolution, no elliptic construction except for the “Because” answer, therefore the dynamic phenomena in UniC are produced by the question/answer relationship. Assertions from UniC can be modeled with any state-of-the-art framework presented in Chapter 1. Thus, UniC is a very good starting point for our model’s development.

We pursued the development of our model by conducting annotation experiments on real-life data in English, Dutch, Mexican Spanish [Cruz Blandon et al., 2019], French, Mandarin

Chinese, and Northern Italian [Amblard et al., 2019], see Chapter 4 for a detailed account. While conducting these experiments, we encountered difficulties we didn't anticipate from our work with toy examples: sometimes, it is difficult to define what a question is; it is even more difficult to determine where an answer starts and ends; finally, if an answer doesn't directly solve the question, is it still an answer? What are the scopes of these phenomena?

From these observations came the conclusion: in order to confront the formal models with real-life data, this real-life data had to be chosen wisely, in a way that the piling-up of real-life phenomena would still allow us to identify which one is causing trouble to the model at the moment. It also appeared that while a lot of data was available on English, even wide-spread languages such as French suffered from a lack of freely available transcribed oral data.

Throughout the literature review, we came upon the Strategic Conversations<sup>40</sup> (STAC) project. The corpus of STAC is composed of chat logs from an online version of the board game *Catan*<sup>41</sup>, played by English speakers. STAC is annotated in Segmented Discourse Representation Theory (SDRT, [Asher et al., 2016]). As STAC is composed of chat logs, it cannot directly be used in order to test the models we develop for real-life oral data; yet, the dialogues from STAC are simple enough and use little enough vocabulary to be very interesting for us as they allow us to keep the amount of observed phenomena under control. As we wanted to use the existing research and the local (French) context, restrain the scope of the study to not encompass too many different phenomena, and be able to focus on questions and answers articulation in real-life dialogue, we decided to create Dialogues in Games (DinG): a corpus of transcriptions of real-life, oral, spontaneous multi-party dialogues between French-speaking players of *Catan*.

The current chapter unravels as follows: first, we present an overview of corpora of composed mostly of transcribed oral, but also written, dialogues in English, in section 3.2.1, then in French, in section 3.2.2.

Then, we introduce the DinG corpus. Section 3.3 gives a precise description of its constitution process, through collection, segmentation, transcription, and anonymization. In section 3.4, we compare DinG to the existing corpora and then outline some of the future development directions for our corpus.

## 3.2 A corpora herbarium

As the subject of our study is dialogue in general and questions and answers in particular, our resources review was first focused on corpora of transcribed oral dialogues, in English, with an explicit marking of interrogatives. Then, as we did not want to work on English only and because of the local (French) context of our lab, we expanded the scope of our review to corpora of transcribed oral French and corpora of French with explicit marking of interrogatives. Finally, while STAC is neither a corpus of transcribed oral dialogues nor of French, it was of high interest to us with respect to studies of questions and answers, as it contains an explicit marking of interrogatives and simple, yet long, multi-party dialogues.

---

<sup>40</sup><https://www.irit.fr/STAC/index.html>

<sup>41</sup><https://www.catan.com/>, visited on 09/03/2021.



### 3.2.1 An overview of corpora of English

Among the numerous corpora of dialogues in English, we consider the Spoken British National Corpus (**S-BNC**, [Love et al., 2017]), the Saarbrücken Corpus of Spoken English (**SCoSE**, [Norrick, 2017]) and the Strategic Conversation corpus (**STAC**, [Asher et al., 2016]). All three corpora have explicit marking of interrogatives.

Corpus	context	# words	# recordings	marking of interrogatives
S-BNC	misc., British English native speakers	11.5 million	1,251	yes
SCoSE p1	informal conversations	32,826	14	yes
STAC	chat logs, Catan	N.A.	59	yes

Table 3.3: English corpora – summary.

**S-BNC** S-BNC contains transcriptions of oral conversations between native speakers of British English. S-BNC was collected between 2012 and 2016 and contains around 11.5 million words through 1,251 recordings. S-BNC contains an explicit marking of interrogatives and appears to us as a corpus of major importance in the scope of our studies, as S-BNC has been the subject of several fundamental studies of dialogue and question-answer relationship, see in particular [Purver et al., 2003], [Ginzburg, 2012].

#### Example 48 (S-BNC, K62, Walsall Local Studies Centre, interview)

*Speech recorded in leisure context.*

PS5B1\_101 : Did, did, did, did they have the union in there at all or (pause)?  
 PS5B2\_102 : Have what dear?

Example 48 has been adapted in format to make it readable by humans. Transcriptions in S-BNC come in XML format with several layers of meta-data and annotation, see figure 3.1. They contain explicit transcription of pauses (without indication of length), onomatopoeia, laughter, and punctuation marking.

5511 ▾	<code>&lt;u who="PS5B2"&gt;</code>
5512 ▾	<code>&lt;s n="102"&gt;</code>
5513	<code>&lt;w c5="VHB" hw="have" pos="VERB"&gt;Have &lt;/w&gt;</code>
5514	<code>&lt;w c5="DTQ" hw="what" pos="PRON"&gt;what &lt;/w&gt;</code>
5515	<code>&lt;w c5="NN1" hw="dear" pos="SUBST"&gt;dear&lt;/w&gt;</code>
5516	<code>&lt;c c5="PUN"&gt;?&lt;/c&gt;</code>
5517	<code>&lt;/s&gt;</code>
5518	<code>&lt;/u&gt;</code>

Figure 3.1: View of the XML code corresponding to PS5B2\_102, example 48, S-BNC.

**SCoSE** SCoSE is a corpus of oral English constituted by the University of Saarland. It consists of eight parts: (1) Complete Conversations, (2) Indianapolis Interviews, (3) Jokes, (4) Drawing Experiment, (5) Kassel Classroom Discourse, (6) Stories, (7) London Teenage Talk, and (8) Musicians' Talk<sup>42</sup>. We only consider SCoSE part 1 for this overview, as the other parts were

<sup>42</sup><https://www.uni-saarland.de/lehrstuhl/gergel/scose.html>, visited on 18/03/21.

composed of data that was too task-oriented (parts 2, 3, 4, 5, 6) or made publically available too late for us to include them in the studies presented in this thesis. Of course, further comparisons should include SCoSE parts 7 and 8 as well. SCoSE part 1 is composed of recordings of informal conversations in a friend and family setting, made by two students from an American university.

#### Example 49 (SCoSE part 1, Addie and Brianne)

161. Brianne but I'm on a limited budget [right now]  
 162. Addie [I KNOW]  
 163. you don't have that much  
 164. you can't [help it]  
 165. Brianne [that's the] thing  
 166. Addie I know (2.0)  
 167. [yeah]  
 168. Brianne [so-]  
 169. Addie that's kind of bad  
 170. Brianne uh-huh

Transcriptions of dialogues in SCoSE (part 1) have numbered speech turns, take into account overlaps, pauses, interruptions, and onomatopoeia. Example 49 shows speech turns from 161 to 170. The end of 161 overlaps with 162, which is materialized by the usage of brackets (no time codes). Speech turn 166 ends with a 2 seconds pause, transcribed by the duration between parenthesis. Brianne's utterance 168 is cut short and overlaps with 167, so a reasonable assumption would be to say that she interrupted herself. Speech turn 170 is a backchanneling onomatopoeia.

**STAC** As mentioned before, the Strategic Conversation corpus (STAC) [Asher et al., 2016] is based on the board game *Catan*. Unlike all the corpora considered in this chapter, STAC is composed of chat logs, in English.

#### Example 50 (STAC, s1-league2-game4)

72 Server *sabercat rolled a 2 and a 5.*  
 73 *sabercat sorry*  
 74 *skinnylinny No worries*  
 75 *sabercat skinny, got sheep?*  
 76 *skinnylinny Nope*  
 77 *skinnylinny Also, neither of you can build a settlement without sheep...*

As STAC is composed of chat logs, we directly have access to the “raw” dialogue, as written by its participants, as opposed to an interpretation given by a transcriber. A first level of annotation that helps to navigate through the logs is the numbering of the chat entries: in example 50, from 72 to 77. The chat logs contain not only the player's utterances but also automatic entries from the server, corresponding to actions happening in the game. In example 50, entry 72 is an automatic message that appears after the player *sabercat* rolled two virtual dices and got a 2 and a 5. This is public information that is available to all players.

In the following section, we present major corpora of transcribed French multilogues.

### 3.2.2 An overview of corpora of French

We start our presentation of French corpora by the ones in the Orféo project<sup>43</sup>, composing the Study Corpus for Contemporary French (Corpus d’Étude pour le Français Contemporain, CÉFC) [Benzitoun et al., 2016]. We selected six corpora:

- **CFPP2000**: the Spoken Parisian French Corpus, Corpus de Français Parlé Parisien<sup>44</sup>, [Branca-Rosoff et al., 2009].
- **CLAPI**: the corpus of spoken language in interaction, Corpus de LAngue Parlée en Interaction<sup>45</sup>, [Balthasar and Bert, 2005].
- **C-ORAL-ROM**: the French part of this multilingual resource of spontaneous speech for Romance languages<sup>46</sup>, [Cresti et al., 2004].
- **CRFP**: the reference corpus of spoken French, Corpus de référence du français parlé, [Delic, 2004].
- **TCOF**: the corpus from the treatment of oral corpora in French project, Traitement de Corpus Oraux en Français<sup>47, 48</sup>, [ATILF, 2020, André and Canut, 2010].
- **TUFS**: the French part of the Tokyo University of Foreign Studies corpus<sup>49</sup>.

Corpus	context	# words	length (hours)
CFPP2000	interviews about Paris	783,017	~ 60
CLAPI	misc.	~ 170,000	17
C-ORAL-ROM	misc.	500,000	40
CRFP	private/professional/public	440,000	36
TCOF	misc.	~ 300,000	23
TUFS	students, misc.	~ 750,000	54

Table 3.4: CÉFC copora – data.

We did not consider other corpora among the ones composing the CÉFC because they were not composed of transcribed dialogues or multilogues. Four additional corpora are included in this overview:

- **Rhapsodie**: a prosodic and syntactic treebank for spoken French<sup>50</sup> [Lacheret-Dujour et al., 2019].
- **FQB**: the French QuestionBank<sup>51</sup>, [Seddah and Candito, 2016].
- **ESLO**: the city of Orleans corpus from the sociolinguistic study in Orleans project, Enquête Sociolinguistique à Orléans<sup>52</sup> [Eshkol-Taravella et al., 2011].

<sup>43</sup><https://repository.ortolang.fr/api/content/cefc-orfeo/10/documentation/site-orfeo/corpus-source/index.html>, visited on 17/05/2021.

<sup>44</sup><http://cfpp2000.univ-paris3.fr/>, visited on 17/05/2021.

<sup>45</sup><http://clapi.icar.cnrs.fr> visited on 13/04/2021.

<sup>46</sup><http://www.elda.org/en/proj/coralrom.html>, visited on 19/05/2021.

<sup>47</sup><http://www.cnrtl.fr/corpus/tcof/>, visited on 17/05/2021.

<sup>48</sup><https://www.ortolang.fr/market/corpora/tcof>, visited on 21/05/2021.

<sup>49</sup>[http://www.coelang.tufs.ac.jp/english/language\\_function.html](http://www.coelang.tufs.ac.jp/english/language_function.html), visited on 13/04/2021.

<sup>50</sup><https://rhapsodie.modyco.fr/francais-parle/>, visited on 23/05/2021.

<sup>51</sup><http://alpage.inria.fr/Treebanks/FQB/>, visited on 17/05/2021.

<sup>52</sup><http://eslo.huma-num.fr/>, visited on 23/05/2021.

- **SLAM**: the corpus from the Schizophrenia and Language, Analysis and Modeling project, [Amblard et al., 2015].

All these resources have been the subject of numerous publications and showcase transcriptions of overlaps, disfluencies, non-verbal expressions, pauses, and segmentation in illocutory units.

Rhapsodie and ESLO are major corpora of spoken French, SLAM is a corpus of confidential dialogues that constitutes an opening for extremely interesting further research projects and interrogations on data privacy. FQB is the state-of-the-art corpus for research on questions in French, which makes it relevant in the perspective of the studies presented in this thesis.

**Note.** All examples in this section come from corpora in French. In order to make them understandable to non-French speakers reading this thesis, all corpora excerpts are followed by translations in English.

**CFPP2000** The Spoken Parisian French Corpus (Corpus de Français Parlé Parisien) is composed of interviews based on surveys about the city and districts of Paris and its surrounding area.

**Example 51 (CFPP2000 [02-01] 02-01\_Lea\_Samvarian\_F\_55\_2e)**

*Interview about the reasons **spk2**'s parents came to Paris.*

**spk1** [50.111] : mm mm donc ils ont souhaité venir à Paris + (oui)  
pa- pour le le travail ?

**spk1 spk2** [55.647] : [1] parce qu'il y avait de la famille parce que [2] pour le travail ++ non pas du tout

**spk2** [58.353] : enfin mon père avait son frère aîné qui était monté à Paris + et son petit frère donc mon papa l'a rejoint pour euh pour travailler à Paris [mm mm] + voilà +

**spk1** [50.111] mm mm so they wished to come to Paris + (yes) fo- for work ?

**spk1 spk2** [55.647] [1] because there was family because [2] for work ++ no not at all

**spk2** [58.353] well my father had his elder brother who had gone up to Paris + and his little brother so my dad joined him to hum to work in Paris [mm mm] + yeah

Interviews from CFPP2000 are available in two written formats, **transcriber** and **HTML**. **Transcriber** files are aligned with the oral recordings while **HTML** ones are more human-readable. Example 51 shows the **HTML** version of a transcription. This example features two speakers, **spk1** and **spk2**. Each speech turn is given with its precise time code (ex: [50.111]) and the locutor(s). At [55.647], there is an overlap between **spk1**'s and **spk2**'s utterance. Both utterances are transcribed, both speakers are written as locutors, and tokens [1] and [2] are used to differentiate the two utterances. Transcriptions from CFPP2000 contain explicit transcriptions of overlaps, pauses (+ for short ones, ++ for long ones), onomatopoeia (ex: mm mm) and laughter. The transcriptions contain both explicit annotation of questions (?) and annotation of rising intonation ([?]).

**CLAPI** The Corpus de Langue Parlée en Interaction (Corpus of Spoken Language in Interaction) is a corpus of real-life data recorded in various settings ranging from private conversations to commercial and professional ones. It is composed of 600 hours of interactional data, transcribed [Bert et al., 2010], of which 67 hours and 194 transcriptions are currently downloadable<sup>53</sup>. The available recordings last between 28 seconds and 3 hours, 19 minutes on average.

### Example 52 (CLAPI, Interaction between four friends during an aperitif)

```
spk1: ben ça dépend en fait c' que tu en fais hein
spk1: tu peux faire de l'enseignement si tu veux
spk1: tu peux faire euh
spk1: 'fin moi par exemple j' vais plutôt essayer de rentrer en communication
      médiation
spk3: quoi dans quoi
```

**spk1** well that depends in fact on what you do of it eh

**spk1** you can teach if you want to

**spk1** you can do hum

**spk1** well me for example I will be more like trying to enter in communication mediation

**spk3** what in what

Example 52 has been adapted in format to make it readable by humans. Transcriptions in CLAPI come in several formats: .trs,.ca,.cha,.doc,.tei,.rtf, figure 3.2 shows an excerpt from a .trs transcription. Transcriptions are made with several layers of meta-data and annotation, see figure 3.2. The transcriptions contain explicit transcription of overlaps, pauses, onomatopoeia, and laughter. The transcriptions do not contain explicit annotations of questions but do contain explicit annotations of prosody.

```
2891 <Turn speaker="spk3" startTime="1515.082" endTime="1518.249">
2892   <Sync time="1515.082"/>quoi dans quoi
2893   <Event type="pronounce" desc="/" extent="previous"/> `fin
2894   <Event type="lexical" desc=":" extent="previous"/>
2895   <Event type="lexical" desc=":" extent="previous"/> communication médiation euh
2896   <Event type="pronounce" desc="/" extent="previous"/>
2897 </Turn>
```

Figure 3.2: View of the .trs file corresponding to spk3's speech turn, example 52.

Example 52 features two speakers, spk1 and spk3. spk1 talks for 4 speech turns, the time code and pronunciation of each is transcribed. Figure 3.2 shows the time code for spk3's speech turn (<Sync time="1515.082"/>), its prosody and its synchronisation with spk1's turns through overlapping excerpts: 'fin, communication médiation euh.

<sup>53</sup>Here: [http://clapi.icar.cnrs.fr/V3\\_Telecharger.php?interface\\_langue=EN](http://clapi.icar.cnrs.fr/V3_Telecharger.php?interface_langue=EN), visited on 04/05/2021.

**C-ORAL-ROM** C-ORAL-ROM is a multilingual resource composed of 4 comparable corpora: spontaneous productions in French, Italian, Portuguese and Spanish. This corpus is the result of a European project coordinated by the University of Florence.

**Example 53 (C-ORAL-ROM, ffamcv02)**

*Discussion about the Sunset, a cafe-nightclub.*

SOP: et tu et c' était quel soir de semaine

ANT: lundi je pense

SOP: hein parce que en plus ça change de trucs tous les soirs donc tu sais jamais quelle sur quelle soirée tu vas tomber en fait

ANT: ouais je j' ai cru comprendre ouais

SOP *and you and what weeknight was it*

ANT *Monday I think*

SOP *huh because it also changes things every night so you never know which party you're landing in in fact*

ANT *yeah I that's what I figured yeah*

Transcriptions from the French part of C-ORAL-ROM are not freely accessible through C-ORAL-ROM's project website, but one can find them as a part of CEFC. There, the transcriptions, in a human-readable format, are aligned with the oral recordings. Example 53 is an excerpt from a three-participants multilogue, only two of them (SOP and ANT) appear in the example. The transcriptions do not explicitly show the time codes, overlaps, and pauses. The transcriptions contain onomatopoeia and hesitations, but no annotation of questions whatsoever.

**CRFP** The reference corpus of spoken French (Corpus de référence du français parlé) is composed of recordings made in 38 different cities, divided according to three different speech situations (private, professional, public) and speaker information (level of studies, age, gender).

**Example 54 (CRFP, PRI-AUX-1)**

*A retired speaker talks about his professional life (teacher).*

L1: oh bé c' est cet cet cet auteur de romans policiers Conan Doyle une traduction de Conan Doyle

L2: ah oui d' accord

L1: euh comment ça s' appelle ce titre-là enfin

L2: peu importe oui oui oui oui

L1 *oh well that's that that that author of detective novels Conan Doyle a translation of Conan Doyle*

L2 *oh yes OK*

L1 *hum what's it called that title there finally*

L2 *never mind yes yes yes yes*





**TUFS** The French part Tokyo University of Foreign Studies corpus is composed of student interviews and was constituted in French universities of Aix-Marseille, Paris, and Bordeaux. The complete corpus features parts in Canadian (with bilingual speakers), Spanish, French, Japanese, Malaysian, and Turkish, both native speakers and language learners. The parts accessible on the TUFS website correspond to 4 recordings from Paris<sup>54</sup> and 4 recordings from Aix-Marseille university<sup>55</sup>. The CEFC gives access to 63 transcriptions.

**Example 56 (TUFS, conversation about differences between China and France)**

*The translation convention is different in this example as we use the translation and presentation given in the TUFS corpus itself.*

French: c'est à dire ici les gens c'est-à-dire ici les gens sont  
vraiment très très liés à leur écriture par rapport à la Chine  
English: In other words, here, people are really very, very attached to their  
writing. Compared to China.

French: tu comprends  
English: You understand?

French: euh non  
English: Well no

French: pourquoi enfin  
English: Why...

Transcriptions on the TUFS website come with Japanese and English translations. The translations are punctuated while original transcriptions aren't. The presentation of the corpus given by Hisae Akihiro and Yuji Kawaguchi<sup>56</sup> offers additional information: transcriptions done before and after 2010 are different in several points. They feature transcription of short and long pauses (+/# and ++/## or ###, respectively before 2010/after 2010). Before 2010, no transcription of overlaps, after 2010 they were transcribed with < and >. No explicit transcription of questions before 2010, explicit transcription after.

**Rhapsodie** Rhapsodie is a treebank composed of 57 oral excerpts of spoken French aligned with an orthographic and phonetic transcription. It also features syntactic and prosodic annotations. Out of the 57 excerpts, 25 are original additions while the 32 others are extracted from previous projects such as CFPP2000 and ESLO. Rhapsodie features both monologues and dialogues. The dialogue part is constituted of 27 excerpts for a total length of a little over 2 hours (7,214 seconds), 24,804 words.

<sup>54</sup>[http://www.coelang.tufs.ac.jp/multilingual\\_corpus/fr2/index.html?contents\\_xml=corpus&menulang=en](http://www.coelang.tufs.ac.jp/multilingual_corpus/fr2/index.html?contents_xml=corpus&menulang=en) visited on 23/05/2021.

<sup>55</sup>[http://www.coelang.tufs.ac.jp/multilingual\\_corpus/fr/index.html?contents\\_xml=corpus&menulang=en](http://www.coelang.tufs.ac.jp/multilingual_corpus/fr/index.html?contents_xml=corpus&menulang=en) visited on 23/05/2021.

<sup>56</sup><https://docplayer.fr/9022270-Presentation-du-corpus-oral-en-francais-de-tufs-et-son-application-pour-1-a.html> visited on 23/05/2021.



**Example 57 (Rhapsodie, D0001)**

\$L1 est -ce que vous pourriez décrire euh les déplacements avec précision une journée

\$L2 eh ben XXX soit on \$-

\$L2 travaille

\$L1 par exemple -\$

\$L2 soit on travaille donc moi ben je vais je je prends le mét~ je prends le métro le matin bon jusqu' au \$-

L1 *could you describe hum the commute precisely [over/during] one day*

L2 *well XXX either we*

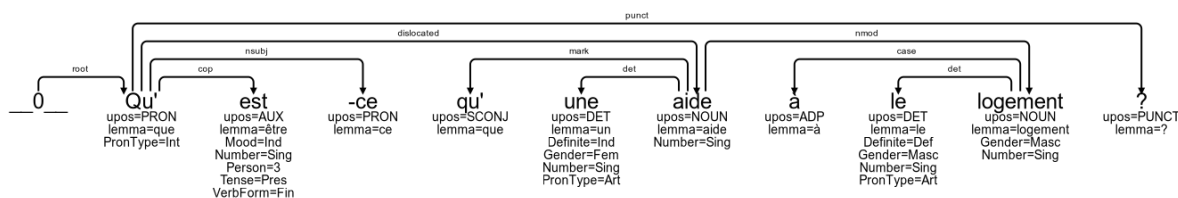
L2 *work*

L1 *for example*

L2 *either we work so me well I go I I take the sub- I take the subway in the morning well until*

Example 57 features explicit transcription of overlaps, onomatopoeia, and disfluencies. Overlaps are materialized through the use of \$ signs coupled with hyphens: L2's second turn overlaps with L1's second turn, but L2's second turn is actually the same as the beginning of her third turn. Rhapsodie is composed of very short excerpts (lasting 5 minutes on average), its main characteristic is a high quality, very detailed, annotation. In particular, Rhapsodie is fully annotated in prosody.

**FQB** The French QuestionBank is a treebank composed of questions in French. Though the FQB does not contain dialogues, we include this corpus here because of our focus on questions in French. Questions in the FQB come from several different sources: translations from English to French of existing test sets for question answering, frequently asked questions section of various websites (official French organizations and cooking forums), see [Seddah and Candito, 2016] for details.

**Example 58 (FQB\_CAF-1, « Qu'est ce qu'une aide au logement ? »)**

“What is housing support?”

Figure 3.4: FQB, « Qu'est ce qu'une aide au logement ? ».

Figure 3.4 shows a Universal Dependency parse of the question « Qu'est ce qu'une aide au logement ? » (“What is housing support?”), coming from the frequently asked questions section of the Social Welfare (CAF) website. The FQB is composed of 2,600 questions.

**ESLO** ESLO gathers transcriptions of recordings of sociological surveys conducted in Orléans, 40 years apart. ESLO1 was recorded between 1969 and 1974 for a total of more than 300 hours. It is constituted of both interviews (more than 182 hours) and free recordings in the city. ESLO2 started in 2008 and is planned to amount to more than 400 hours of recordings.

**Example 59 (ESLO1, ESL01\_ENT\_001\_C)**

```
spk1: oui bien sûr
spk2: n'est ce pas ?
spk2: oui oui
```

**spk1** *yes of course*

**spk2** *isn't it?*

**spk2** *yes yes*

Example 59 is an adapted version of an ESLO1 transcription in .trs format, shown in figure 3.5. Transcriptions in ESLO1 come with several layers of meta-data, closer to TCOF than CLAPI. They contain explicit indications of timecodes and overlaps. Line 207 in figure 3.5 is an example of an overlap transcription: spk1 and spk2 are talking at the same time (start-Time="41.518" endTime="42.326") and lines 210 and 213 give us the respective utterances. Transcriptions in ESLO1 also contain explicit annotation of questions, prosody, laughter, and onomatopoeia.

```
207 ▾ <Turn speaker="spk1 spk2" startTime="41.518" endTime="42.326">
208 <Sync time="41.518"/>
209 <Who nb="1"/>
210 oui bien sûr
211
212 <Who nb="2"/>
213 n'est-ce pas ?
214
215 </Turn>
216 ▾ <Turn speaker="spk2" startTime="42.326" endTime="42.686">
217 <Sync time="42.326"/>
218 oui oui
219
220 </Turn>
```

Figure 3.5: View of the .trs file corresponding to example 59.

**Example 60 (ESLO2, ESL02\_REPAS\_1247\_C)**

```
spk1: tu voudrais une fourchette et une ?
spk2: oui
spk1: des petites cuillères peut-être ?
spk2: non
```

**spk1** *you would like a fork and a ?*

**spk2** *yes*

**spk1** *some teaspoons maybe?*

**spk2** *no*

Example 60 is an adapted version of an ESLO2 transcription in .trs format, shown in figure 3.6. Transcriptions in ESLO2 contain at least the same information as in ESLO1.

```

124 ▾          <Turn speaker="spk1" startTime="78.945" endTime="80.352">
125          <Sync time="78.945"/>
126 tu voudrais une fourchette et une ?
127
128          </Turn>
129 ▾          <Turn speaker="spk2" startTime="80.352" endTime="82.374">
130          <Sync time="80.352"/>
131 oui
132
133          <Sync time="80.879"/>
134          </Turn>
135 ▾          <Turn speaker="spk1" startTime="82.374" endTime="84.410">
136          <Sync time="82.374"/>
137 des petites cuillères peut-être ?
138
139          </Turn>
140 ▾          <Turn speaker="spk2" startTime="84.410" endTime="86.256">
141          <Sync time="84.410"/>
142 non

```

Figure 3.6: View of the .trs file corresponding to example 60.

**SLAM** The Schizophrenia and Language, Analysis and Modeling (SLAM) [Amblard et al., 2015, Rebuschi et al., 2013] corpus is composed of 80 long semi-directed interviews between patients with schizophrenia under medication, patients with schizophrenia without medication, or control patients (without any known diagnosis) and psychologists. This corpus is not freely available as it contains medical and sensitive data [Boritchev and Amblard, 2021]. Anonymized parts of the corpus can be found throughout publications and can be used to conduct further studies.

### Example 61 (SLAM)

**Ps:** du coup tu faisais quoi avant ?

**Th:** j'ai fait une licence d'arts plastiques et euh...

**Ps:** ok

**Ps** *so what were you doing before?*

**Th** *I was studying for a bachelor in visual arts and hum...*

**Ps** *ok*

Transcriptions from SLAM contain no timecodes. Overlaps are materialized, but not in a precise way. As shown in example 61, questions are explicitly transcribed, along with onomatopoeia and hesitations. A particular focus in the transcription of SLAM was made on disfluencies.

The presentation of SLAM concludes our overview of the major English and French corpora that have at least one of the characteristics that we were looking for: corpora of transcribed oral dialogues and/or with an explicit marking of interrogatives.

One of the main difficulties we encountered while compiling this overview was that a large part of the considered corpora was not fully and/or easily accessible. Data privacy issues are of highest importance and so is the possibility for researchers to access to complete datasets, if only to produce clean and reproducible research. We wanted to produce a new dataset of transcribed oral dialogues in French, yet do so in a way that would respect the privacy of participants and also allow us to share as much data as possible as widely as possible. The STAC corpus then gave us the idea to use a board game as an excuse and an activity for participants to engage in conversations while revealing as little personal details as possible. As STAC already used *Catan*, we decided to use the same game, at least to begin with, to develop DinG, in view of future comparison works.

### 3.3 Dialogues in Games (DinG)

Dialogues in Games (DinG) is a corpus of manual transcriptions of real-life, oral, spontaneous multi-party dialogues between French-speaking players of *Catan*<sup>57</sup>. *Catan*, or *Settlers of Catan*, is a board game for three to four players in which the main goal for each participant is to make their settlement prosper and grow, using resources that are scarce. Bargaining over these resources is a major part of the gameplay and constitutes the core of DinG’s data.

#### Example 62 (A dialogue from DinG)

**Yellow**<sub>1</sub> *Tu veux bien me donner un mouton?*

**Yellow**<sub>1</sub> *“Would you like to give me a sheep?”*

**Blue**<sub>2</sub> *Je veux bien euh un blé*

**Blue**<sub>2</sub> *“I would like uh a wheat”*

**Yellow**<sub>3</sub> *Et tu me donnes un mouton?*

**Yellow**<sub>3</sub> *“And you give me a sheep?”*

**Blue**<sub>4</sub> *Et je te donne un mouton*

**Blue**<sub>4</sub> *“And I give you a sheep”*

**Yellow** and **Blue** are players of *Catan*, designated by the colour of their game tokens. In example 62, **Yellow** and **Blue** are bargaining over resources: **Yellow** seeks a sheep; **Blue** offers to give them one, but only in exchange for wheat. **Yellow** secures the bargain in speech turn **Yellow**<sub>3</sub>; **Blue** confirms.

Dialogues from DinG are unconstrained, as the players don’t have to follow any rule or specific guideline, apart from playing the game. As bargaining over the resources is part of the gameplay, the players have to speak in order to play, so the dialogues are the ones naturally occurring in this particular setting. As the players have to speak to play, they do not discuss

<sup>57</sup>Copyright ©2017 CATAN Studio, Inc. and CATAN GmbH. All rights reserved.

personal subjects outside the game setting, which makes it possible to completely anonymize the corpus by removing the players' names, see below for a longer discussion.

An average game of *Catan* lasts at least 30 minutes, thus DinG contains long interactions, going beyond informative exchanges. The corpus was designed to study human-human dialogue based on attested, spontaneous, and unconstrained oral data in French. We want to study different types of dialogue-encountered phenomena; on one hand, the mechanisms underlying the combination of dialogue turns, in order to produce a computational model of dialogue, on the other hand, the dynamics of interactions, in order to account for meta-levels of human-human interactions.

The proximity we share with the STAC corpus appears to us as a great opportunity. Indeed, modeling natural language at discourse level is a task for which annotated resources are scarce. The Parallel Meaning Bank<sup>58</sup> constitutes a notable exception and is annotated in DRT<sup>59</sup>. The STAC corpus is annotated in SDRT<sup>60</sup>, which means that the proximity our two corpora share is an opportunity to progress on comparisons between written and oral discourse on the same topics while opening on dialogue.

Developing a corpus of spoken French (DinG) that is close topic-wise and vocabulary-wise to an informal written chat English one (STAC) allows us to study fine-grained modeling phenomena, as much regarding the correspondence between the two languages as for the logical models that we use.

We envision our corpus as a corpus of spontaneous dialogues in French with a quality transcription. Its nature allows for large dissemination and high cross-domain reusability. Its length allows for a study from different perspectives, such as shown above. Furthermore, we want to present it as a corpus that follows the good practices of corpus collection, usable for collection of other corpora, especially ones that cannot be disseminated largely because of the nature of the data they contain (medical-condition related for example). It is our case, as our team develops corpora such as Schizophrenia and Language, Analysis and Modeling<sup>61</sup> (SLAM) and the corpus of the Computerized tools to assist in the Diagnosis of Mental Illnesses project<sup>62</sup> (Outils informatisés d'aide au Diagnostic des Maladies mentales, ODiM) that contain sensitive data and cannot be largely published. The ODiM corpus follows the same transcription process as DinG. For the sake of science reproducibility, we developed DinG as a free-to-share corpus to showcase our studies, our tools, and our formal models.

Recording during a game of *Catan* allows us to capture long spontaneous interactions that (almost) don't contain personal data. [Amblard et al., 2014, Grouin et al., 2015] show that long interactions contain enough information to reduce the identification process to a (very) small amount of people. As the production of transcriptions of oral data is a very costly process, we want our corpus to be as widely sharable as possible. Therefore, following the recommendations of [Leidner and Plachouras, 2017], we decided to take care of ethical aspects by thinking them through the entire process of data collection and publication. The process was developed under the supervision and the validation of the Operational Committee for the Evaluation of Legal and Ethical Risks<sup>63</sup> (OCELER) of the INRIA, in due respect of the General Data Protection

---

<sup>58</sup><https://pmb.let.rug.nl/>, visited on 16/04/2021.

<sup>59</sup>DRT is a classical formalism for discourse representation, for more on DRT see Chapter 1, section 1.3.1.

<sup>60</sup>SDRT is a detailed formalism for relations in discourse, for more on SDRT see Chapter 1, section 1.3.1, and Chapter 2, section 2.2.4.

<sup>61</sup><https://team.inria.fr/semagramme/fr/slam/>, see section 3.2.2 for a complete presentation.

<sup>62</sup><https://team.inria.fr/semagramme/odim/>, more about this project below.

<sup>63</sup><https://www.inria.fr/en/operational-committee-assesment-legal-and-ethical-risks>, visited on 21/04/2021

Regulation<sup>64</sup> (GDPR). For now, we remove all mention of the names of the participants in the transcriptions and we publish transcriptions only. In future work, we are considering anonymizing the oral data, following approaches such as [Qian et al., 2017]. If and when we manage to do so, we will contact the participants again as they will have to sign a new consent sheet for their data to be published.

As DinG is composed of long human-human interactions, it can be very largely used for dialogue studies in numerous fields. In particular, dialogues from DinG can be a starting point for medical research groups: studying data produced by people without any known diagnosis can help identify conversational disorders when they occur. Conversely, it can help not to tag a phenomenon as a conversational disorder by acknowledging the fact that this phenomenon occurs in a non-pathological setting as well. In fact, data from DinG is also used in the Sémagramme team (LORIA, Inria Nancy Grand-Est) to help develop and showcase SLODiM<sup>65</sup>, a tool for dialogue analysis in a medical (psychiatric) setting.

### 3.3.1 A game of *Catan*

A typical game of *Catan* takes place between 3 to 4 players and lasts between 30 and 90 minutes. The board is built using 19 hexagon terrain tiles of different colors: bright green pastures, yellow fields, grey mountains, brown carriers, dark green forests, and a desert tile. These tiles constitute the island of Catan, surrounded by water. See figure 3.7 for an example.



Figure 3.7: A photo of the game board during a game of *Catan*.

Most of the players we recorded for DinG never played *Catan* before. Therefore, each recording was preceded by a phase of rules explanation conducted by an observer. After this phase, the rule book was handed to the participants for them to be able to play autonomously. The players use game pieces of different colors: red, white, yellow, blue. In the following, we identify the players with the color of their game pieces and we call them by the first letter of the name

<sup>64</sup><https://gdpr-info.eu/>, visited on 15/08/21.

<sup>65</sup><https://slodim.loria.fr/>, visited on 11/03/2021.



of the color: **R** for red, **W** for white, **Y** for yellow, **B** for blue. As the recordings of DinG took place in a noisy environment (see section 3.3.3 for details), we used Other (**O**) as an identifier for all oral data that was not produced by one of the players.

### 3.3.2 Corpus description

The corpus is composed of 10 recordings of games that last 70 minutes on average. The shortest recording is almost 40 minutes long (DinG8), the longest lasts a little over 1h44m (DinG1). Most of the recordings (all but n°4, 5, and 6) were split into two parts because of a food break occurring during the game, as the recordings took place during university game nights. This division was kept in the transcription process, as it was easier for transcribers to work with shorter pieces of audio. The data presented in table 3.5 is computed on the merged recordings, one per game. The complete data per partial recording and per participant can be found in tables A.1, A.2, A.3, in the Annexes. The last line of table 3.5 shows the coefficients of variation (*CV*), a standardized measure of dispersion of a dataset, which is computed as the ratio of the standard deviation  $\sigma$  to the mean  $\mu$ . A *CV* < 100% corresponds to a dataset with low variance.

Name	Length (min)	Length (turns)	# questions	# turns /minute	# questions /minute	% questions among turns
DinG1	<b>104.33</b>	<b>3,572</b>	<b>506</b>	34.24	4.85	<b>14.17</b>
DinG2	86.31	2,969	290	34.40	3.36	9.77
DinG3	53.7	1,716	126	31.96	2.35	<b>7.34</b>
DinG4	75.93	2,985	333	39.31	4.39	11.16
DinG5	78.41	3,012	362	38.41	4.62	12.02
DinG6	84.02	3,130	265	37.25	3.15	8.47
DinG7	96.34	3,293	340	34.18	3.53	10.32
DinG8	<b>39.92</b>	1,627	196	<b>40.76</b>	<b>4.91</b>	12.05
DinG9	41.71	<b>795</b>	<b>69</b>	<b>19.06</b>	<b>1.65</b>	8.68
DinG10	41.13	<b>476</b>	<b>41</b>	<b>11.57</b>	<b>1.00</b>	8.61
Global data	701.8	23,575	2,528	33.59	3.60	10.72
CV	34%	47%	57%	29%	40%	20%

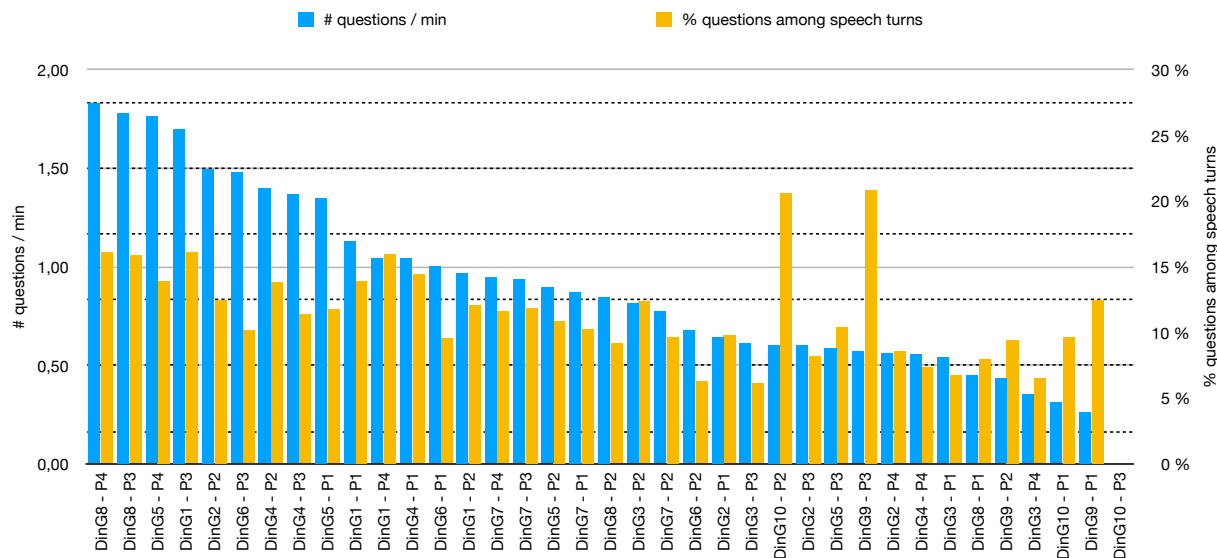
In boldface, the data discussed in the following paragraphs.

Table 3.5: DinG data – observations per game, average on whole corpus and coefficients of variation (*CV*).

DinG1 is the longest both with respect to time and amount of speech turns; it also contains the biggest amount of questions. While DinG9 and 10 are not the shortest in terms of time, their amount of speech turns and questions are significantly (more than 10%) smaller than DinG8’s (shortest in terms of time). This observation is supported by the fact that DinG9 and 10 present the smallest amount of speech turns per minute, while DinG8 presents the greatest: DinG8 lasts less time but DinG8’s players talked at least twice more than DinG9 and DinG10’s ones. Similarly, DinG8 presents the highest amount of questions per minute while DinG9 and DinG10 show the smallest ones.

The focus returns on DinG1 when we look at the percentage of questions among all the speech turns, as this game presents the highest percentage. The smallest percentage is shown by DinG3; more about this below, in the discussion of figure 3.8. DinG is homogeneous in terms of all the measures we considered in table 3.5, as all the coefficients of variation stay under 60%.

While the amount of questions (identified as the utterances marked with a ‘?’) varies quite a lot from one recording to another, the percentage of questions among turns stays very similar (under 30%); see the discussions of figure 3.8.



Amount of questions per minute compared with the percentage of questions per speech turn, for each participant (P) of the study, sorted by descending number of questions per minute.

Figure 3.8: Quantitative data, DinG.

We computed, for each participant (**O** excluded), the number of questions per minute of game along with the percentage of questions among all the speech turns. Figure 3.8 shows a visualization of the comparison between the two sets of values. We observe that while the difference in behavior between the two values is small for most of the participants (P), P2 of DinG10, P3 of DinG9, P2 of DinG9, P1 of DinG10, and P2 of DinG9 show a significant difference. For these five participants, the percentage of questions per speech turn is relatively high (with respect to the values of the other participants), while the amount of questions per minute of game stays around 0.5. One of the possible explanations for this is that overall, the participants spoke very little during DinG9 and DinG10 compared to the other games, see the discussion above. Indeed, the highlighted participants were all part of these two games, and the reason P3 of DinG10 doesn’t appear in this list is that P3 asked 0 questions during the whole game.

### 3.3.3 Collection process

The recording part of the corpus collection took place during university game nights. All participants are native French speakers. 33 people participated in the recording process, 12 women and 21 men. All participants but 3 had a master’s degree or higher. Each participant only appears once in the corpus. We collected as little personal data as possible, but we can say that the average age of the participants is around 25 years old, and all the participants are native French speakers. All the participants signed an informed consent sheet, acknowledging they were giving us the right to record personal data (their voices) and share transcriptions of it. It was important to us to stress the fact that their consent was retractable at any point in the process.

As we wanted the participants to feel as relaxed and natural as possible, the recordings were conducted in the room where the rest of the game night took place. Recording during the game



nights raised some technical challenges, in particular, because different people were playing different games in the same physical space. Yet, it allowed us to record in a way that made the participants very comfortable: most of them report afterward that they forgot the recording devices after the first fifteen minutes of playing. All recordings were conducted by a non-player observer, whose duties were to explain the experiment, find volunteers and supervise the smooth running of the process. In particular, the observer had to manage the microphone and monitor the level of surrounding noise. We needed to address the technical challenge of recording our participants in a clear enough way for transcription, without recording other people's conversations. In order to do so, we used *H2 next handy recorder* by ZOOM<sup>66</sup>, in XY (90° recording mode) setting. See figure 3.9 for a schematic view of the recording setting. Part of the observer's duty was to manage the surrounding noises to ensure a good quality of the recording.

### 3.3.4 Data processing

This section describes the data processing that was necessary in order to produce DinG. Once a game is recorded, the raw audio file is given to transcribers<sup>67</sup>. As our corpus was recorded in a noisy environment, transcribers had to pre-process the audio signal before starting to properly work on it. The preprocessing is done using Audacity<sup>68</sup> and aims to reduce the peaks in the sound signal (corresponding to loud noises such as rolling the dices) in order to then be able to uniformly amplify the whole signal and make the voices clearer. The following treatments were applied:

**High-pass filter** Cuts the lowest frequencies. A lot of peaks come from thumps such as the ones produced by a foot banging on a table leg, manipulations of the microphone, *etc.* Applying a relatively low high-pass filter ( $\sim 200Hz$ ) clears the signal from the biggest peaks and sometimes alleviates surrounding noises such as music or electronic devices sounds, that can cover the participants' voices. Using a high-pass filter with a setting that is too high would result in suppressing the lowest voices and sounds such as expirations, « hum ».

**Click removal** Cuts the very sharp and short peaks such as the ones coming from tapping, clicking of a pen, shuffling of cards, dice sounds. It doesn't always work, particularly if the sounds are too powerful or coupled with low-frequency sounds. This last case is why using a high-pass filter before click removal is important.

**Limiter** After the first two steps, it is possible to proceed to a general flattening of the signal. Audacity allows to set the limiter between 1 and 10dB: once the limiter is set, it cuts everything above the selected amplitude. One should not overdo it, as cutting too much might result in a deterioration of the signal, creating clicks. As the data we're interested in for DinG is mostly conversation, it is possible to apply huge limitations of 8 to 10dB.

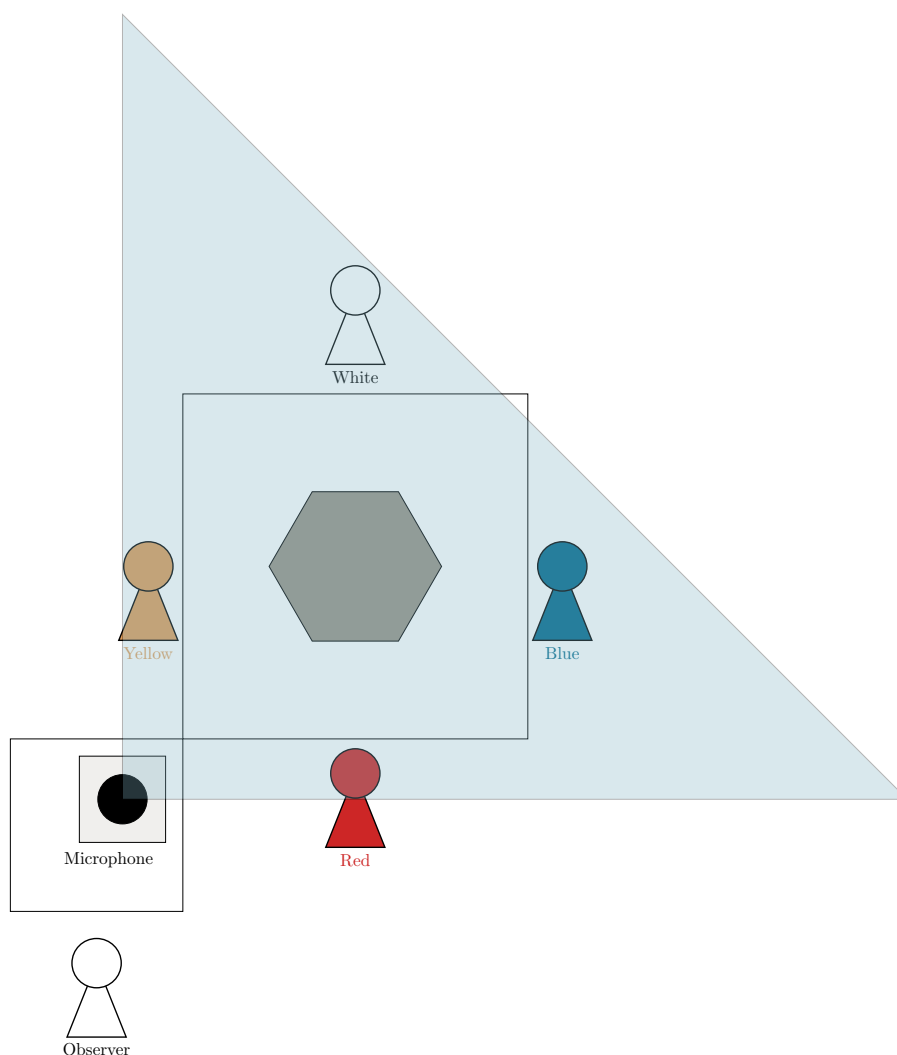
**Distortion** An alternative to the use of a limiter: if the output of the limiter is not giving a satisfying result, it is possible to use a distortion, with the pre-setting cubic curve (odd harmonics). This will change the sound a little but this is not a problem for DinG's data, as we are mostly interested in the content of what is being said, as opposed to voice pitch for example.

---

<sup>66</sup><https://zoomcorp.com/en/us/handheld-recorders/h2n-handly-recorder/>, visited on 09/03/21.

<sup>67</sup>We thank immensely Esteban Marquer, Julien Botzanowski, Laurine Jeannot, Léa Dieudonat, Srilakshmi Balard, and Lucille Dumont for their great work.

<sup>68</sup><https://www.audacityteam.org/>, visited on 09/03/21.



The microphone is standing on an isolating compact polystyrene platform on a table adjacent to (but different from) the table on which the game takes place, in order to reduce as much as possible sound effects induced by vibration from the players moving around, rolling the dices. The microphone's recording area (the light blue triangle on the figure) is oriented towards the players.

Figure 3.9: Recording process – schematic view.

**Amplification** After the previous manipulations give us a cleaned signal, it can be amplified. The default setting of Audacity amplifies the signal up to right before the saturation point. If the values get negative, that means that the signal is already saturating and has to be lowered to reach  $0dB$ .

These treatments diminish the background noise while sharpening the voices of the participants. The next step is manual transcription. We wrote a transcription guide, adapted from [Blanche-Benveniste and Jeanjean, 1987]. This guide ensures that the different parts (segmentation and transcription) are homogeneously realized by different transcribers.

Before choosing a transcription tool, we conducted a comparative study based on ergonomics, quality, and general characteristics of different specialized software. We also took into considera-

tion the free availability of the software and their codes. We tested EXPRESSSCRIBE<sup>69</sup>, ASTALI<sup>70</sup>, YOUTUBE<sup>71</sup>, TRANSCRIBERJS<sup>72</sup>, OTRANSCRIBE<sup>73</sup> and ELAN<sup>74</sup>. These software were evaluated on each of the following aspects: sound manipulation, navigation inside the recording, supported formats, text manipulation, transcription tools, speaker annotation, dialogue act annotation, management of overlaps, noises, and inaudible fragments. The tool that got the best evaluation was ELAN [Wittenburg et al., 2006] because of its overall ergonomic design (for segmentation and transcription) in one tool while giving access to a visualization of the sound signal. Thanks to ELAN, DinG contains time-code alignment and disambiguation of speakers' overlaps. ELAN is a multimodal annotation tool for audio and video. Among the different possibilities that ELAN offers, we used mainly segmentation and transcription to produce the final version of DinG.

6 transcribers took part in the project. 5 of them were recruited among natural language processing students, one is an expert in production and synchronization of subtitles. They were trained for the task on a 5 minutes excerpt from DinG, that they all annotated and got to discuss with us and, when possible, between them. The transcribers were paid for the task. We counted 30 hours of work for the transcription of a recording of 1.5 hours. The transcribers followed different strategies to complete their work, from minute-by-minute segmentation & transcription in parallel to the full segmentation and then full transcription, speaker by speaker, of the whole recording. Figure 3.10 shows the complete data collection and processing pipeline for DinG.

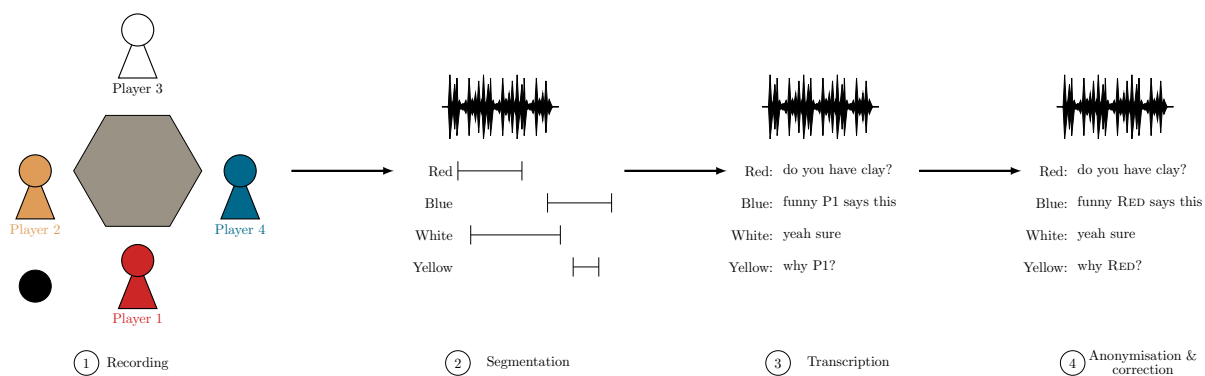


Figure 3.10: Data collection and processing pipeline.

**Manual segmentation in speech turns** The first step in the process of getting the transcription is a manual segmentation of the recording in speech turns, called *segments* in ELAN. We define a segment as a speech turn that is composed of a pseudo-sentence, an onomatopoeia, a noise, or any combination of the above. A speech turn is a theoretical linguistic unit corresponding to the verbal production of a speaker, see for example [Schegloff et al., 1974]. Using ELAN allows us to process overlaps in a very simple and visual way, see figure 3.11 for an example. Thus, each segment constitutes a coherent linguistic unit.

<sup>69</sup><https://www.nch.com.au/scribe/index.html>, visited on 09/03/21.

<sup>70</sup><http://ortolang108.inist.fr/astali/fr/>, visited on 09/03/21.

<sup>71</sup><https://support.google.com/youtube/answer/2734796?hl=en>, visited on 09/03/21.

<sup>72</sup><https://ct3.ortolang.fr/trjs/>, visited on 09/03/21.

<sup>73</sup><https://otranscribe.com/>, visited on 31/03/21.

<sup>74</sup><https://archive.mpi.nl/tla/elan>, visited on 09/03/21.

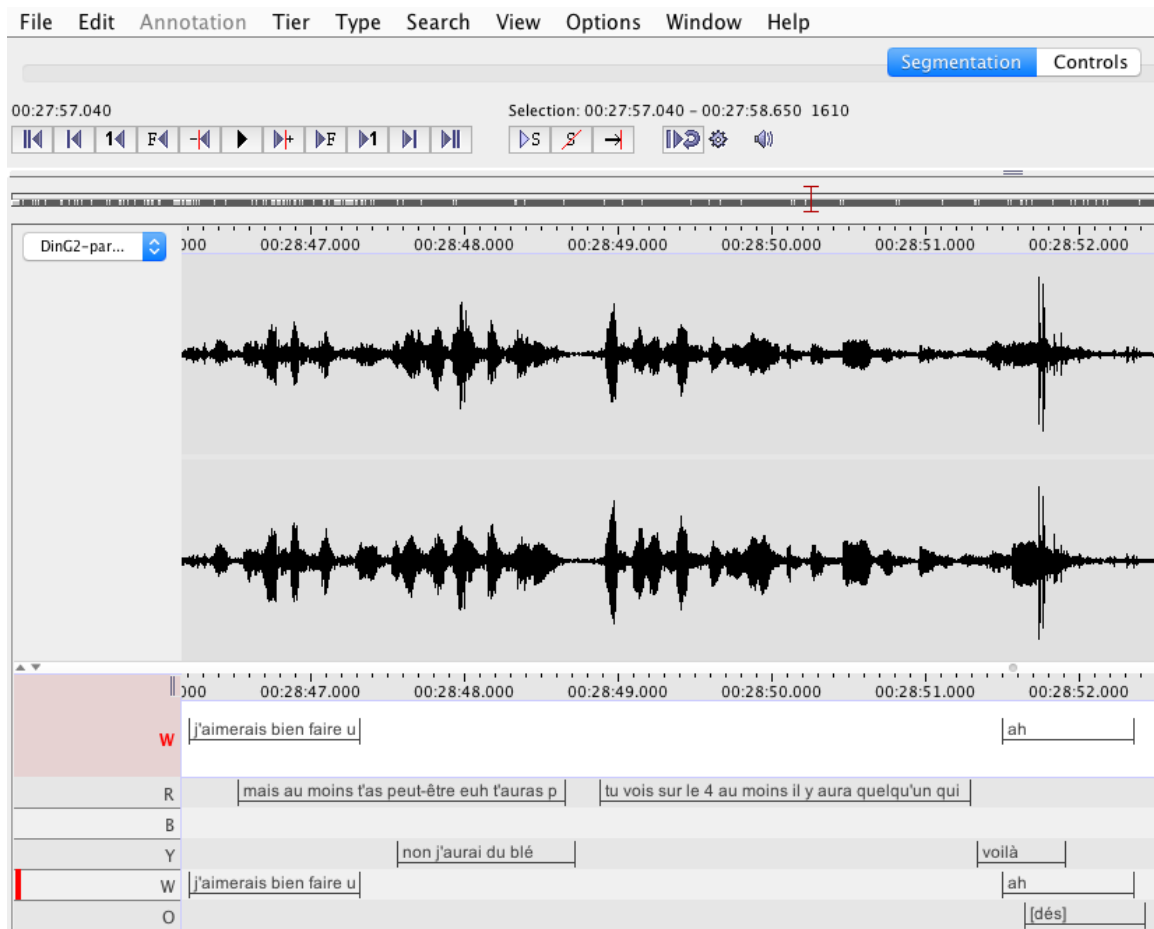


Figure 3.11: View of ELAN 5.6-FX in segmentation mode, working on DinG2.

**Transcription** The transcription guide sets the norms to follow. The guide is inspired by [Blanche-Benveniste and Jeanjean, 1987]. The main modifications are adaptations to the subject of our observation and the object of our research: (1) we specified the noise tags in order to adapt them to the board game context by adding tags such as [dice], [tokens]; (2) we added an explicit transcription of interrogative marks in order to account for utterances that were perceived (by the transcribers) as questions (rising intonation, answers given in the following dialogue turns). The transcription guide will be made available online. Furthermore, we produced a transcription and a segmentation, to preserve the dynamic aspect of interaction, for example by explicitly visualizing overlaps. Several automatic annotations were produced using SLODiM<sup>75</sup>, in particular for disfluencies and syntax. Transcribers are asked to respect scrupulously what is recorded/heard/said – they are not supposed to correct the language but to produce a faithful written version of French as it is used by the speakers. The writing of onomatopoeias is normalized via lists (« euh », « hum », *etc.*). Pauses are explicitly marked with their approximate duration (ex: (0.2s)). Dysfluencies are also kept, in particular repetitions and beginning of words, that are marked with a hyphen (ex: « ca-(interrup-) carte » // “ca-(interrup-) card”). The transcription does not contain any punctuation except the interrogation point, which is used to annotate rising intonations that correspond to questions in the recording’s oral context. As we

<sup>75</sup><https://academia.slodim.fr/>

are interested in questions and answers in dialogue, having an explicit annotation of questions is particularly useful for us. Figure 3.12 shows the transcription interface of ELAN.

The transcribers who participated in the project have all received training on the same 5 minutes excerpt. Everyone did an individual segmentation and transcription before pooling and comparing the results. The inter-annotator agreement for transcriptions is calculated on the transcription of a 5 minutes excerpt of DinG2, pre-segmented. Two independent annotators<sup>76</sup> (not working on the project before) have received empty segments for the excerpt and filled them with transcriptions, following the transcription guide. First, we computed the agreements for the full transcriptions, see the first two lines of table 3.6. It is important to stress that inter-annotator agreement on transcriptions is always low, as the amount of possible transcriptions is very large; yet, even taking this into account, the results we got were very low (under 0.3). Then, we computed the agreements for the transcriptions from which we removed the noises and the pauses. This produced lines 3 and 4 of table 3.6, with results higher than 0.5, which is usually considered to be a good agreement for transcriptions. This difference leads us to the conclusion that the quality of the recordings might be insufficient to grant an objective transcription of noises, on one hand, and also that transcriptions of the duration of pauses can vary from one transcriber to another.

	$\kappa_{ipf}$	Raw agreement
With noise, including unlinked/ unmatched annotations	0.28	0.28
With noise, excluding unlinked/ unmatched annotations	0.28	0.28
Without noise, including unlinked/ unmatched annotations	0.52	0.55
Without noise, excluding unlinked/ unmatched annotations	0.53	0.55

$\kappa_{ipf}$  is a modified version of Cohen’s  $\kappa$ , computed using an *iterative proportional fitting* algorithm, it includes the unmatched annotations in the agreement calculation. The raw agreement is computed by dividing the number of agreeing cases by the total number of cases [Holle and Rein, 2015].

Table 3.6: Interrater agreement for transcription before/after noise tags and pauses removal, calculated with ELAN, following [Holle and Rein, 2013].

**Anonymization** It was of major importance for us to be able to distribute our resource while preserving the participants’ private data. The last step in the transcription process is anonymizing the transcription. Each of the players is identified with the colour of their game pieces: Red (**R**), White (**W**), Yellow (**Y**) or Blue (**B**). If a name is pronounced out loud, it is replaced in the transcription by the name of the corresponding color, in upper case. Outside noises and speakers

<sup>76</sup>We thank greatly Amandine Lecompte and Samuel Buchel for their contribution to our work.

No	Type 1 : default-lt
	W
891	j'aimerais bien faire une colonie moi
	R
892	mais au moins t'as peut-être euh t'auras potentiellement
	Y
893	non j'aurai du blé
	R
894	tu vois sur le 4 au moins il y aura quelqu'un qui aura quelque chose
	Y
895	voilà
	W
896	ah
	O
897	[dés]
	R
898	alors 7 c'est bon tu vas pouvoir enlever ça [rire]

Figure 3.12: View of ELAN 5.6-FX in transcription mode, working on DinG2.

are assigned to an outside speaker called Other (O).

**Super-annotation** Once the transcription is done, it is given to a super-annotator. The super-annotator proofreads all the transcriptions, checks the recording when verification is asked by the transcriber through the mark  $\wedge\wedge$ . The goal of super-annotation is to homogenize the corpus via the correction of typos (those that do not correspond to what is being said), standardization of the noise tags, of the transcription of onomatopoeias, of the writing of numbers<sup>77</sup> and of the anonymization. If the name of the player has been said out loud, it is replaced by the name of their color, in upper-case letters (ex: « c'est juste que ROUGE a fait 6 » // “it's just that RED has got a 6”). If the players directly use the name of the color instead of the player's name, the first letter of the transcription is capitalized (ex: « c'est juste que Rouge a fait 6 » // “it's just that Red has got a 6”). If the name of the color is used directly to reference the color of the game pieces, it is kept in lowercase letters (ex: « un blé pour le rouge » // “a wheat for the red one”). Table 3.7 shows the raise in transcriptions quality obtained after super annotation, computed on non-specialists' transcriptions: the first two lines of table 3.7 are the same as in table 3.6,

<sup>77</sup>In French, « un » corresponds both to “a” and to “one”. The guide asks for disambiguation, as much as possible. Following this idea, we transcribe « 1 », « 2 », etc. for amounts and « un » when the usage corresponds to the determiner.

the last two correspond to the results obtained by calculating the agreement on the proofread transcriptions.

	$\kappa_{ipf}$	Raw agreement
Original, including unlinked/ unmatched annotations	0.28	0.28
Original, excluding unlinked/ unmatched annotations	0.28	0.28
After super-annotation, including unlinked/unmatched annotations	0.35	0.35
After super-annotation, excluding unlinked/unmatched annotations	0.35	0.35

$\kappa_{ipf}$  is a modified version of Cohen’s  $\kappa$ , computed using an *iterative proportional fitting* algorithm, it includes the unmatched annotations in the agreement calculation. The raw agreement is computed by dividing the number of agreeing cases by the total number of cases [Holle and Rein, 2015].

Table 3.7: Interrater agreement for transcription before/after super-annotation, calculated with ELAN, following [Holle and Rein, 2013].

The corpus is available on Gitlab: <https://gitlab.inria.fr/semagramme/DinG>. It is distributed under the Attribution ShareAlike Creative Commons license (CC BY-SA 4.0). Each game is available as a numbered .txt file, exported from ELAN. We chose to export the transcriptions as **Traditional Transcript Text** in order to generate text files that would be both readable by human observers and could easily be manipulated by scripts.

The files correspond to linearised versions of the games. Each segment appears on one line, that starts by the number of this segment in the transcription, then the letter that identifies the speaker (**R**, **W**, **Y**, **B** or **O**), then the transcription of what has been said. The next line contains the time codes of the beginning and end of the segment. When two following speech turns do not overlap, the gap between the two is calculated automatically and written in brackets on the next line. These elements are shown in example 63, see in particular the gap between 009 and 010.

### Example 63 (Excerpt from DinG transcription, DinG6)

```
009 Y  j'aimerais bien faire 7 pour une fois
      00:00:14.438 - 00:00:15.880
      (0.64)

010 R  en fait t'as (te-) t'étais contente parce que juste tu as fait un double
      6 et qu'en général c'est cool dans les jeux [rire]
      00:00:16.518 - 00:00:21.910

011 Y  ouais c'est ça
      00:00:21.712 - 00:00:22.718

012 R  [rire]
      00:00:21.915 - 00:00:23.219
```

009 **Y** I would like to get a 7 for once

010 **R** in fact your have (y-) you were happy because simply you got a double 6 and generally it's cool in games [laugh]

011 **Y** yeah that's it

012 **R** [laugh]

Now that we have presented the complete constitution process for *DinG*, we want to put *DinG* in perspective of the other corpora showcased in section 3.2.

### 3.4 *DinG* in the corpora landscape

In the previous sections, we presented several corpora of English and French that we find to be of interest in the scope of our studies of dialogue in general and questions and answers in dialogue in particular. Table 3.8 summarizes the quantitative information we have about these corpora along with information about *DinG*.

Corpus	language	length (h)	# recordings	# EDU/ speech turns	# words	marking of '?'
SCoSE p1	English	~ 3	14	5,886	32,826	yes
STAC	English	N.A.	59	10,678	N.A. <sup>78</sup>	yes <sup>79</sup>
S-BNC	English	~ 987	1,251	1,248,110	11,500,000	yes
TUFS <sup>80</sup>	French	~ 1	4	205	1,829	yes
FQB <sup>81</sup>	French	N.A.	N.A.	<sup>82</sup> 2,289	23,901	yes
Rhapsodie <sup>83</sup>	French	~2	51	2,667	29,063	prosody
C-ORAL-ROM <sup>84</sup>	French	21	152	6,306	244,671	no
CRFP <sup>85</sup>	French	~ 33	124	22,638	375,483	no
<b>DinG</b>	<b>French</b>	<b>~12</b>	<b>10</b>	<b>23,575</b>	<b>121,010</b>	<b>yes</b>
TCOF <sup>86</sup>	French	26	112	25,333	352,762	no
SLAM	French	N.A. <sup>87</sup>	80	31,575	375,000	yes
CFPP2000 <sup>88</sup>	French	~ 48	43	49,865	618,013	yes
CLAPI <sup>89</sup>	French	117	324	123,771	1,057,574	prosody
ESLO <sup>90</sup>	French	305	443	310,686	3,686,571	yes
Average <sup>91</sup>	-	~ 203	~ 221	~ 141,539	~ 1,524,808	-

Where EDU stands for *elementary discourse unit*, the smallest semantic unit of segmentation for SDRT analysis (see Chapter 2, section 2.2.4 for a presentation). Note that the numbers given in table 3.8 are different from the values that can be found across various publications involving these corpora. We chose to present a larger amount of coherent numbers to get a bigger vision of the corpora landscape: in different publications where the corpora are mentioned, it is usual to find 2 or 3 of the measures presented in our table, but seldom all 4 metrics (length, amount of recordings, number of speech turns, amount of words) together. Thus, we computed most of the values in our table, using `bash` scripts, from the downloadable parts of the corpora<sup>92</sup>. Inside the table, we sorted the corpora, separately for English and for French, by ascending amount of EDUs/speech turns.

Table 3.8: Quantitative corpora description.



Table 3.9 summarizes the qualitative information we have about these corpora along with information about DinG. As pointed out by Farah Benamara, it is important to have an idea about the contents of the corpora in terms of contexts in which the interactions occur. In the context of our studies of questions and answers in dialogues, the topic/domain of a dialogue may have a huge influence on the distribution of question/answer pairs, along with the type of questions and answers that will be used.

Corpus	language	contents	type	marking of ‘?’
SCoSE p1	English	informal conversations	oral	yes
STAC	English	chat logs, <i>Catan</i>	written	yes
S-BNC	English	misc., British English	oral	yes
TUFS	French	students, misc.	oral	yes
FQB	French	question answering, FAQ	written	yes
Rhapsodie	French	misc.	oral	prosody
C-ORAL-ROM	French	misc.	oral	no
CRFP	French	private/professional/public	oral	no
<b>DinG</b>	<b>French</b>	<b><i>Catan</i></b>	oral	<b>yes</b>
TCOF	French	misc.	oral	no
SLAM	French	patients w/ schizophrenia, psychologists	oral	yes
CFPP2000	French	interviews about Paris	oral	yes
CLAPI	French	misc.	oral	prosody
ESLO	French	sociological surveys	oral	yes

Table 3.9: Qualitative corpora description.

In table 3.8, we observe that while DinG is small in terms of duration, it contains twice more speech turns than STAC. DinG contains 2,528 questions (table 3.5), which makes it comparable to the FQB (2,289 question), a corpus specifically designed to showcase questions in French. As DinG is composed of long interactions, even though the amount of recordings composing DinG is small (more than 20 times smaller than the average amount of recordings in the corpora considered in the table), the amount of DinG’s speech turns is only 6 times smaller than the average one. The same observation can be made for the number of words, 13 (and not 20) times smaller.

Table 3.10 presents a summary of the transcription characteristics for the oral corpora, similar

<sup>78</sup>The number of words in STAC corpus is not fully relevant as the corpus contains server messages, not produced by the human participants.

<sup>79</sup>STAC does contain an explicit marking of interrogatives, but the usage of “?” is different from the one found in transcriptions of oral interactions. Ex: “I have one clay but I want it for myself?”

<sup>80</sup>Data found in references: 54 hours, ~750,000 words.

<sup>81</sup>Data coming from Grew-match, [Guillaume, 2021], <http://match.grew.fr/>.

<sup>82</sup>For FQB, we counted the total amount of sentences constituting the corpus.

<sup>83</sup>Data found in references: 27 recordings, 24,804 words.

<sup>84</sup>Data found in references: 40 hours, 500,000 words

<sup>85</sup>Data found in references: 36 hours, 440,000 words

<sup>86</sup>Data found in references: 23 hours, ~300,000 words.

<sup>87</sup>Information about the length of the corpus is currently non available because of the medical restrictions surrounding this corpus. A reasonable approximation is to say that it is at least 40h long.

<sup>88</sup>Data found in references: ~ 60 hours, 783,017 words.

<sup>89</sup>Data sent by Carole Etienne, whom we thank immensely. The data found in references was ~ 600 hours.

<sup>90</sup>Data found in references: 700 hours.

<sup>91</sup>Without DinG.

<sup>92</sup>We thank immensely Pierre Lefebvre, whose engineering work is a major ingredient in the results presented in this chapter.

Corpus	Transcription of interrogatives	Pauses	Onomatopoeia	Laughter	Overlaps	Disfluencies (hesitations, interruptions)
S-BNC	✓	✓	✓	✓	✗	✓
SCoSE p1	✓	✓	✓	✓	✓	✓
CFPP2000	✓(fine-grained)	✓	✓	✓	✓	partly
CLAPI	prosody	✓	✓	✓	✓	partly
C-ORAL-ROM	✗	✗	✓	✗	✗	✓
CRFP	✗	✗	✓	✗	✗	✓
TCOF	✗	✓	✓	✓	✗	✓
TUFS	partly	✓	✓	✓	partly	✓
Rhapsodie	prosody	✗	✓	✗	✓	✓
ESLO	✓(and prosody)	✗	✓	✓	✓	✗
SLAM	✓	✗	✓	✗	✓	✓
DinG	✓	✓	✓	✓	✓	✓

Table 3.10: Comparison of the characteristics of (oral) corpora transcriptions. The ✓ stand for “has the feature”, ✗ for “doesn’t have it”.

in form to DinG. We have excluded STAC and FQB from this comparison because they are composed of written data. DinG is composed of detailed transcriptions whose characteristics fit well in the usual practices of oral corpora transcription. Transcribing with ELAN allowed us to precisely annotate overlaps, which are not transcribed in S-BNC, C-ORAL-ROM, CRFP, and TCOF. As DinG is designed for the study of question and answer relationship in dialogue, it is important to note that it can be easily compared to all the aforementioned corpora but C-ORAL-ROM, CRFP, and TCOF in terms of transcription of interrogatives. Finally, as DinG’s transcription contain annotations of laughter and disfluencies, they can be used for studies outside the scope of this thesis.

The aim of this comparison is to show that while DinG is relatively small in terms of duration and amount of recordings, it is comparable to state-of-the-art corpora in terms of the amount of speech turns and of words. DinG features detailed annotation of phenomena such as interrogatives, pauses, onomatopoeia, laughter, overlaps, and disfluencies, which makes it an interesting addition to the landscape of corpora of spoken French for further research projects. There is of course much space for progress, in particular in terms of integration of DinG to existing projects and datasets, such as CLAPI.

### 3.5 Conclusion of the chapter

This chapter thoroughly presents the *Dialogues in Games* corpus. The development of this corpus has been a major part of this thesis, yet we hope that the scientific journey of DinG is only starting. We developed the corpus and put it in perspective of the existing resources. When choosing the game(s) on which to focus, we specifically focused on *Catan* in order to produce a resource that could be used as support for annotation and analysis with the (slightly adapted) SDRT schemas that were developed for STAC. We hope that this analysis will lead to a meaningful comparison between the two corpora. Furthermore, we are eager to discover what parallels one will be able to draw between the two corpora, the two languages, the two communication modalities.

We envision three perspectives for further development of DinG: its extension, its annotation, and its usage. A first step would be to transform it to fit the TEI format<sup>93</sup> [Parisse and Liégeois, 2020]. Another path we envision is through the anonymization of the recordings through approaches such as the ones described in [Qian et al., 2017]. This would require contacting all the participants to draw a new consent form and might result in a slight shrinking of the corpus. Once both the transcriptions and the oral data available, a synchronization work would have to take place to enrich the resource, which will have to be validated in order to be further distributed.

Transcription constitutes a first level of linguistic annotation. One of our goals would be to offer other annotations, at different linguistic levels: morphosyntactic, part-of-speech, disfluencies, syntactic (through universal dependencies, for example). We would also like to annotate on layers specific to dialogue: dialogue transactions, connectives, argumentation structures, in particular, throughout the annotation schemata that were developed for the STAC project. Another interesting annotation-based investigation would lie in the classification and analysis of non-answer responses such as (but not only) clarification requests: utterances that follow questions, do not answer them, but do not ignore them either.

Finally, this corpus can be used as a starting point for fine-grained analysis on the mechanisms underlying the articulations of questions and answers in French. The theoretical use of this corpus is the object of this thesis. A first step towards a broader practical application would be the inclusion of DinG in the French Question banks [Judge et al., 2006, Seddah and Candito, 2016].

Recording people playing only one game restricts the range and variety of interactions one can observe in the corpus. As it happens, we consider this to be a current strength of DinG for our logical modeling purposes: the task of modeling real-life interactions is hard enough without broadening it just yet. Still, as the name of the corpus suggests, we would very much like for our corpus to thrive in the future and be expanded by adding recordings from other board games.

This chapter had a goal other than introducing DinG – more largely, its purpose was to talk about real-life data, its challenges, and its perks. DinG, in its status of real-life corpus of non-constrained oral French, is used in this thesis to develop and test a logical model of human-human dialogues. DinG presents both very well behaved examples with respect to standardized French grammar and pronunciation (see example 64) and very complex ones, such as example 65. Real-life dialogues combine several layers of complicated dialogical phenomena such as clarification requests, non-sentential utterances, ellipses, etc.

#### Example 64 (DinG8, almost standard French)

*In Catan, profit depends on the number on which the die falls.*

**Blue**<sub>1</sub> *C'est quoi qui rapporte le plus de thune ?*

**Blue**<sub>1</sub> *“What brings in the most cash?”*

**Yellow**<sub>2</sub> *Le 6*

**Yellow**<sub>2</sub> *“The 6”*

#### Example 65 (DinG6, complex utterance construction)

**Yellow**<sub>1723</sub> *personne veut toujours échanger du bois ?*

**Yellow**<sub>1723</sub> *“no one wants still exchange wood?”*

**Blue**<sub>1724</sub> *non*

**Blue**<sub>1724</sub> *“no”*

<sup>93</sup><https://tei-c.org/Guidelines/>, visited on 21/04/2021.

This concludes the chapter about real-life data corpora. The next chapter showcases the methods that allow us to build bridges between the neat world of formal models and the (often) messy world of real-life data.



# Dialogue annotation

## Contents

---

<b>4.1</b>	<b>A linguistic-based classification . . . . .</b>	<b>116</b>
4.1.1	Classification schema . . . . .	116
4.1.2	Annotation experiments – English, French, Italian, and Mandarin Chinese . . . . .	120
4.1.3	The need for a simpler schema . . . . .	126
<b>4.2</b>	<b>Towards an automation of the classification . . . . .</b>	<b>126</b>
4.2.1	The annotation schemas . . . . .	126
4.2.2	Annotation experiments – English, Dutch, Spanish . . . . .	128
4.2.3	Results and evaluation . . . . .	129
<b>4.3</b>	<b>Annotation experiments on DinG . . . . .</b>	<b>130</b>
4.3.1	Automatic annotation . . . . .	130
4.3.2	Human annotation . . . . .	132
4.3.3	Discussion of chosen examples . . . . .	134
4.3.4	Results of the human annotation . . . . .	137
<b>4.4</b>	<b>Conclusion and future work . . . . .</b>	<b>140</b>

---

Chapter 2 shows a theoretical overview of dialogue, focusing in particular on utterances in context and specifically on questions and answers classification. We present existing typologies and give insights into the way they can be put to use for formal semantics modeling. Chapter 3 focuses on data, specifically on transcriptions of real-life oral dialogues. We introduce our corpus, called Dialogues in Games (DinG), we detail its construction process and we compare it with existing corpora of English and French dialogue.

The aim of the current chapter is to constitute a middle ground between Chapter 2 and Chapter 3 by showing how theories developed in Chapter 2 can be applied to real-life data alike to the one we presented in Chapter 3. We start with a first annotation schema for questions and answers in dialogue, which was developed by taking linguistic theories and classifications as a starting point, in section 4.1. After giving the definitions of the various categories of the classification, we present the results of annotation experiments, conducted on corpora in English, French, northern Italian, and Mandarin Chinese. The review of this first experiment shows numerous shortcomings, which orient us towards the development of alternative annotation schemas.

The next section of this chapter focuses on a simplified schema for the annotation of questions and answers, developed by taking data observations as a starting point and with the perspective of an automation of the annotation task in mind, see section 4.2. We explain and justify the simplification decisions that were taken and show the results of the annotation experiments conducted on corpora in English, Spanish, and Dutch. Finally, we apply the annotation schema for questions to the corpus of questions from DinG in section 4.3.

## 4.1 A linguistic-based classification

**Note.** This section corresponds to the main developments obtained while working in close collaboration with Maxime Amblard, Marta Carletti, Léa Dieudonat, and Yiting Tsai, presented in [Amblard et al., 2019].

The work presented in this section has been conducted in the perspective of a detailed work on the question-answer relationship, focusing in particular on coherence between questions and answers. Because of this, the methodology was often to consider the utterances as question-answer pairs and outline the classification from there. This approach is different from the one taken in most of Chapter 2.

### 4.1.1 Classification schema

We classify the questions and the answers according to their *form* and their *function*, following [Freed, 1994]. The form of an utterance is defined by its syntactic form – such as syntactic inversions – and the lexical items that it contains (*wh*-words, ‘yes’, ‘no’, *etc.*), but also through the type of utterances it can be associated with (e.g. what type of answer does a question accept). The function of an utterance is close to the concept of Austin’s illocutionary force [Austin, 1975]: it is defined by the intention of the speaker.

Two criteria were used to establish the question form classification. First, we conducted an analysis of the lexical items the question contains. This criteria allowed us to identify the questions of the following form: *wh*-questions, *disjunctive questions*, *auxiliary questions*. We have presented *wh*-questions and disjunctive questions in Chapter 2; the first criteria consists in identifying the presence of *wh*-word for the *wh*-questions, and the presence of an “or” for a disjunctive question. As for auxiliary questions, they are identified and defined by the presence of an auxiliary verb: “be”, “can”, “could”, “dare”, “do”, “have”, “may”, “might”, “must”, “need”, “ought”, “shall”, “should”, “will”, “would” [Jurafsky and Martin, 2000].

The second criterion that was used to establish the question form classification is classifying the questions according to the type of answer they accepted. This criterion allowed us to classify the following types of questions: *yes/no questions*, *inclusive disjunctive questions*, *exclusive disjunctive questions*, *action auxiliary questions* and *epistemic auxiliary questions*.

#### Definition 29 (Inclusive and exclusive disjunctive questions)

A question  $Q_d$  containing a disjunction (“or”) is disjunctive inclusive if it is possible to answer  $Q_d$  with an answer encompassing all the options proposed by  $Q_d$ .  $Q_d$  is disjunctive exclusive if the only acceptable answers are those corresponding to one and only one of the options proposed by  $Q_d$ .

Example 66 shows two very similar disjunctive questions  $\mathbf{A}_1$  and  $\mathbf{A}'_1$ . In the case of  $\mathbf{A}_1$ , all the answers proposed in  $\mathbf{B}_2$  are acceptable:  $\mathbf{A}_1$  is disjunctive inclusive. For  $\mathbf{A}'_1$ , holding the hypothesis that sugar and stevia should not be used concurrently, the answer  $\mathbf{B}'_2$  “Both” or “Yes”

is not an acceptable one: it might yield clarification requests or at minimum, a slight eyebrow raising.  $A'_1$  is disjunctive exclusive.

**Example 66 (Examples of inclusive and exclusive disjunctive questions)**

$A_1$  Do you want sugar or milk in your coffee?     $A'_1$  Do you want sugar or stevia<sup>94</sup> in your coffee?

$B_2$  Sugar/Milk/Neither/Both/?Yes                       $B'_2$  Sugar/Stevia/Neither/#Both/#Yes

Where the ‘?’ preceding “Yes” in  $B_2$  corresponds to a doubtful usage: it goes slightly against the usage in English, but can still be understood by speakers in context.

**Definition 30 (Action auxiliary questions)**

Action auxiliary questions are auxiliary questions (questions containing auxiliary verbs) that ask the addressee for an action. The answer to an action auxiliary question should be an action, either physical ( $A$  says “could you pass me the salt?”,  $B$  passes the salt) or communicational (“can you repeat louder?” “IT WAS LAST YEAR [...]”).

**Definition 31 (Epistemic auxiliary questions)**

Epistemic auxiliary questions are auxiliary questions (questions containing auxiliary verbs) that interrogate the addressee’s knowledge and/or beliefs (“do you believe in magic?”/“can you speak French?”).

Epistemic auxiliary questions are very similar in syntactic form to action auxiliary questions. The main difference between the two lies in the type of answer they accept, which makes the second criterion to establish the question form classification particularly important in this case. It is important for us to distinguish the two auxiliary questions forms because of the question-answer pair perspective: as the two question forms accept very different types of answers, we need to be able to distinguish one from another to be able to study the coherence between questions and answers and be able to pinpoint incoherences.

Table 4.1 summarizes the question forms we have established based on the two criteria presented above. Starting from this list of question forms, we were able to build a list of corresponding answer forms. The answer forms corresponding to yes/no-questions and *wh*-questions are quite straightforward: those are *yes/no answers* and *wh-answers* (answers that correspond to the thematic role interrogated by the *wh*-word in the *wh*-questions, see Chapter 2, section 2.6 for a detailed discussion). Then, we need answer forms that will account for three other types of answers: the ones containing lexical items conveying uncertainty such as “maybe”, the ones containing clear statements of an incapacity to answer the question (e.g. “I don’t know”), and finally, we need a category for the answers that do not have any specific syntactic form or contain any characteristic lexical item. This gives us the following three forms of answers: *uncertain*, *unknown*, *other*. The forms of answers are summarized in table 4.2.

Now that we have established a list of forms for questions and answers, we can focus on the functions of the utterances: when the utterance is said, what is the goal of the speaker? The list of functions of question is presented in table 4.3. Some of the functions are directly derived from the form of the utterance: for example, a yes/no-question very often asks for a confirmation,

<sup>94</sup>Stevia is a plant-based widely-used sugar substitute. For this example to work well, we suppose that people do not mix sugar with a sugar substitute, though it has been pointed out to us that everything is possible.



Name	Tag	Description	Examples
Yes/no	YN	Contains do-support, inversions	“do you want clay?” “have you finished?”
Wh	WH	Starts with a <i>wh</i> -word	“how many cards do you have?”
Disjunctive inclusive	DQ_I	Contains an “or”, accepts inclusive answers	“do you want sugar or milk in your coffee?”
Disjunctive exclusive	DQ_E	Contains an “or”, doesn’t accept inclusive answers	“do you want sugar or stevia in your coffee?”
Action auxiliary	AUX_A	Contains an auxiliary, asks for an action	“could you open a window?”
Epistemic auxiliary	AUX_E	Contains an auxiliary, asks for an action	“do you know who has clay?”

Table 4.1: Forms of questions.

Name	Tag	Description	Examples
Yes/no	YN	Contains linguistic cue for confirmation or denial	“yes” “yeah”, “sure”, “no”, “nope”, “not at all”
Wh	WH	Corresponds to a thematic role	“Charlie”, “4”, “at 5”, “in Paris”
Uncertain	UNC	Contains linguistic cue for uncertainty	“I’m not sure”, “maybe”
Unknown	UNK	Contains linguistic cue for absence of knowledge	“I don’t know”, “can’t know”
Other	O	Cannot be classified as any of the above	“just go to the tea shop!”

Table 4.2: Forms of answers.

which corresponds to the *ask-confirmation* function; an action auxiliary asks for an action to be performed, which corresponds to the *ask-performance* function. Others are more subtle: a question can be of yes/no- or *wh*- form, but be phatic in function; for example, “why is this happening to me?” contains the *wh*-word “why”, but it is easy to imagine a context in which this question would not actually call for a response.

The other functions in table 4.3 are inspired by real-life data observations, developed in section 4.1.2. A speaker can complete the turn of another speaker; this type of utterance can take multiple forms but always bear the same intention, so the same function. Similarly, it can happen in dialogue that a speaker repeats something that has been said by another person, maybe even in a different conversation. Though this repeated utterance can take virtually any form, its main function is that of a reported speech as it will be perceived as such by the addressee; a proof of this lies in the fact that when an addressee fails to perceive an utterance as a reported speech, the resulting incomprehension can trigger a clarification discussion.

Table 4.4 presents the functions of answers. Phatic and reported speech are common with questions. The other functions correspond to the detailed lists of questions forms and functions, as each of those triggers a specific type of expected answer; for example, a question of form *auxiliary-action* and of function *ask-performance* is likely to yield an answer of form *other* and of *perform* function. Table 4.5 summarizes the classification schema we presented in this section.

Name	Tag	Description	Examples
Completion suggestion	CS	The speaker proposes a possible end to another speaker's utterance	"I don't have any..." / "clay?"
Phatic	PHA	The goal of this utterance is metacommunicational	"what ?", "why are you doing this?"
Ask confirmation	ASK_CONF	The speaker asks for a confirmation of an informational content	"do you have clay?", "is it my turn?"
Ask wh	ASK_WH	The goal of this utterance is to get an answer corresponding to a <i>wh</i> -query	"where is the clay?", "how do we win?"
Ask performance	AUX_PERF	The speaker asks for an action to be performed	"could you open a window?"
Reported speech	RS	The speaker rephrases a question another speaker asked	"she was asking where the sheep were"

Table 4.3: Functions of questions.

Name	Tag	Description	Examples
Refuse	REF	The speaker refuses the content of the question	"no", "nope", "not at all"
Accept	ACC	The speaker accepts the content of the question	"ok"
Phatic	PHA	The goal of this utterance is metacommunicational	"well of course this happens", "come on!"
Give confirmation	GIVE_CONF	The speaker confirms the content of the question	"yes", "sure"
Give uncertainty	GIVE_UNC	The speaker conveys uncertainty	"I'm not sure", "maybe"
Give unknown	GIVE_UNK	The speaker conveys absence of knowledge	"I don't know", "can't know"
Reported speech	RS	The speaker repeats something another speaker said	"she said the sheep was there"
Give wh	GIVE_WH	The speaker gives an answer that corresponds to the <i>wh</i> -content asked by the question	"Charlie", "4", "at 5", "in Paris"
Perform	PERF	The speaker performs the required action	"it's 3 o'clock"

Table 4.4: Functions of answers.

	Questions	Answers
<b>Forms</b>	YN, WH, DQ_I, DQ_E, AUX_A, AUX_E	YN, WH, UNC, UNK, O
<b>Functions</b>	CS, PHA, ASK_CONF, ASK_WH, ASK_PERF, RS	REF, ACC, PHA, GIVE_CONF, GIVE_UNC, GIVE_UNK, RS, GIVE_WH, PERF

Table 4.5: Forms and functions of questions and answers.

### 4.1.2 Annotation experiments – English, French, Italian, and Mandarin Chinese

Once the annotation schema has been defined, it was used by three annotators on a subset of the Saarbrücken Corpus of Spoken English [Norrick, 2017]<sup>95</sup>, see Chapter 3, section 3.2.1 for a complete presentation. The annotators worked on 90 question-answer pairs extracted from transcriptions. After the annotation was completed, the results were put in common to decide on a golden annotation.

	Question form	Question function	Answer form	Answer function
A1 + A2	0.94	0.91	0.90	0.88
A1 + A3	0.94	0.88	0.85	0.85
A2 + A3	0.89	0.88	0.84	0.79

Table 4.6: Inter-annotator agreement scores (Cohen’s  $\kappa$ ) between annotators 1, 2, and 3 (A1, A2, A3) for the annotation of the corpus of 90 question-answer pairs from SCoSE.

The results presented in table 4.6 are very high, but this can be explained by the fact that all three annotators took part in the design of the annotation schema and the related discussions, in particular for potentially problematic examples. The first inter-annotator agreement scores they achieved were around 0.65.

After the annotators trained together on the English corpus of question-answer pairs, each of them annotated a corpus in another language: in French, in (northern) Italian, and in Mandarin Chinese.

**French** The annotations in French were conducted on a corpus of 79 question-answer pairs extracted from the adult interactions corpus of TCOF<sup>96, 97</sup>, [ATILF, 2020, André and Canut, 2010], see Chapter 3, section 3.2.2 for a complete presentation. The transcriptions extracted from this corpus do not have explicit punctuation marking, in particular, they do not feature question marks. The questions had to be annotated both using linguistic cues such as *wh*-words and prosody that the annotator could hear through the corresponding audio file.

**Italian** The annotations in Italian were conducted on a corpus of 78 question-answer pairs extracted from the Archivio di Parlato Italiano (API, [Crocco and Savy, 2003]). The annotated dialogue is a spontaneous conversation between two interlocutors from Naples, speaking in a northern variation of Italian. One of the main difficulties in using the annotation schema presented in the previous section for this corpus lies in the fact the questions, in Italian, do not

<sup>95</sup><https://www.uni-saarland.de/lehrstuhl/gergel/scose.html>, visited on 18/03/21.

<sup>96</sup><http://www.cnrtl.fr/corpus/tcof/>, visited on 17/05/2021.

<sup>97</sup><https://www.ortolang.fr/market/corpora/tcof>, visited on 21/05/2021.

involve any inversion or morphological insertions. As for the answers, it appears that in real-life Italian dialogues, interjections such as «eh» are widely used to respond to questions. However, the meaning of «eh» changes according to intonation and can express doubts, confirmation, surprise, disappointment, exaggeration, etc. This specificity constitutes a huge difficulty for a systematic answer annotation.

**Mandarin Chinese** The annotations in Mandarin Chinese were conducted on a corpus of 287 question-answer pairs extracted from the NCCU Corpus of Spoken Taiwan Mandarin [Chui and Lai, 2008]. The corpus is composed of open-topic conversations, recordings, and transcriptions. The transcriptions contain markings of speech overlap and pauses, but do not have explicit marking of questions. The annotator had to proceed to the annotation while listening to the recording. One of the main difficulties during the annotation comes from the fact that the speakers in the NCCU corpus mix Mandarin, Southern Min, Hakka, Japanese, and English through the conversations. Southern Min and Hakka are two major Chinese dialects spoken in Taiwan. Because of this, some additional annotations were made on a corpus of 22 question-answer pairs extracted from the PolyU Corpus of Spoken Chinese<sup>98</sup>, in which the participants were specifically requested to speak in standard Mandarin only.

	English	French	Italian	Chinese
Interrogative words	✓	✓	✓	✓
Prosody	✓	✓	✓	✓
Inversion	✓	✓	✗	✗

Table 4.7: Linguistic cues for question annotation across English, French, Italian and Mandarin Chinese.

Table 4.7 summarizes the linguistic cues for the annotation of question forms that can be used for each of the considered languages. Tables 4.8, 4.9, 4.10, and 4.11 and the corresponding visualisations in figures 4.1, 4.2, 4.3, and 4.4 show the statistic distribution of the annotation tags for English (EN), Italian (IT), French (FR) and Mandarin Chinese (CN).

	EN	IT	FR	CN
YN	76,0%	70,6%	62,4%	69,5%
WH	17,6%	22,7%	29,6%	28,3%
DQ_I	0%	0%	2,0%	0,6%
DQ_E	3,6%	2,7%	0%	1,3%
AUX_A	0%	4,0%	4,0%	0%
AUX_E	2,8%	0%	2,0%	0,3%

Table 4.8: Statistic distribution of question forms, in percentage, approximated to the highest closest decimal, across English (EN), Italian (IT), French (FR) and Mandarin Chinese (CN) according to table 4.5.

**Question forms** It is interesting to observe in table 4.1 and figure 4.1 that the YN tag is high in all 4 languages (> 62%). The next most frequent category for all the languages is WH, between 17 and 30%. The next categories are less consensual: English, Italian, and Mandarin Chinese

<sup>98</sup>[https://chaaklau.github.io/poly\\_corpus\\_static/#man](https://chaaklau.github.io/poly_corpus_static/#man), visited on 24/07/21.

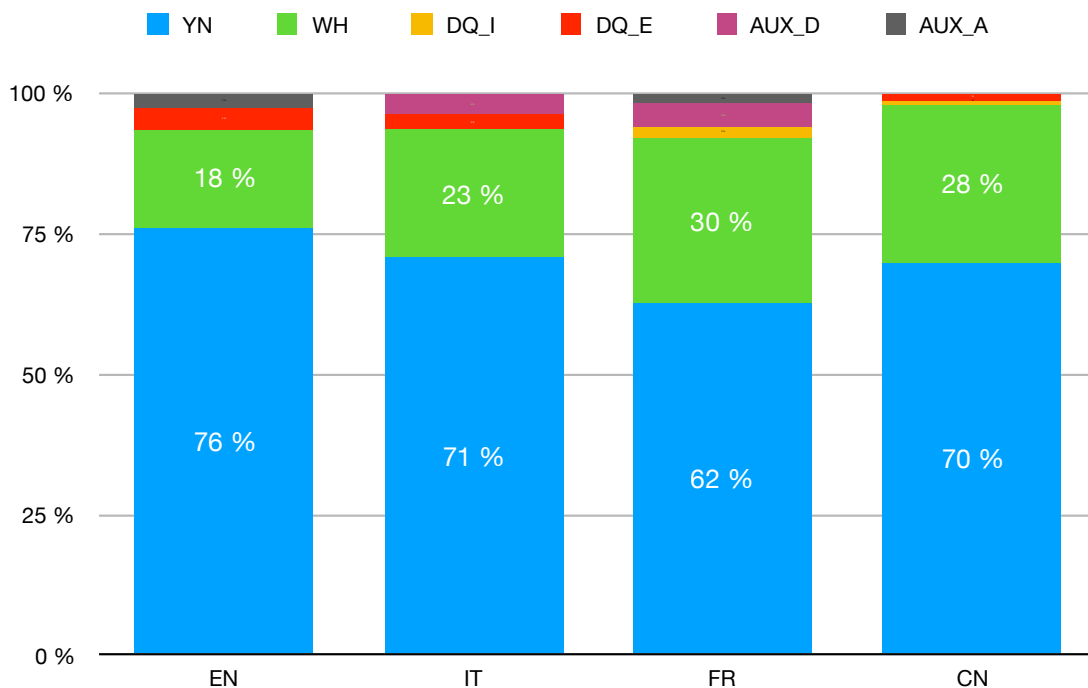


Figure 4.1: Visualization of the statistic distribution of question forms (in percentage) across English (EN), Italian (IT), French (FR) and Mandarin Chinese (CN) according to table 4.5, presented in table 4.8.

corpora contain a significant amount of disjunctive exclusive questions ( $> 1\%$ ) while French does not; at the same time, the observation is the opposite for disjunctive inclusive questions. There are two possible interpretations for this result: either it is characteristic of the chosen corpora, or the annotator for French did not interpret the disjunctive questions in the same way as the two others.

	EN	IT	FR	CN
<b>YN</b>	39,4%	58%	42,6%	63,8%
<b>UNC</b>	4,3%	1,4%	2,0%	9,7%
<b>UNK</b>	0,7%	0%	3,0%	0,5%
<b>WH</b>	15,8%	28,4%	28,7%	26,0%
<b>O</b>	39,8%	12,2%	23,7%	0%
<b>PERF</b>	0%	0%	0%	0%

Table 4.9: Statistic distribution of answer forms, in percentage, approximated to the highest closest decimal, across English (EN), Italian (IT), French (FR) and Mandarin Chinese (CN), according to table 4.5.

**Answer forms** Table 4.9 and figure 4.2 show that the YN tag is also very high in all 4 languages ( $> 39\%$ ) for answers. The next most frequent category for Italian, French, and Mandarin Chinese is WH, between 26 and 29%. For English, the second most frequent category is NONE/O, with the surprising score of almost 40%. Less than 1% of answers were annotated as PERF across all languages, which can easily be explained by the fact the performing answers can often be

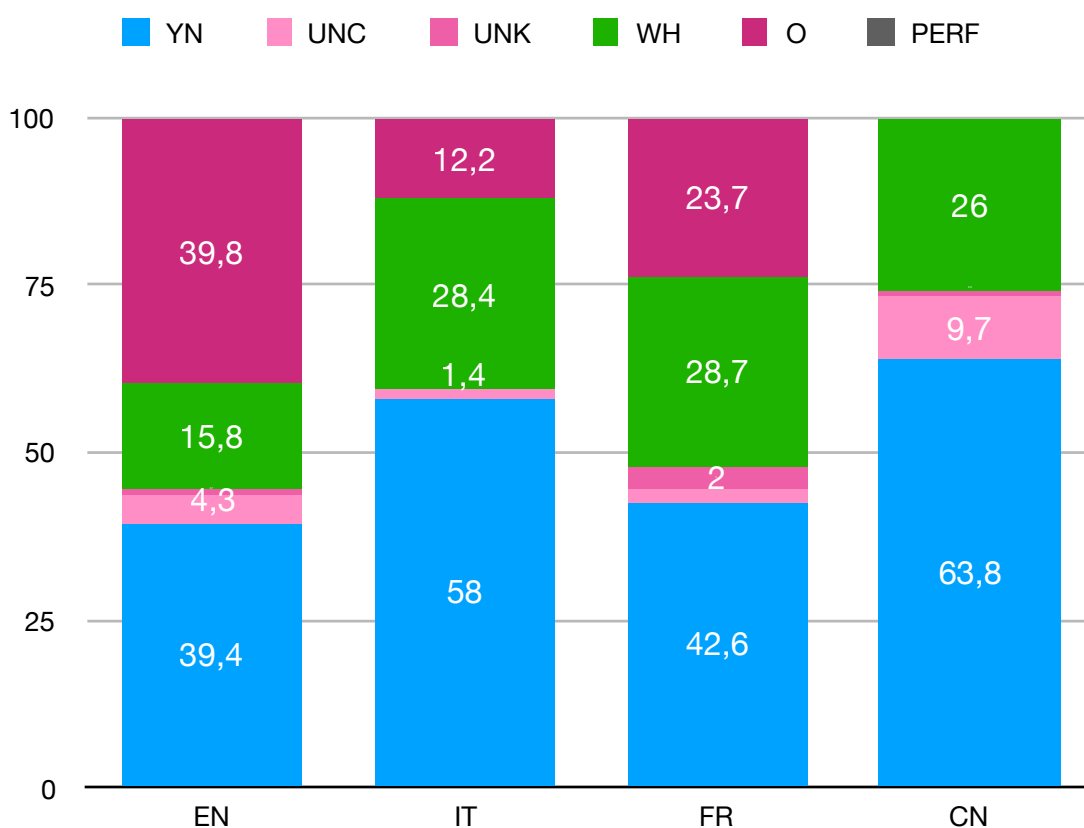


Figure 4.2: Visualization of the statistic distribution of answer forms (in percentage) across English (EN), Italian (IT), French (FR) and Mandarin Chinese (CN), according to table 4.5.

executed without any corresponding linguistic realization (for example, by passing the salt as an answer to the question “can you pass the salt please?”). **UNCERTAIN** appears to be a more frequent tag than **UNKNOWN** for English, Italian, and Mandarin Chinese, **UNKNOWN** being almost absent for Italian and Mandarin Chinese. In French, **UNCERTAIN** and **UNKNOWN** score respectively 2 and 3%.

	EN	IT	FR	CN
<b>CS</b>	2,9%	2,7%	1,0%	1,7%
<b>PHA</b>	37,3%	0%	3,0%	20,3%
<b>ASK_CONF</b>	35,0%	69,3%	66,3%	55,0%
<b>ASK_WH</b>	15,8%	25,3%	29,7%	22,7%
<b>ASK_PERF</b>	0%	2,7%	0%	0%
<b>RS</b>	9,0%	0%	0%	0,3%

Table 4.10: Statistic distribution of question functions, in percentage, approximated to the highest closest decimal, across English (EN), Italian (IT), French (FR) and Mandarin Chinese (CN) according to table 4.5.

**Question functions** Following the observation presented in table 4.8 and figure 4.1, it is not surprising to notice in table 4.10 and figure 4.3 that the **ASK\_CONF** tag is the most used one for Italian, French and Mandarin Chinese (> 55%). It is more interesting to see that it is not the

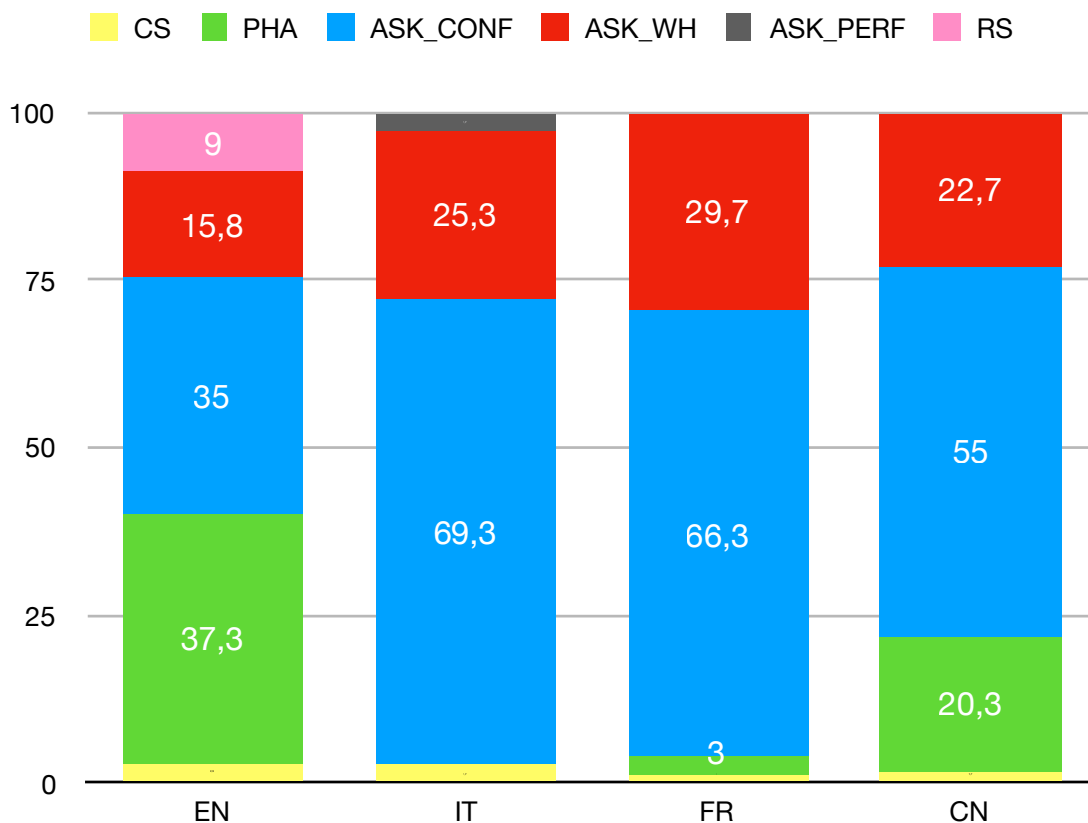


Figure 4.3: Statistic distribution of question functions (in percentage) across English (EN), Italian (IT), French (FR) and Mandarin Chinese (CN) according to table 4.5.

case for English, but this observation can be explained by the fact that the PHA tag is higher than the ASK\_CONF tag for English. ASK\_CONF and PHA are both functions that can be taken by yes/no-questions, which as we have seen in table 4.8 and figure 4.1 constitute more than 62% for all 4 languages. It is, again, impossible to say whether the amount of phatic questions annotated for each language is specific to the chosen corpus or the interpretation of the person who annotated the corpus. The RS tag is significantly used only in the annotation of English and the ASK\_PERF one, only for Italian.

**Answer functions** Table 4.11 and figure 4.4 show three highest categories across the 4 languages: GIVE\_CONF, GIVE\_WH, and NONE. The other tags correspond to 11.2% of the answers for the English corpus, 5.1% for the Italian one, 5.1% for the French one, and 23.5% for the Mandarin Chinese one. The ACCEPT tag is significant ( $> 1\%$ ) only for Italian. The REFUSE tag is relatively high for English and Mandarin Chinese as in both cases, it accounts for more than half of the answers that were left from tagging GIVE\_CONF, GIVE\_WH, and NONE. The PHA tag is significant for English and Mandarin Chinese only and even there, in very small proportions. GIVE\_UNC is common to the 4 languages whereas GIVE\_UNK is significant only for French. GIVE\_PERF is significant only for Italian, where it accounts for more than half of the answers that were left from tagging GIVE\_CONF, GIVE\_WH, and NONE. RS was used for English (less than 1% of tags) and Mandarin Chinese.

	EN	IT	FR	CN
<b>ACCEPT</b>	0,2%	1,4%	0%	0%
<b>REFUSE</b>	5,7%	0%	1,0%	12,7%
<b>PHA</b>	1,4%	0%	0%	2,6%
<b>GIVE_CONF</b>	44,3%	61,1%	52,0%	31,4%
<b>GIVE_UNC</b>	3,2%	1,4%	2,0%	6,9%
<b>GIVE_UNK</b>	0%	0%	3,0%	0%
<b>PERF</b>	0%	2,8%	0%	0%
<b>GIVE_WH</b>	10,0%	26,4%	21,0%	25,8%
<b>NONE</b>	34,4%	6,9%	21,0%	19,3%
<b>RS</b>	0,7%	0%	0%	1,3%

Table 4.11: Statistic distribution of answer functions, in percentage, approximated to the highest closest decimal, across English (EN), Italian (IT), French (FR) and Mandarin Chinese (CN) according to table 4.5. The NONE tag was added to account for the answer forms that couldn't be assigned any tag from the tag list.

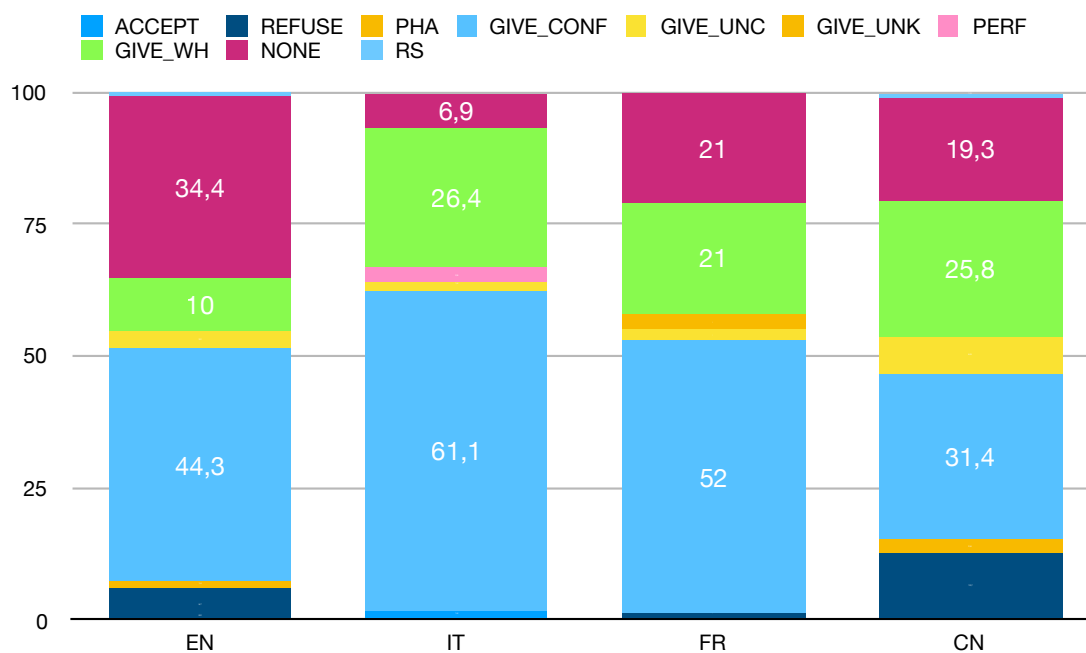


Figure 4.4: Statistic distribution of answer functions (in percentage) across English (EN), Italian (IT), French (FR) and Mandarin Chinese (CN) according to table 4.5. The NONE tag was added to account for the answer forms that couldn't be assigned any tag from the tag list.

These observations conclude our first annotation experiment. The next section reviews the annotation schema along with the results of our experiment. We sum up the highlights and shortcomings of this first annotation schema and outline the next steps of our experiments.



### 4.1.3 The need for a simpler schema

The considerations developed in the previous section call for a twofold observation. First, while the question and answer form tagging schema appears to be well-tailored for English, it does not straightforwardly transpose to other languages. Then, the multiplication of tags for question and answer function seems to introduce unnecessary complexity in the annotation schema that either has no linguistic requisite for in the corpora chosen for the study or was not correctly implemented by the annotators who participated themselves in the design of the annotation schema. Either way, the conclusion of this section is that while this schema is linguistically inspired and tailored for English, it is not suitable for a larger use such as cross-linguistic exploration or automation of the annotation: if the schema cannot be correctly implemented by the human annotators that participated in its design, it is likely that it will be difficult to algorithmically implement it for automatic annotation in its current state.

The next section outlines a simpler annotation schema, designed while keeping the automation of the annotation and the cross-linguistic transferability in mind.

## 4.2 Towards an automation of the classification

**Note.** This section takes roots in the main developments obtained while working in close collaboration with Maxime Amblard, María Andrea Cruz Blandón, Gosse Minnema, and Aria Nourbakhsh, presented in [Cruz Blandon et al., 2019].

### 4.2.1 The annotation schemas

The annotation schema for questions and answers in dialogue presented in this section takes roots in two existing schemas: first, [Freed, 1994], which categorizes questions along an information continuum that ranges from questions purely asking for factual information to questions that convey, rather than request, (social) information. Within this continuum, questions are divided into classes that are defined based on a combination of formal (syntactic) and functional criteria. Both of these ideas are also used in our scheme: our question types are also distinguished by whether they ask or convey information (‘phatic questions’ and ‘completion suggestions’ fall into the latter category) and are defined as combinations of specific forms and functions.

The second annotation schema we took into account is [Stolcke et al., 2000], adapted from DAMSL, see Chapter 2, section 2.2.1 for a detailed presentation of DAMSL. This schema includes a set of eight different question types that has considerable overlap with our set of question types. Our schema contains two sets of tags: the question types and the answer types.

**Question annotation** The question tagset was designed in a corpus-driven way, starting with two basic types and expanding the tagset based on corpus data. The starting corpus was SCoSE, see Chapter 3, section 3.2.1 for a complete presentation. Our beginning assumption is that the corpora would contain at least two well-known and well-defined categories of questions: yes/no-questions and *wh*-questions. This assumption is supported by the results of the cross-lingual experiments presented in section 4.1.2.

Next, we looked for questions in our corpora that did not correspond to either of these two types and extended the scheme to fit them. First, there are questions that are similar to *wh*-questions or yes/no-questions in usage but have a different form: for example, *wh*-in-situ questions like ‘You saw what?’, or yes/no questions without inversion such as ‘You saw him?’. We decided not to introduce new categories for these on the basis of their semantics and pragmatics.

A second group of questions has the syntactic characteristics of a yes/no-question or a *wh*-question, but are used with different pragmatics and/or semantics. For example, the asker of the question suggests a way to complete the utterance of the previous speaker, and the expected answer would confirm or deny this suggestion. This is subtly different from a prototypical yes/no-question because the asker of the question does not necessarily ask their interlocutor to confirm the truth value of the suggestion. We call these types of questions *completion suggestions*<sup>99</sup>.

### Example 67 (Completion suggestion, SCoSE)

**A**<sub>1</sub> it includes heat and uhm, I think

**B**<sub>2</sub> water?

In example 67, the interrogative intonation used by **B** in **B**<sub>2</sub> can refer either to the completed question (“does it include heat and water?”) or on the contrary to the completion suggestion itself (“is the word you’re looking for ‘water’?”).

The third group of questions takes the appearance of yes/no-question or a *wh*-question, respectively, but the context and intonation of the utterance make clear that the asker is not actually interested in the confirmation or denial of the proposition. Instead, such questions can have various so-called *phatic functions*, i.e. their semantic content is less important than their social and rhetorical functions ([Freed, 1994], [Senft, 2009]). We call this type of questions *phatic questions*: examples include “right?”, “oh yeah?”, “you know?”. Our definition of phatic questions is somewhat broader than the phatic information question described in [Freed, 1994]; in particular, our definition counts rhetorical questions as phatic questions, which is not the case in Freed’s original scheme.

Finally, some questions containing a disjunction (e.g. ‘Do you go on Monday or on Tuesday?’) are semantically and pragmatically similar to *wh*-questions, but are syntactically closer to yes/no-questions. This kind of question, like yes/no-questions, exhibits subject-auxiliary inversion (in English) but does not ask for the confirmation or denial of the proposition that it expresses. Instead, it expects the answerer to provide some missing information with the set of options to choose from. We call this type of question *disjunctive questions* (sometimes also called alternative questions in the literature). We do not draw a distinction between disjunctive inclusive and disjunctive exclusive in this schema as the experiments presented in section 4.1.2 show us that the difference between the two is often given by the prosody of the utterance, making the classification more open to interpretation of the annotator and less direct to automate using transcriptions only.

Tag	Name
YN	yes/no-question
WH	<i>wh</i> -question
DQ	disjunctive question
CS	completion suggestion
PQ	phatic question

Table 4.12: Set of question tags.

The tagset we obtain for the questions types is summarized in table 4.12. It is significantly simpler than the one presented in table 4.5; this entails that the annotations we will obtain using

<sup>99</sup>As pointed out by Jonathan Ginzburg, this category of questions is called *filler* in [Purver et al., 2003].

this new schema will be less precise than the ones obtained with the previous one. However, a simpler schema can also mean that the annotation task is simpler, therefore the annotations contain less interpretation bias/fewer errors, and the annotators can work faster, therefore, annotate more examples and get statistically more significant results.

**Answer annotation** The main idea underlying our answer annotation scheme is that question types restrict their answers: for example, yes/no-questions are prototypically answered by linguistic variations on ‘yes’ or ‘no’, and *wh*-questions ask for a constituent corresponding to the interrogated *wh*-word. Table 4.13 summarizes the answer types we consider along with their corresponding question types. Among these types of answers, there may be overlaps. For example, a ‘deny the assumption’ answer can be thought of as a negative answer because it is possible that they share the same grammatical and semantic structure. Different factors including the context and prosody are relevant to decide between overlapping tags.

Tag	Name	Corresponding question tags
PA	positive answer	YN, CS
NA	negative answer	YN, CS
WHA	<i>wh</i> -answer	DQ, WH
PHA	phatic answer	YN, WH, DQ, CS, PQ
UA	uncertainty answer	YN, WH, DQ, CS, PQ
UT	non-topical answer	YN, WH, DQ, CS, PQ
DA	deny the presupposition	YN, WH, DQ, CS, PQ

Table 4.13: Set of answer tags and corresponding question tags.

Some questions are not followed by an answer. We distinguish between two situations. First, there are questions that receive a reply that, while not providing the information asked for in the question, clearly do respond to it. In such cases, the response is tagged as unrelated topic (UT) because it is about a different topic but still responds to the question. By contrast, when there is no response at all, no answer should be annotated.

The answers expressing uncertainty and absence of knowledge are grouped under the common tag UA for “uncertainty answer”. Merging the insight from answer forms and functions from the tagset presented in 4.5 allows us to produce a clearer and simpler schema, that is hopefully easier to use and can avoid the problems we ran into in the experiment phase, in particular in the cross-lingual setting.

## 4.2.2 Annotation experiments – English, Dutch, Spanish

Once the annotation schema was defined an annotation guide was written. It can be found here: [https://github.com/andrea08/question\\_answer\\_annotation](https://github.com/andrea08/question_answer_annotation)<sup>100</sup>. The annotation guide specifies a precedence order for question tagging, in order to establish a smooth and homogeneous annotation process. The precedence order dictates what category should be checked first when trying to assign a tag to a question: (1) *Wh*-question, (2) Disjunctive question, (3) Yes/no-question, (4) Completion suggestion, (5) Phatic question.

The annotation schema was then used by three annotators on a corpus of 701 questions and 483 answers across three languages: English, (Mexican) Spanish, and Dutch. The three

<sup>100</sup>Note that the annotation process presented in this thesis is slightly simplified compared to the original one from the paper, for the sake of readability.

corpora were the Saarbrücken Corpus of Spoken English (SCoSE), a corpus of face-to-face conversations; the CallFriend corpus for Spanish [Canavan and Zipperlen, 1996], a corpus of phone conversations; and the Spoken Dutch Corpus (CGN) [Oostdijk, 2001], also composed of phone conversations. The questions and answers were distributed as follows: 422 questions and 289 answers from the SCoSE corpus; 87 questions and 72 answers from the CGN corpus; and 192 questions and 122 answers from the CallFriend corpus. The difference between the amount of questions and the amount of answers can be explained by several reasons: first, not all questions obtain answers; then, sometimes a person uses several questions to convey one interrogation, which then asks for only one answer. A descriptive analysis of the annotations shows that yes/no-questions constitute the most common tag in the three corpora: 40% for Spanish, 42% for English, and 64% for Dutch.

### 4.2.3 Results and evaluation

To evaluate the annotations, the inter-annotator agreement was calculated based on a subset of the SCoSE corpus. Cohen’s  $\kappa$  [Cohen, 1960] for question types annotation (for all annotators combined) is of 0.63; for answer annotation, it is of 0.49. 0.63 would generally be considered to be a ‘moderate’ level of agreement [Landis and Koch, 1977]. A large share of the disagreements came from phatic questions; distinguishing these from other question types sometimes relies on subtle pragmatic and semantic contextual judgments. Agreement for answer types is lower than for question types because question types restrict answer types and hence question type disagreements can cause answer type disagreements. In order to improve the annotation guidelines, we systematically examined all of the disagreements, most of which fell into one of four categories:

1. Simple attention-related mistakes, such as missing a question or choosing an (obviously) wrong tag.
2. Disagreements as a consequence of a previous disagreement, as the ones with answer types that are linked with questions types.
3. Missing instructions in the annotation guidelines for handling particular situations, e.g. annotating utterances containing interruptions.
4. Utterances whose interpretation was ambiguous and depends on subtle prosody or contextual cues for which it is hard to formulate a general rule.

The annotation guide was updated in consequence, when possible.

Tables 4.14 and 4.15 show the distribution of each of the tags in our annotation schema, grouped by language. There are some differences between the languages, but in most cases, the global pictures are quite similar. For example, for question types (table 4.14), the most common tag in each language is YN (yes/no question), followed by WH (*wh*-question) or PQ (phatic question). However, exactly how prevalent YN is depends on the language: in the Dutch corpus, it accounts for almost 60% of the questions, while in Spanish and English corpora this is about 40%. As pointed out by Jonathan Ginzburg, one of the possible explanations for such as prevalence of yes/no-questions could be the absence of distinction between tag questions such as “isn’t it?” and genuine yes/no-questions.

In the distribution of answer types (table 4.15), there are several obvious patterns: for example, the Spanish and Dutch corpora have a much larger percentage of WHA (*wh*-answer) tags than the English one; this difference is caused by the larger number of WH-questions in these two languages. Something that is shared between all the three corpora is that in call cases, there

	YN	WH	DQ	CS	PQ
<b>English</b>	42.2%	23.5%	1.2%	1.7%	31.5%
<b>Spanish</b>	39.9%	33.0%	1.6%	1.1%	24.5%
<b>Dutch</b>	64.4%	26.4%	1.2%	0%	8.1%

Table 4.14: Statistic distribution of question tags (in percentage) across English, Spanish, and Dutch corpora.

are far more positive answers than negative answers. Interestingly, however, this difference is much larger in the Dutch corpus than in the other languages.

	PA	NA	WHA	PHA	UA	UT	DA
<b>English</b>	24.6%	12.8%	15.6%	25.6%	5.2%	15.6%	0.7%
<b>Spanish</b>	37.2%	13.3%	31.0%	6.2%	8.0%	2.7%	1.8%
<b>Dutch</b>	47.2%	1.4%	25.0%	8.3%	6.9%	9.7%	1.4%

Table 4.15: Statistic distribution of answer tags (in percentage) across English, Spanish, and Dutch corpora.

The results obtained with this new annotation schema are more encouraging than with the one presented in section 4.1, in particular regarding the annotation schema for questions. The annotation of answers still appears to be problematic because of the cumulative effect of annotation errors. A direction for further work should take this specificity into account and provide a way to bypass it.

### 4.3 Annotation experiments on DinG

Following the developments we presented in the previous sections, we annotated the 2,528 questions from the DinG corpus (see Chapter 3, table 3.5 for details by recording) using the annotation schema for questions present in table 4.12. As the questions were explicitly transcribed through the interrogative point marking, we were able to automatically retrieve all of them along with a small context containing the two preceding and the two following utterances. The questions were first automatically assigned tags, then annotated by hand to correct and augment this first annotation.

#### 4.3.1 Automatic annotation

The automated annotation of questions from DinG, called hereafter *utterance number n*, follows the next rules, in this order:

1. If the utterance number  $n + 1$  is affirmative (starts with the word « oui », « ouais » or « ok ») or negative (starts with the word « non »), the question in utterance  $n$  is automatically tagged as a yes-no question: YN.
2. If the utterance contains a French *wh*-word, the question is automatically tagged as a *wh*-question: WH. See Chapter 5, section 2.6.1, table 2.8 for the list of *wh*-words in French.
3. If the utterance contains « ou », the question is automatically tagged as a disjunction: DQ.

4. After first experiments with human annotators, we added a last category of automatic annotation: if the complete utterance is « (xxx) ? », which is interpreted as “the person who was transcribing could not figure out any words but still picked up an interrogative/rising intonation”, the question should automatically be tagged as a question that we cannot assign a category to: N/A.

As phatic questions and completion suggestion are categories that are highly context-dependant, we did not pre-annotate them automatically. Some of the utterances contain more than one interrogative points, which corresponds to several questions asked in a row, without pauses between them and with several rising intonation points. In this case, the automatic annotation only annotated once. The automatic annotation was able to assign a tag to 772 out of 2504 utterances containing at least one interrogation point, which corresponds to  $\sim 31\%$ .

The automatic annotation was systematically wrong in several cases:

- Several times, a *wh*-question is directly followed by a negation (see example 68). Following 1. in the automatic annotation rules, the question was assigned the YN tag in this case, while it should have been a WH.
- A lonely « *quoi ?* » (“what?”) is most of the time phatic (if it is the only content of the utterance). Following 2. in the automatic annotation rules, the question was assigned the WH tag in this case, while it should have been a PQ.

**Example 68 (Wh-question followed by a ‘no’, DinG4)**

<b>R</b> <sub>1148</sub>	<i>oui moi je prends</i>
<b>R</b> <sub>1148</sub>	“yes me I’ll take it”
<b>B</b> <sub>1149</sub>	<i>non mais</i>
<b>B</b> <sub>1149</sub>	“no but”
<b>W</b> <sub>1150</sub>	<i>qui veut une pierre contre un mouton? [rire]</i>
<b>W</b> <sub>1150</sub>	“who wants a rock for a sheep? [laugh]”
<b>R</b> <sub>1151</sub>	<i>non mais vraiment</i>
<b>R</b> <sub>1151</sub>	“no but really”
<b>R</b> <sub>1152</sub>	<i>oui oui oui je euh je prends totalement</i>
<b>R</b> <sub>1152</sub>	“yes yes yes I uh I’ll totally take [it]”

A way to interpret example 68 is that **R** started the utterance **R**<sub>1151</sub> without listening to **W**<sub>1150</sub>, to answer **B**<sub>1149</sub>. This hypothesis is supported by the content of **R**<sub>1152</sub>, which is an answer to **W**<sub>1150</sub>. It is important to note that in the current chapter, we stay agnostic as to ways to take multi-thread conversations into account in dialogue annotation, leaving this up to the annotator’s insight.

### 4.3.2 Human annotation

The annotation process was conducted by 10 people<sup>101</sup>, among which 3 did the full annotation and 7 annotated subparts of the questions corpus. All the annotators but 2 are native speakers of French. Annotating the whole corpus took 6 hours to one human annotator, which is why most of the annotators only went through part of the task.

The annotations were performed using spreadsheets, see figures 4.5 and 4.6. All the questions are automatically retrieved from transcriptions of DinG and displayed within a 5-utterance window: the utterance containing the question mark is surrounded by the previous two utterances and the next two utterances. Part of the questions were automatically pre-annotated, through the process presented in section 4.3.1. Figures 4.5 and 4.6 show automatically annotated examples: the human annotators worked with spreadsheets that already contained ‘1’ in the columns corresponding to the automatic annotation. They could leave the ‘1’ in the column where it was when they agreed with the automatic annotation or move it to another column when disagreeing.

File	ID	Question	YN	WH	CS	DQ	PQ	N/A
ding3-1.txt.ufo	961	R:6						
ding3-1.txt.ufo	962	R:c'était limite hein						
ding3-1.txt.ufo	963	Y:combien j'en ai ?		1				
ding3-1.txt.ufo	964	Y:4						
ding3-1.txt.ufo	965	Y:hum						

Figure 4.5: Screen-shot of the spreadsheet used to annotate questions from DinG with an automatic annotation of the central utterance as a *wh*-question.

#### Example 69 (English translation of the example presented in figure 4.5)

R 6

R it was borderline eh

Y how many do I have?

Y 4

Y hum

In example 69, figure 4.5, **Yellow** is asking a *wh*-question introduced by the French *wh*-word « combien » (“how much”). The *wh*-word is identified by the automatic tagger, the ‘1’ is put in the WH column. The human annotator can validate this tagging by not changing it and proceeding to the next question.

<sup>101</sup>We thank greatly Maxime Amblard, Amandine Decker, Bruno Guillaume, Pierre Lefebvre, Chuyuan Li, Léo Mangel, Siyana Pavlova, Guy Perrier, and Valentin Richard for their time and their contribution to our work.



File	ID	Question	YN	WH	CS	DQ	PQ	N/A
ding3-1.txt.ufo	985	O:[dés]						
ding3-1.txt.ufo	986	W:[rire]						
ding3-1.txt.ufo	987	W:pourquoi c'est toujours comme ça ?		1				
ding3-1.txt.ufo	988	O:[dés]						
ding3-1.txt.ufo	989	R:10						

Figure 4.6: Screen-shot of the spreadsheet used to annotate questions from DinG with a (wrong) automatic annotation of the central utterance as a *wh*-question.

#### Example 70 (English translation of the example presented in figure 4.6)

O [dice]

W [laugh]

W why is it always like that?

O [dice]

R 10

In example 70, figure 4.6, **White** is asking a *wh*-question introduced by the French *wh*-word « pourquoi » (“why”). The *wh*-word is identified by the automatic tagger, the ‘1’ is put in the WH column. Yet, the human annotator can see, from the surrounding context, that **White**’s question, while *wh*- in its form, is actually phatic. **White** is in fact (fake?) complaining about something related to the game, most likely the result of the roll of the die. The human annotator needs to correct this tagging by moving the ‘1’ to the PQ column.

The human annotators encountered several difficulties throughout the annotation. Annotator 1 was the first to conduct the full annotation. Their experience led us to implement the following decisions in the next annotations:

- Annotate question-tags (« j’ai fini, c’est ça ? » / “I finished, isn’t it?”) as YN.
- Separate the « ou pas » (“or not”) question tag from the others and tag questions finishing by « ou pas » as disjunctive, because they contain an “or” (« ou »). This decision is arguable and could be replaced by a tagging of this type of questions as YN.
- We annotate as phatic, unless clear from context that it’s not, the following questions: short questions such as « quoi ? » (“what?”), « c’est bon ? » (“all good?”), rhetorical/theatrical questions such as « sérieusement ? » (“seriously?”), « encore ? » (“again?”), « pourquoi c’est encore le 7 ? » (“why is it again the 7?”).
- A question in speech turn  $n$  is tagged as a completion suggestion only if the speech turn  $n - 1$  is uttered by a speaker different from  $n$ ’s one. This comes from the fact that if a person pauses in the middle of a question, it can be split into two speech turns and create the false impression of a completion suggestion.



These decisions eased the work of the following annotators, however, they didn't solve all the difficulties. The following section presents some of the cases that raised difficulties during the annotation process.

### 4.3.3 Discussion of chosen examples

This section shows examples that were highlighted as problematic or difficult to tag by the annotators. We examine the different cases and list solutions that could be envisioned for future development and usage of the annotation schema.

#### Example 71 (Request, DinG6)

**B**<sub>1813</sub>      *eh (xxx) (0.2s) c'est bon pour moi*

**B**<sub>1813</sub>      *"uh (xxx) (0.2s) it's ok for me"*

**R**<sub>1814</sub>      *c'est à qui là ?*

**R**<sub>1814</sub>      *"whose turn is it now?"*

**Y**<sub>1815</sub>      *alors je voudrais du bois (0.1s) s'il vous plaît ?*

**Y**<sub>1815</sub>      *"so I would like some wood (0.1s) please?"*

**Y**<sub>1816</sub>      *donc il me faudrait*

**Y**<sub>1816</sub>      *"so I would need"*

**R**<sub>1817</sub>      *fais un 7 là pour eh changer un peu de place*

**R**<sub>1817</sub>      *"make a 7 there to uh move around a little"*

Example 71 shows a question that fits neither of the 5 tags of the annotation schema. The utterance to be annotated is **Y**<sub>1815</sub>; it has been transcribed as a question by the transcriber as the annotation rises at the end of the word "please", but in fact, it is not a question but a request. We left the treatment of imperatives altogether out of our work so far, and it should be the object of future developments. We can for example take [Lascarides and Asher, 2003] as a starting point.

#### Example 72 (N/A question, DinG5)

**W**<sub>1654</sub>      *comment tout arrive*

**W**<sub>1654</sub>      *"how everything arrives"*

**B**<sub>1655</sub>      *du bois*

**B**<sub>1655</sub>      *"wood"*

**Y**<sub>1656</sub>      *non mais à part du bois?*

**Y**<sub>1656</sub>      *"no but besides wood?"*

**R**<sub>1657</sub>      *il a combien de bois lui du coup?*

**R**<sub>1657</sub>      *"so how much wood does he have now?"*

**R**<sub>1658</sub>      *il en a plein là non?*

**R**<sub>1658</sub>      *"he has a lot of them there<sup>102</sup>, right?"*

Example 72 is one of the most common examples in DinG that we cannot assign a tag to because we don't have enough context from the surrounding utterances to interpret the meaning of the question or even its scope. The utterance to be annotated is  $\mathbf{Y}_{1656}$ ; it does not contain a *wh*-word nor an “or”, the following utterance does not contain either explicit agreement or disagreement, it doesn't complete the previous utterance and the context does not give us an explicit indication of the other speakers interpreting  $\mathbf{Y}_{1656}$  as phatic. The N/A tag was added specifically to account for this type of example.

**Example 73 (« ou » VS « où », DinG4)**

$\mathbf{B}_{2840}$  *je te rends ça*

$\mathbf{B}_{2840}$  “I give this back to you”

$\mathbf{W}_{2841}$  *ah exact*

$\mathbf{W}_{2841}$  “ah right”

$\mathbf{R}_{2842}$  *tu es à côté où ?*

$\mathbf{R}_{2842}$  “where close are you?/[are you close or?]”

$\mathbf{W}_{2843}$  *il tu dois avoir 2 cases de (xxx) fin tu dois avoir 2 (xxx) d'écart à chaque fois*

$\mathbf{W}_{2843}$  “he you must have 2 squares of (xxx) well you must have 2 (xxx) a gap [of 2] each time”

$\mathbf{B}_{2844}$  *(xxx)*

$\mathbf{B}_{2844}$  “(xxx)”

Example 73 showcases a difficulty that is inherent to French and in particular to the transcription of oral French. In French, « ou » (“or”) and « où » (“where”) are homophones and almost homographs: the pronunciation of the two is exactly the same, and the difference in spelling between the two resides in the usage of a grave accent, and it is not rare for it to be forgotten by native speakers when they write. Usually, it is easy to disambiguate between the two from context, but the case of  $\mathbf{R}_{2842}$  is precisely one where the disambiguation cannot be done by reading the transcription only. It is possible that listening to the recording could help us, because, in the case of “or”, R would be likely to let the end of the sentence drag a little longer, to convey the implicit alternative. Yet, as the transcriber explicitly gave the transcription with the grave accent, the interpretation they made seems to be the one of “where”. Keeping this idea in mind,  $\mathbf{R}_{2842}$  can be annotated as WH.

**Example 74 (Implicit *wh*-question, DinG2)**

$\mathbf{R}_{103}$  *BLEU et toi c'est comment ?*

$\mathbf{R}_{103}$  “BLUE and you [what is your name]?”

$\mathbf{Y}_{104}$  *JAUNE*

$\mathbf{Y}_{104}$  “YELLOW”

$\mathbf{R}_{105}$  *JAUNE moi c'est ROUGE et ?*

$\mathbf{R}_{105}$  “YELLOW mine is RED and?”

$\mathbf{W}_{106}$  *BLANC*

$\mathbf{W}_{106}$  “WHITE”

$\mathbf{R}_{107}$  *donc euh*

$\mathbf{R}_{107}$  “so hum”

Example 74 is interesting for us in that it shows a question that should clearly, from context, be annotated as WH, but that contains no *wh*-word. This excerpt takes place at the beginning of a recording and the players do not know each other. **Blue** states their name and asks **Yellow** for their. When **Yellow** answers, **Red** repeats **Yellow**'s name to get familiar with it, states theirs, and asks the fourth player for theirs in **R**<sub>105</sub>. In **R**<sub>105</sub>, the question that we are annotating is in fact completely implicit: the tag we assign is the one corresponding to the question “and you, what is your name?”. Because of this, we tag **R**<sub>105</sub> as WH; yet, without the context of the two previous and two next utterances, we wouldn't be able to do so.

**Example 75 (A negative yes/no- or disjunctive question, DinG1)**

<b>R</b> <sub>927</sub>	<i>j'en ai pas</i>
<b>R</b> <sub>927</sub>	“I don't have any”
<b>Y</b> <sub>928</sub>	<i>personne n'a du blé ?</i>
<b>Y</b> <sub>928</sub>	“no one has any wheat?”
<b>Y</b> <sub>929</sub>	<i>ni de caillou ?</i>
<b>Y</b> <sub>929</sub>	“nor stone?”
<b>R</b> <sub>930</sub>	<i>il est pas tombé depuis longtemps hein</i>
<b>R</b> <sub>930</sub>	“it has not fallen for a long time eh”
<b>W</b> <sub>931</sub>	<i>hum</i>
<b>W</b> <sub>931</sub>	“um”

Example 75 is one of the examples that caused problems to the annotators. Indeed, **Y**<sub>929</sub> on its own can be seen as negative yes/no-question: it is introduced by the negative particle « ni » (“nor”) and could be answered by « si » (a strong version of « oui » (“yes”) for interro-negatives) or « non » (“no”). However, when considering **Y**<sub>929</sub> within its context, in particular paired with **Y**<sub>928</sub>, it appears that it could also be seen as a disjunctive question. The subtlety lies in the fact that « personne n'a du blé ni de caillou ? » (“no one has any wheat nor stone?”) should be tagged YN as a statement that has to be confirmed or contradicted. Yet, the prosodic contour, given by the two interrogation points used by the annotator and the fact that they specifically transcribed two utterances instead of one, leaves the opportunity for other tags to be used.

**Example 76 (A disjunctive or *wh*-question, DinG2)**

<b>B</b> <sub>1398</sub>	<i>j'ai pas de pierre</i>
<b>B</b> <sub>1398</sub>	“I have no stone”
<b>R</b> <sub>1399</sub>	<i>ah bon [rire] (xxx)</i>
<b>R</b> <sub>1399</sub>	“oh really [laugh] (xxx)”
<b>W</b> <sub>1400</sub>	<i>attends j'ai pas compris tu veux des moutons ou tu veux quoi ?</i>
<b>W</b> <sub>1400</sub>	“wait I didn't understand you want sheep or you want what?”
<b>R</b> <sub>1401</sub>	<i>elle a des moutons</i>
<b>R</b> <sub>1401</sub>	“she has sheep”
<b>B</b> <sub>1402</sub>	<i>euh non j'aimerais du blé j'aimerais du blé</i>
<b>B</b> <sub>1402</sub>	“uh no I would like some wheat I would like some wheat”

Example 76 shows a question that is tagged as DQ according to the linguistic cue, as it contains the word « ou » (“or”). Yet, the second part of the question appears to be an embedded *wh*-question: « tu veux quoi ? », especially in oral French, can be translated both as “you want what?” (with a stress on the “what”, used as a pronoun) or as “what do you want?”. The interpretation of  $\mathbf{W}_{1400}$  is then that first, a disjunctive question was started, but as it unfolds,  $\mathbf{W}$  realized they didn’t know what the second alternative should be and added a *wh*-twist to replace the second alternative in the disjunctive question. As in our annotation, we favor annotation from the top-most part of the utterance, we tag  $\mathbf{W}_{1400}$  as DQ.

#### 4.3.4 Results of the human annotation

The annotation results obtained by the three annotators who worked on the whole corpus are presented in table 4.16. The inter-annotator agreement scores are shown in table 4.17. Annotator 1 was the first to annotate the whole corpus; after they turned in their annotation, the annotation guide was adjusted based on corpus observations we will come back to below. In particular, the definition of completion suggestions (CS) in the context of DinG was clarified. This explains the fact that Annotator 1 tagged no questions as CS, while Annotators 2 and 3 found a few.

	YN	WH	CS	DQ	PQ	N/A	Total
<b>Automatic</b>	494	262	-	16	-	-	772
<b>Annotator 1</b>	1,588	608	0	71	89	100	2,456
<b>Annotator 2</b>	1,389	602	17	115	364	26	2,513
<b>Annotator 3</b>	1,345	572	7	106	458	23	2,511
<b>Average amount</b>	1,441	594	8	97	304	50	2,493
<b>Average percentage</b>	57.78 %	23.82%	0.32 %	3.90 %	12.18%	1.99%	100%

The averages are calculated on the results of the three human annotations.

Table 4.16: Annotation results for DinG’s questions annotations.

It is also interesting to note that the annotators did not annotate the same amount of questions even though they were all working with the same corpus. Our hypothesis is that this results from the utterances that contain two interrogation points: even though the convention was to annotate each of the questions separately, some examples seemed to be open to interpretation. Example 77 shows an utterance transcribed with two interrogation marks.  $\mathbf{R}_{1074}$  was attributed two tags: a ‘1’ in the YN column, because the expected answer for the “and some clay?” part of the utterance is a ‘yes’ or a ‘no’; and a ‘1’ in the PQ column, as the second part of the utterance is used for metacommunication purpose, to put an emphasis on the first part or perhaps on the fact that the consent of the addressee matters.

#### Example 77 (Utterance transcribed with two ‘?’, DinG2)

$\mathbf{R}_{1072}$	<i>eh 1 mouton et autre chose ouais ouais</i>
$\mathbf{R}_{1072}$	<i>“uh 1 sheep and something else yeah yeah”</i>
$\mathbf{Y}_{1073}$	<i>1 mouton et autre chose oui</i>
$\mathbf{Y}_{1073}$	<i>“1 sheep and something else yes”</i>

<sup>102</sup>Or “now”, we do not have enough context for the translation.

<b>R</b> <sub>1074</sub>	<i>et de l'argile ? ça te va ?</i>
<b>R</b> <sub>1074</sub>	<i>"and some clay? is that okay for you?"</i>
<b>Y</b> <sub>1075</sub>	<i>ça me va parfaitement</i>
<b>Y</b> <sub>1075</sub>	<i>"it is perfectly okay for me"</i>
<b>R</b> <sub>1076</sub>	<i>voilà on va faire ça</i>
<b>R</b> <sub>1076</sub>	<i>"okay then we'll do that"</i>

In general, it seems that tagging the questions as polar or *wh*- is an easier task than assigning the CS, DQ, or PQ tags. CS and PQ are categories that correspond to the pragmatics of dialogue, they are highly open to interpretation. The case of disjunctive questions is more interesting: they constitute on average less than 4% of the corpus, and they are quite likely to be confused with yes/no questions because of a frequently encountered mixed form such as the one in example 78.

**Example 78 (Yes/no-question with an embedded disjunction, DinG1)**

<b>B</b> <sub>568</sub>	<i>ah mais en plus ah</i>
<b>B</b> <sub>568</sub>	<i>"ah but also ah"</i>
<b>R</b> <sub>569</sub>	<i>non moi je je</i>
<b>R</b> <sub>569</sub>	<i>"no I I I"</i>
<b>R</b> <sub>570</sub>	<i>j'achète du mouton quelqu'un veut du (1s) blé ou du bois?</i>
<b>R</b> <sub>570</sub>	<i>"I'm buying sheep does anyone want (1s) wheat or wood?"</i>
<b>B</b> <sub>571</sub>	<i>non</i>
<b>B</b> <sub>571</sub>	<i>"no"</i>

● YN ● WH ● CS ● DQ ● PQ ● N/A

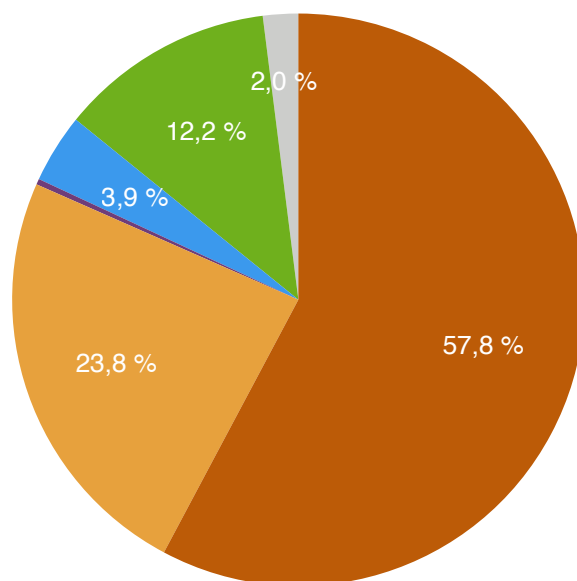


Figure 4.7: Distribution of the question annotation tags for DinG, on average, in percentage

$Y_{572}$	<i>du mouton?</i>
$Y_{572}$	<i>“some sheep?”</i>

In  $R_{570}$  in example 78,  $R$  asks a question that begins as a yes/no-question with a do-support (“does anyone want”), but continues with a disjunctive part: “wheat or wood?”. The decision was taken to follow the top-most form and annotate this type of question with the YN tag.

Table 4.17 presents the inter-annotator agreement scores for the three annotators that annotated all of the questions. As Cohen’s  $\kappa$  measures agreement between two annotators only, we also computed Fleiss  $\kappa$  for all three annotators. All the scores are quite high ( $\kappa > 0.61$ ), but it is particularly interesting to notice that the agreement between annotator 2 and annotator 3 is higher than 0.8. Annotators 2 and 3 performed the annotation after the aforementioned modifications of the annotation guidelines, inspired by annotator 1’s experience.

	Average Cohen $\kappa$	Fleiss $\kappa$
A1 + A2	0.651	-
A1 + A3	0.615	-
A2 + A3	<b>0.804</b>	-
A1 + A2 + A3	-	0.693

Table 4.17: Inter-annotator agreement scores for DinG’s questions annotations, where A1, A2 and A3 are the three annotators that annotated all of the questions.

Three annotators annotated only the first half of the questions. Partial annotator 2 annotated in parallel with annotator 1, so the adjustments in the annotation guide mentioned above took place after partial annotator 2 turned their annotation in. Table 4.18 presents the inter-annotator agreement scores for the three partial annotators. All the scores are high ( $\kappa > 0.77$ ), but it is particularly interesting to notice that the agreement between partial-annotator 1 and partial-annotator 3 is higher than 0.87. Partial-annotators 1 and 3 performed the annotation after the aforementioned modifications of the annotation guidelines, inspired by annotator 1’s experience, while partial-annotator 2 annotated with the same guidelines as annotator 1.

	Average Cohen $\kappa$	Fleiss $\kappa$
P1 + P2	0.797	-
P1 + P3	<b>0.872</b>	-
P2 + P3	0.771	-
P1 + P2 + P3	-	0.813

Table 4.18: Inter-annotator agreement scores for the first half of DinG’s questions annotations, where P1, P2 and P3 are the three annotators that annotated half of the questions.

The next step for the annotation experiments on DinG is to produce a golden annotation corpus. It will straightforwardly contain all the annotations for which the 6 annotators agree (from the first half of the questions). Then, we need to examine the annotations from the first half of the questions for which all annotators but A1 and P2 agree. In practice, we are waiting for more complete annotations before building and publishing the gold corpus.

## 4.4 Conclusion and future work

This chapter presents the annotation experiments we conducted on questions and answers in dialogue, using different annotation schema and different corpora in several languages. We started, in section 4.1, with a linguistic-inspired annotation schema which we applied to question and answers pairs corpora extracted from dialogues in English, French, northern Italian, and Mandarin Chinese. This annotation schema is composed of four parts: we annotate, separately, question forms, answer forms, question functions, and answer functions. The practical results of the annotations on the different corpora, produced by four annotators, show that some of the categories are both more relevant and easier to use for the annotators than others. In particular, the fact that the schema was very detailed for answer functions did not prove itself particularly useful or relevant for the cross-lingual annotations.

Next, we experimented with a simplified, data-inspired annotation schema in section 4.2. The schema contains only 5 tags for questions and 7 tags for answers. We applied this schema to question and answers pairs corpora extracted from dialogues in English, Spanish, and Dutch. The practical results of the annotations on the different corpora, produced by three annotators, show that the annotation of questions goes well, but the annotation of answers is problematic in that it depends on the annotation of questions, so any mistake made while annotating a question gets propagated and multiplied in answer annotation.

This brings us to experiments on the corpus of questions extracted from DinG in section 4.3. This corpus is annotated using the tagset for questions introduced in section 4.2, augmented with the tag N/A, used for questions that do not correspond to any of the other types. The annotation is partly automated, as the human annotators work with data that has been pre-tagged and have to either validate or correct the automatically assigned tags. This experiment gives us a complete analysis of the types of questions encountered in our corpus.

**Towards a comparison between STAC and DinG** The next step for our annotation experiments on DinG is to use the annotation schema provided by [Asher et al., 2016], designed for data from STAC, composed of chat logs of multilogues between people playing *Catan* online. This annotation schema aims for the identification of the discourse relations between dialogue acts. Dialogue acts in this schema are one of the following: *offer*, *counter-offer*, *acceptance*, *refusal*, *other*. It is important to note that all the dialogue acts, questions and answers and others alike, are classified as a part of this list. The authors of [Asher et al., 2016] mention the fact that STAC data contains a significant portion of “other”; our hypothesis is that for DinG, the results would be similar.

Then, the discourse structure annotation schema provided by [Asher et al., 2016] contains 16 categories for discourse relations, based on SDRT: `comment`, `clarification_question`, `elaboration`, `acknowledgment`, `continuation`, `explanation`, `conditional`, `question-answer_pair`, `alternation`, `Q-Elab`, `result`, `background`, `narration`, `correction`, `parallel`, `contrast`. Our idea is first to use the annotation schema on DinG, first exactly as it is presented in [Asher et al., 2016]. Then, we would like to integrate a detailed annotation of different types of questions (and answers) in the first layer of annotation (the dialogue acts) and see how it modifies the results on the second layer (the discourse relations). Finally, we would like to come up with a mixed schema for dialogue acts and dialogue relations that would acknowledge the complexity of dialogue data and question-answer mechanisms in particular. We hope that this mixed schema could be used in formal approaches such as the ones presented in the next chapter, Chapter 5.

## Part III

# Semantic processing of dialogue





# Semantic processing & dialogue

## Contents

---

<b>5.1</b>	<b>Negotiation phases</b>	<b>145</b>
<b>5.2</b>	<b>Relative response set</b>	<b>147</b>
5.2.1	Semantic representation of questions	148
5.2.2	The relevant response set	150
5.2.3	Answers & the effect on the relative response set	151
5.2.4	Model review and conclusions	154
<b>5.3</b>	<b>Semantic representation of <i>wh</i>-questions</b>	<b>155</b>
5.3.1	Syntax and semantics of <i>wh</i> -interrogatives	156
5.3.2	Controlling <i>wh</i> -extraction and quantifier raising	161
5.3.3	Discussion and shortcomings	163
5.3.4	Model review and conclusions	164
<b>5.4</b>	<b>Dynamic interactions and the dialogue context</b>	<b>164</b>
5.4.1	Dialogue gameboards	165
5.4.2	Formal settings	166
5.4.3	Dialogue participant's epistemic state	169
5.4.4	Dialogue dynamics and context updating	170
5.4.5	A complete example: Cheryl's birthday	173
5.4.6	Model review and conclusions	175
<b>5.5</b>	<b>Conclusion of the chapter</b>	<b>176</b>

---

All the preceding chapters of this thesis introduced various concepts, tools, and observations, that help us to pave the way towards a compositional, dynamic, and logic model of dialogue. The goal of this last chapter is to present the steps we took on this way, along with the difficulties we encountered, and future work directions.

The first chapter of this thesis introduces notions of formal semantics of discourse and dialogue. Chapter 2 presents more general ideas from dialogue processing and annotation. We showcase the quantity and extent of phenomena one can encounter when working with dialogue and discourse, even with toy examples, in Chapter 1. Then, in Chapter 3 we present a corpus of real-life data called DinG and show first data analysis and annotations in Chapter 4. The goal of the current chapter is to display our take on a formal semantic model of (real-life) dialogue.

**Note.** This chapter corresponds to the main developments obtained while working in close collaboration with Maxime Amblard and Philippe de Groote.

This chapter is shaped as an hourglass: we begin our discussion of dialogue by considering long interactions and splitting them into smaller pieces, called *negotiation phases*; inside those negotiation phases, we focus on the relationship between questions and answers and in particular between *wh*-questions and their answers. This, in turn, brings us to formal considerations on the semantic representation of *wh*-questions. Growing on the considerations we develop around *wh*-questions, we can widen back the scope of our study and model the interaction of utterances with the dialogue context.

Section 5.1 introduces the *negotiation phases* model of dialogue; we view the dialogue as the concatenation of a number of smaller parts, that we call negotiation phases, each organized around a particular topic and using a common dialogue context to grow information and store it. Inside the negotiation phases, we focus on the way questions and answers interact. To analyse part of these interactions, we introduce an imaginary (non-linguistic) tool called the *relative response set* in section 5.2. Intuitively, it corresponds to the semantic glue that links non-complete or vague answers to the *wh*-question they attempt to solve.

Then, we concentrate on the formal foundations of our model, in order to build a syntax-semantics interface for *wh*-questions and grow our model of dialogue on this basis. This task is the object of section 5.3. Next, we expand our focus to come back to negotiation phases. Section 5.4 introduces a formal model of the interactions between utterances that compose the negotiation phases and the dialogue context.

The first challenge encountered when working with real-life dialogues such as the ones from DinG is that the data is composed of long interactions: the average recording from DinG lasts  $\sim 70$  minutes, which roughly corresponds to 2,358 speech turns (see Chapter 3, section 3.3.2). This brings us to the idea of subdividing the interactions into smaller pieces, that we call negotiation phases.

Inside negotiation phases, one can find particular phenomena such as sequences of questions, simultaneous presence of a piece of information and its negation, along with usage of Non-Sentential Utterances [Fernández et al., 2007]. Beyond mechanical aspects of dialogue, one needs to consider meta-levels: links with argumentation (argument-mining, topoi studies), see for example [Breitholtz, 2014], along with simultaneous dialogues management [Asher et al., 2016]. We focus in particular on questions and answers and on the mechanisms that allow them to articulate together. First, we draw a distinction between the polar questions and the *wh*-questions. Then, we define the *relevant response set* with respect to a *wh*-question  $Q$ ,  $\mathcal{R}_Q$ . This set constitutes an artificial tool to materialize relevance phenomena that have no explicit linguistic realization and take place between answers and *wh*-questions.

Next, we shift back to a formal semantics perspective such as introduced in Chapter 1. In particular, we focus on a computational approach and develop a syntax-semantics interface for *wh*-questions. Section 5.3 presents a compositional method for logical computation of the semantic representations of *wh*-questions. We use compositional NDES as presented in [Champollion, 2017] joined with IS [Ciardelli et al., 2018]. Compositional NDES is introduced in Chapter 1, see in particular section 1.6. We implement our encoding using the Abstract Categorical Grammars (ACGs) framework [de Groote, 2001, Pogodalla, 2016], presented in Chapter 1, section 1.5.3.

Finally, we present a formal model of dialogue in a dynamic and epistemic setting, concentrating on the interactions between utterances (inside the negotiation phases) and the dialogue context. This model appeals to ideas derived from classical epistemic logic in order to model the

knowledge states of the dialogue participants [Van Ditmarsch et al., 2007], and includes context-updating mechanisms based on Continuation Style Dynamic Semantics (CSDS) [de Groot, 2006, Lebedeva, 2012]. CSDS is introduced in Chapter 1, see in particular section 1.5.

We conclude this chapter with an extended discussion of different parts of our model. We put our work in perspective of existing studies and observe the phenomena our model currently accounts for. We establish a list of forthcoming benchmarks and objectives our future work should be directed towards, and detail the path we will take to get there.

## 5.1 Negotiation phases

When working on long dialogue transcriptions, we stumble upon computational limitations. How far away can two linked dialogue acts be? How do we logically account for a piece of information that has been shared and then followed by its negation?

We came up with a computational vision of a dialogue by considering it as a combination of sequences of utterances and a dynamic context, constantly updated by new utterances. The information that we exchange is not exchanged in a continuous way, speech acts occur one after another, and information is discussed before an agreement is reached. This agreement phase can be void (when one participant directly accepts the other participant’s input in the conversation) but can also take several speech turns. During the discussion, each participant can use data that has previously been discussed and agreed upon in the history of this particular discussion. We call this part of dialogue the *negotiation phase*. This view of dialogue as composed of a concatenation of phases can in fact be related to the ideas of *transactions* introduced in Roulet’s model of conversation [Roulet et al., 1987].

We chose to separate context modeling from negotiation phase computation as we aim to proceed step by step in order to have full control over the computation process. We focus here on a view of the context of a conversation as a common (storage) ground, where information accessible to *both* participants of the dialogue appears.

Consider a dialogue involving two participants **A** and **B** (see figure 5.1). If **A** utters something, **B** can argue for or against it. In order to build her argumentation, **B** can both use information coming from **A**’s original utterance along with information coming from the dialogue context (see dashed arrows in figure 5.1). **A** can then reply, building up on **B**’s utterance but also eventually using the dialogue context (dashed arrows as well). When **A** and **B** reach an agreement (which can also be about dropping the original subject), the negotiation phase ends. We can then compute its semantic representation and store it in the dialogue context for further use in the dialogue (see dotted arrows in figure 5.1). Example 79 shows a negotiation phase encountered in DinG.

### Example 79 (Negotiation Phase, DinG6)

**Red**<sub>1</sub>        *un bois contre un argile*

**Red**<sub>1</sub>        “a wood for a clay”

**Yellow**<sub>2</sub>    *oui mais on sait pas ce que tu veux et ce que tu veux donner*

**Yellow**<sub>2</sub>    “yes but we don’t know what you want and what you want to give”

**Red**<sub>3</sub>        *ben je viens de le dire*

**Red**<sub>3</sub>        “well I just said it”

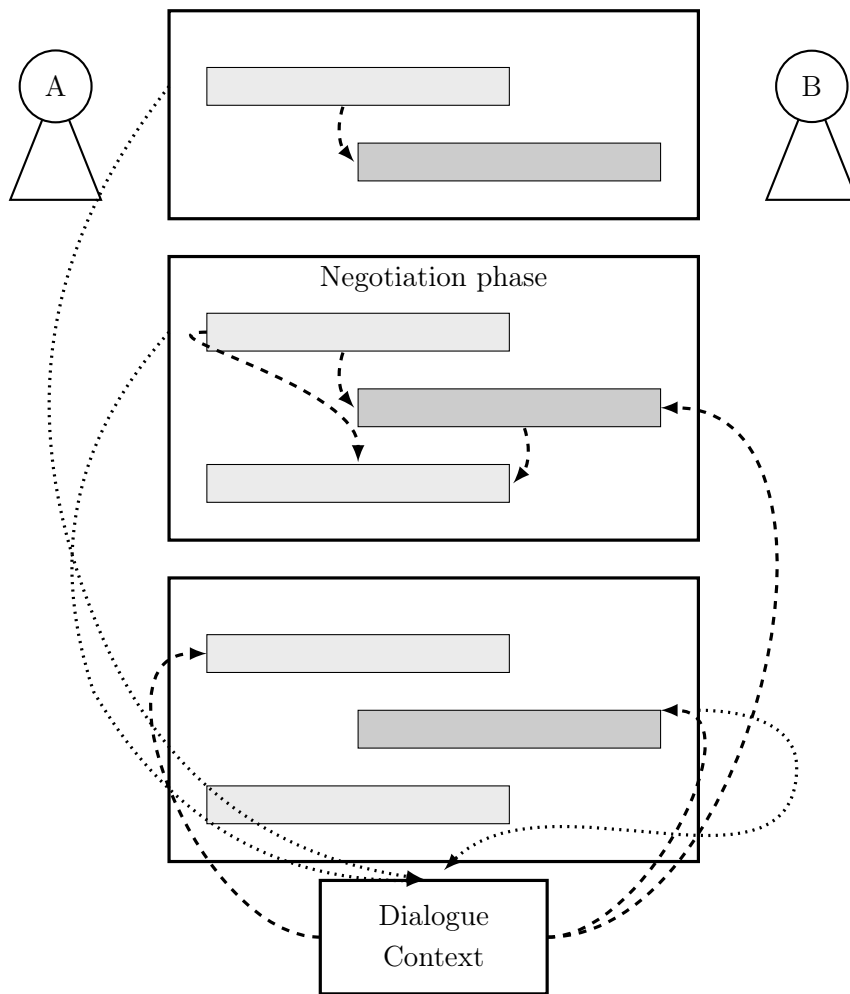


Figure 5.1: Subdivision of dialogue in negotiation phases.

**Yellow<sub>4</sub>** *un bois contre un argile tu veux donner un bois contre un argile ou tu veux recevoir un bois contre un argile*

**Yellow<sub>4</sub>** *“a wood for a clay you want to give a wood for a clay or you want to get a wood for a clay”*

**Red<sub>5</sub>** *ah je veux recevoir de l’argile contre un bois*

**Red<sub>5</sub>** *“oh I want to get some clay for a wood”*

**Red<sub>6</sub>** *je donne un bois et je veux bien récupérer un argile*

**Red<sub>6</sub>** *“I give a wood and I would like to get a clay”*

The negotiation phase in example 79 unravels as follows: **Red** utters **Red<sub>1</sub>**. **Yellow** does not directly accept **Red<sub>1</sub>**. Instead, **Yellow** makes an assertion about **Red<sub>1</sub>** that happens to be a clarification request, in order for **Red** to specify her utterance. Turns 3 to 5 are dedicated to a discussion about **Red<sub>1</sub>**, aiming to clarify it (**Red** doesn’t understand **Yellow**’s interpretation problem, **Yellow** makes it explicit, **Red** finally understands). **Red<sub>6</sub>** contains the result of this negotiation phase: an explicit and clarified version of **Red<sub>1</sub>**.

We can also imagine a minimal negotiation phase, such as the one presented in example 80:

**Example 80 (Minimal negotiation phase, fabricated)** .

**beginning of the negotiation phase**

**B<sub>1</sub>** *I get a clay.*

**end of the negotiation phase**

Let us consider the effect of **B** uttering **B<sub>1</sub>** on the dialogue context. Depending on the amount of information we want to keep, we will store in the context either only the semantic representation of **B<sub>1</sub>** (denoted  $\llbracket B_1 \rrbracket$ ), or  $\llbracket B_1 \rrbracket$  along with the information that it was uttered by **B**. We store  $\llbracket B_1 \rrbracket$  directly without waiting for it to be negotiated because we know that this negotiation phase is minimal: the discussion of this subject end there, and if it is brought up again in the following conversation (ex: “Didn’t you say that you got a clay, before?”), it will have to be introduced by reference words such as “before”, triggering some operators that will fetch  $\llbracket B_1 \rrbracket$  along with whatever linked information from the context, and modify  $\llbracket B_1 \rrbracket$  as needed.

In the following, section 5.2 and 5.3 focus on the modeling of phenomena happening inside the negotiation phases, in particular between (*wh*-)questions and answers, while section 5.4 presents a formal semantics model of interactions between the negotiation phases and the context.

## 5.2 Relative response set

**Note.** The following section grows on ideas introduced in [Boritchev and Amblard, 2020], presenting the main contents of the article along with our further thoughts and developments.<sup>103</sup>

As we have shown in Chapter 1, sentences can be represented in a logical, compositional and dynamic way. We rely on these notions to grow and define the representation of utterances and the links between them, in order to materialize the implicit mechanisms happening in dialogue, in particular for notions of compatibility between two successive utterances. Our goal is to produce semantic representations of utterances. Several strategies can be used to achieve this; our approach is part of a compositional vision of treatments. [Boritchev and Amblard, 2019] offers a process to compute compositional representations of questions by using NDES as presented in [Champollion, 2017].

To each declarative sentence  $S$  we associate a first order logic NDES formula  $\llbracket S \rrbracket$ , following [Champollion, 2015]. The translation process, well-defined for declarative sentences, allows us to compute the semantic representation of utterances that syntactically behave as declarative sentences: part of the assertions, part of the answers. The object of the following pages is the computation of semantic representations of interrogative sentences.

**Definition 32 (Simple normal form of NDES FOL formulae)**

*A FOL NDES formula is in a simple normal form if it is in prenex normal form (i.e written as a string of quantifiers and bound variables, followed by a quantifier-free part), where the quantifier-free part contains only conjunctions and negations.*

For the sake of simplicity, we will only focus in this section on NDES formulae in a simple normal form, presented in definition 32.

<sup>103</sup>We thank Philippe de Groote and Clément Beysson-Cabaret for the numerous discussions that resulted in the current version of this section.

### 5.2.1 Semantic representation of questions

Chapter 2 of this thesis features an extensive discussion of the typology of questions, see in particular section 2.3.1. The following section relies on the definitions given in Chapter 2 to develop the semantic representation of questions according to their type.

A polar, or *yes/no*-question requests confirmation or denial of its informational content. Therefore, it is possible to see it as a declarative sentence (with the same content) that will be either directly accepted, or negated and then accepted, depending on whether the answer is “Yes” or “No”. We denote the declarative content of a polar question  $Q_p$  as  $decl(Q_p)$ .  $decl(Q_p)$  is obtained from  $Q_p$  by resolving the deictic elements of  $Q_p$  (such as “we”, “you”, “here”, “there”, “now”, etc.) and then stripping  $Q_p$  from its explicit questioning markings (such as inversions).

#### Example 81 (Declarative content of a polar question)

Consider the following polar question **Blue**<sub>1032</sub>, encountered in *DinG6*:

**Blue**<sub>1032</sub> *Quelqu’un veut m’échanger de l’argile ?*

**Blue**<sub>1032</sub> “Does someone want to trade clay with me?”

The declarative content of **Blue**<sub>1032</sub>, denoted  $decl(\mathbf{Blue}_{1032})$ , is obtained first by resolving the deictic elements:

(6) *Does someone want to trade clay with **Blue**?*

Then, we reverse the questioning *do*-inversion:

(7) *Someone wants to trade clay with **Blue***

Which is the declarative content of the polar question **Blue**<sub>1032</sub>.

Now we can define semantic representations of polar questions in a way that is very similar to what we already know how to do for declarative sentences. The semantic representation of a polar question is the semantic representation of its declarative content.

#### Example 82 (Semantic representation of a polar question)

Following example 81, we consider the question **Blue**<sub>1032</sub> (*DinG6*):

**Blue**<sub>1032</sub> *Quelqu’un veut m’échanger de l’argile ?*

**Blue**<sub>1032</sub> “Does someone want to trade clay with me?”

The semantic representation of **Blue**<sub>1032</sub> is the semantic representation of  $decl(\mathbf{Blue}_{1032})$ :

$$\begin{aligned} \llbracket \mathbf{Blue}_{1032} \rrbracket &= \llbracket \text{Does someone want to trade clay with me?} \rrbracket \\ &= \llbracket decl(\mathbf{Blue}_{1032}) \rrbracket \\ &= \llbracket \text{Someone wants to trade clay with } \mathbf{Blue} \rrbracket \\ &= \exists e. \exists x. \exists y. trade(e) \wedge Ag_1(e, x) \wedge Th(e, y) \wedge clay(y) \wedge Ag_2(e, \mathbf{Blue}) \end{aligned}$$

In example 82, we identify a trading event that is built upon two agents, **Blue** and an unknown agent (the one we’re looking for), and a theme, the clay. The trading event is represented with the predicate  $\mathbf{trade}$ <sup>104</sup>. The formula is in simple normal form: it is in prenex form, as all the

<sup>104</sup>See [Babonnaud, 2019] for a discussion about choices of predicate ontologies.

quantifiers are grouped at the beginning of the formula, and the quantifier-free part contains only conjunctions.

Of course, in order to operationalize the computation of the semantic representation of polar questions, a few elements are missing. To be able to compute the declarative content of a polar question, we need a tool that resolves deictics and another that identifies and removes the questioning markings. To solve the deictics, we also need access to the dialogue context. These considerations are out of the scope of the current section, as we focus here on the semantics of links between questions and answers, inside negotiation phases. The next sections include discussions of the dialogue context.

We have now settled on a way to compute the semantic representation of declarative sentences and of polar questions. Both types of computations result in FOL NDES formulae in simple normal form. Our idea for the computation of semantic representation of *wh*-questions is that we see *wh*-questions as declarative sentences with missing content, corresponding to the part that is being interrogated. The part that is being interrogated is identified by the *wh*-word used in the *wh*-question. Chapter 2, section 2.6 gives us a correspondence between *wh*-words and thematic roles, thus we can identify the part that is being interrogated through the corresponding thematic role.

In NDES terms, this means that the semantic representation of a *wh*-question is a FOL NDES formula in simple normal form such that the quantifier-free part of the formula contains a predicate matching the thematic role corresponding to the missing content, and the string of quantifiers is prefixed by a  $\lambda$ -abstraction on a variable that is the second argument of the interrogated thematic role predicate. Similarly to what we do to compute the semantic representation of polar questions, we resolve the deictic elements before computing the semantic representation of a *wh*-question.

### Example 83 (Semantic representation of a *wh*-question)

Consider the following *wh*-question **Yellow**<sub>1001</sub>, encountered in *DinG7*.

**Yellow**<sub>1001</sub> *alors où est-ce que je fais une ville ?*

**Yellow**<sub>1001</sub> “so where do I make a city?”

First, we solve the deictic element « je » (“I”):

(8) So where does **Yellow** make a city?

The *wh*-word in **Yellow**<sub>1001</sub> is “where”, corresponding to the thematic role *Location* (*Loc*). The semantic representation of **Yellow**<sub>1001</sub> is the following FOL NDES formula in simple normal form<sup>105</sup>:

$$\llbracket \mathbf{Yellow}_{1001} \rrbracket = \lambda x. \exists e. \exists y. \text{make}(e) \wedge \text{Ag}(e, \mathbf{Yellow}) \wedge \text{Th}(e, y) \wedge \text{city}(y) \wedge \text{Loc}(e, x)$$

In example 83, we identify a making event that is built upon an agent, **Yellow**, a theme (the city), and an unknown location. The  $\lambda$ -abstraction binds the variable corresponding to the unknown location. In general, the  $\lambda$ -abstraction in the semantic representation of a *wh*-question accounts for the querying part of the *wh*-question, as opposed to the other, presupposed, semantic content of the question. In example 83, the presupposed semantic content is the fact of **Yellow** making a city.

Now that we have a method to compute the semantic representation of declarative sentences as well as polar and *wh*-questions, we can work on questions and answers in dialogue. The next section focuses on *wh*-questions and their answers, in particular through notions of answerhood.

<sup>105</sup>We do not take the phatic “so” into account for the semantic representation.



### 5.2.2 The relevant response set

We propose to introduce an intermediate level of description between *wh*-questions and answers. In human-human dialogue, answers can sometimes be easily and directly linked to the questions they should solve, but it is also often the case that an answer either solves its question partially, in a limited way or not at all but still stays within the scope of the subject raised by the question. Intuitively, our idea is to materialize the semantic glue that links answers to the corresponding *wh*-questions, even when the answers are partial, limited, borderline.

To define what constitutes an answer to a *wh*-question  $Q$ , we introduce the *relevant response set*, denoted  $\mathcal{R}_Q$ . It is the set of entities that match the constraints introduced by the predicates in the semantic representation of the question  $Q$ . In order to define  $\mathcal{R}_Q$  formally, we first need to define the *co-predication of variables*, a syntactic notion for FOL.

#### Definition 33 (Co-predicated variables)

Consider a FOL formula  $F$ , a binary predicate  $p$  and two variables  $x$  and  $y$ .  $x$  and  $y$  are co-predicated in  $F$  if  $p(x, y)$  or  $p(y, x)$  occur in  $F$ .

For example, consider the FOL formula  $H$ :

$$H = \forall x.\exists y.\forall z.p(x, y) \wedge q(y, z)$$

$x$  and  $y$  are co-predicated in  $H$ , and so are  $y$  and  $z$ , but  $x$  and  $z$  are not co-predicated, because there is no predicate that takes  $x$  and  $z$  as arguments, and the co-predication relation is not transitive. We use the co-predication of variables to define the syntactic construction of a question's relevant response set.

Consider a *wh*-question  $Q$ .  $Q$ 's *relevant response set*, denoted  $\mathcal{R}_Q$ , is syntactically built from  $\llbracket Q \rrbracket$  through the following process:

1. We identify the *central variable* of the question. The central variable is the one corresponding to the interrogated thematic role, i.e the one that bears the  $\lambda$ -abstraction.
2. Then, we strip  $\llbracket Q \rrbracket$  of everything but the quantifiers and predicates applied to the central variable and to all variables co-predicated with the central variable in  $\llbracket Q \rrbracket$ . We also keep all the predicates applied to a co-predicated variable and a constant.

#### Example 84 (Relative response set of $\mathbf{Yellow}_{1001}$ (example 83))

Example 83 gives us the semantic representation of  $\mathbf{Yellow}_{1001}$ :

$$\llbracket \mathbf{Yellow}_{1001} \rrbracket = \lambda x.\exists e.\exists y.make(e) \wedge Ag(e, \mathbf{Yellow}) \wedge Th(e, y) \wedge city(y) \wedge Loc(e, x)$$

The central variable of  $\mathbf{Yellow}_{1001}$  is  $x$ . To write the relative response set of  $\mathbf{Yellow}_{1001}$ , denoted  $\mathcal{R}_{\mathbf{Yellow}_{1001}}$ , we keep only the quantifiers and predicates applied to  $x$  and to all variables co-predicated with  $x$  in  $\llbracket \mathbf{Yellow}_{1001} \rrbracket$ . The variable co-predicated with  $x$  in  $\llbracket \mathbf{Yellow}_{1001} \rrbracket$  is  $e$ , through the *Loc* predicate.

Thus, we keep  $\lambda x$  as it concerns the central variable, we keep  $\exists e$  as  $e$  is co-predicated with  $x$  in  $\llbracket \mathbf{Yellow}_{1001} \rrbracket$ , but we drop  $\exists y$  because  $y$  is not co-predicated with  $x$  in  $\llbracket \mathbf{Yellow}_{1001} \rrbracket$ . Then, we keep  $make(e) \wedge Ag(e, \mathbf{Yellow}) \wedge Loc(e, x)$ , because  $\mathbf{Yellow}$  is a constant. We discard everything else. This gives us  $\mathcal{R}_{\mathbf{Yellow}_{1001}}$ :

$$\mathcal{R}_{\mathbf{Yellow}_{1001}} = \lambda x.\exists e.make(e) \wedge Ag(e, \mathbf{Yellow}) \wedge Loc(e, x)$$

$\mathcal{R}_{\text{Yellow}_{1001}}$  defines the set of all locations in which there exists a making event whose agent is Yellow. Our goal, in this section, is not to define an algorithm to build the representation, but to be able to infer  $\mathcal{R}_Q$  from the semantic representation of the question  $Q$  itself. The idea we develop in this section is that the question-answer relationship, for *wh*-questions, is built around  $\mathcal{R}_Q$ . Indeed, a *wh*-question’s semantic representation is a  $\lambda$ -abstraction of a FOL formula, of the general form  $\lambda x.P(x)$ , where  $P$  is a FOL proposition. The  $\lambda$ -term  $\lambda x.P(x)$  defines the set  $\{x|P(x)\}$ : the set of entities  $x$  such that  $P(x)$  is true. Thus, a *wh*-question’s semantic representation gives us the definition of a set, in the mathematical sense, of possible ways to fill the missing information. This set is what we call the relative response set  $\mathcal{R}_Q$  of the question. Next, our goal is to model different types of answers, among which the ones that do not entirely solve the question under discussion but rather give additional information that narrows the set of possible ways to fill the missing information, i.e  $\mathcal{R}_Q$ . Applying the semantic representation of this type of answers to  $\mathcal{R}_Q$  should therefore result in a subset of  $\mathcal{R}_Q$ .

### 5.2.3 Answers & the effect on the relative response set

Coming back to the typology of answers given in [Ginzburg, 2012] and presented in this thesis in Chapter 2, section 2.3.1, we want to give an overview of the effect of different types of answers on the relative response set. As said previously, following [Ginzburg, 2012], we distinguish 5 types of notions for the answers: (1) *simple answerhood*, (2) *aboutness*, (3) *strong exhaustive answer*, (4) *partially resolving answer*, and (5) *question/question relations* (answer to a question by a question), see table 5.1.

Category	Description
Simple Answer	Single-item utterance, concerning an instantiation of a question or its negation
Aboutness	Conditionals, modalities, disjunctions, quantifiers; gives additionnal info, doesn’t resolve the issue
Strongly exhaustive answer	Entails all the (true) simple answers
Potentially resolving answer	Either brings a simple answer or shows that there is no answer
Question	Answer that is an interrogation

Table 5.1: Taxonomy of answers, adapted from [Ginzburg, 2012].

Consider again the question **Yellow**<sub>1001</sub> (example 83): “so where do I make a city?”. Example 85 gives different fabricated answers to this question, one per type presented in table 5.1.

#### Example 85 (Different answers to **Yellow**<sub>1001</sub> (example 83), fabricated)

- (9) a. *in the woods*  
 b. ***Blue** will build a city in the woods*  
 c. *in the woods or in the field*  
 d. *if **Blue** builds a city in the woods, nowhere*  
 e. *when will you make a city?*

In example 85, (9a) instantiates the question **Yellow**<sub>1001</sub> by giving a possible location where **Yellow** can make a city, (9a) is a *simple answer*. (9b) gives additional information but doesn't resolve the issue, it is an answer *about Yellow*<sub>1001</sub>. (9c) is *strongly exhaustive* under the hypothesis (which we take here) that the two only possible locations for **Yellow** to make a city are the woods and the field; in this case, (9c) gives all the simple answers (among which (9a)). (9d) is a conditional that either brings a simple answer (if **Blue** doesn't build a city in the woods, the woods will be available for **Yellow** to make a city there!) or shows that there is no answer (through the consequence “nowhere”). Therefore, (9d) is a *potentially resolving answer*. (9e) illustrates the frequent type of answer that is another question, in this case, concerning another thematic role (**Temporality**).

The previous developments presented in this section lead us to the definition of the relative response set  $\mathcal{R}_Q$  for a question  $Q$ . Our goal in the following is to detail the effect of different types of answers on  $\mathcal{R}_Q$ , as a set.  $\mathcal{R}_Q$  is a set that is larger than the one defined by  $\llbracket Q \rrbracket$ , as it is obtained by loosening constraints on the elements of the set. Intuitively, the action of answers on  $\mathcal{R}_Q$  should result in a set  $\mathcal{R}$  obtained from  $\mathcal{R}_Q$  by adding constraints to the elements of  $\mathcal{R}_Q$ . These new constraints should both be defined by the semantic content of the answers and their type. More formally, we need a mechanism that allows dynamic changes in a  $\lambda$ -term while keeping track of a previous context. CSDS [de Groote, 2006, Lebedeva, 2012] provides such a mechanism.

If we have a mean of computation for  $\mathcal{R}$  from  $\mathcal{R}_Q$  and the semantic representation of an answer  $A$  ( $\llbracket A \rrbracket$ ), we have a model for a question-answer pair inside a negotiation phase. If this question-answer pair signs the end of the negotiation phase, we need to obtain a formula (a truth value) to store in the context. Therefore, we need an operation that would transform a  $\mathcal{R}$  into a formula. This type of operation should be a closure operator, very similar to the one used in NDES as presented in [Champollion, 2017] to close the scope of the existential quantifier on the event variable. Indeed, we need to close the  $\lambda$ -term to obtain a truth-conditional formula.

Now that we have listed the formal ingredients needed for the articulation of the relative response set model, we can discuss the different constructions of  $\mathcal{R}$  from  $\mathcal{R}_Q$  according to the type of considered answers. We illustrate the discussion on the sentences from examples 83 and 85.

**Simple answerhood** A simple answer instantiates the question (positively or negatively). The  $\mathcal{R}$  obtained from a simple answer should be a singleton, a set containing only the instantiation of the question given by the answer.

**Example 86 ( $\mathcal{R}$  corresponding to (9a), example 85)**

As (9a) is the simple answer “in the woods”, the corresponding  $\mathcal{R}$  is a singleton:

$$\mathcal{R}_{(9a)} = \{\text{woods}\}$$

**Aboutness** An answer about the question gives additional information but doesn't solve the question. Adding information amounts to adding constraints on the elements of  $\mathcal{R}_Q$ , so the resulting  $\mathcal{R}$  should be a subset of  $\mathcal{R}_Q$ .

**Example 87** ( $\mathcal{R}$  corresponding to (9b), example 85)

(9b) is “**Blue** will build a city in the woods”. The semantic representation of (9b) is the following<sup>106</sup>:

$$\exists e'. \exists z. \exists l. \text{build}(e') \wedge \text{Ag}(e', \mathbf{Blue}) \wedge \text{Th}(e', z) \wedge \text{Loc}(e', l) \wedge \text{city}(z) \wedge \text{woods}(l)$$

Example 84 gives us  $\mathcal{R}_{\mathbf{Yellow}_{1001}}$ :

$$\mathcal{R}_{\mathbf{Yellow}_{1001}} = \lambda x. \exists e. \text{make}(e) \wedge \text{Ag}(e, \mathbf{Yellow}) \wedge \text{Loc}(e, x)$$

$\mathcal{R}$  corresponding to (9b) is obtained from  $\mathcal{R}_{\mathbf{Yellow}_{1001}}$  by adding the constraints introduced by the semantic representation of (9b) under the scope of  $\lambda$  in  $\mathcal{R}_{\mathbf{Yellow}_{1001}}$ . Thus,

$$\begin{aligned} \mathcal{R}_{(9b)} = \lambda x. \exists e. \exists e'. \exists z. \exists l. & \text{make}(e) \wedge \text{Ag}(e, \mathbf{Yellow}) \wedge \text{Loc}(e, x) \\ & \wedge \text{build}(e') \wedge \text{Ag}(e', \mathbf{Blue}) \wedge \text{Th}(e', z) \wedge \text{Loc}(e', l) \wedge \text{city}(z) \wedge \text{woods}(l) \end{aligned}$$

**Strongly exhaustive answer** A strongly exhaustive answer entails all the true simple answers. The identification and computation of predicate entailment are outside the scope of this section, thus for the sake of simplicity, we consider that the  $\mathcal{R}$  obtained from a strongly exhaustive answer contains at least the union of all the  $\mathcal{R}$ 's corresponding to true simple answers.

**Example 88** ( $\mathcal{R}$  corresponding to (9c), example 85)

(9c) is “in the woods or in the field” and is a strongly exhaustive answer under the hypothesis that the two only possible locations for **Yellow** to make a city are the woods and the field. The corresponding  $\mathcal{R}$  is the union of the  $\mathcal{R}$ 's for the simple answers “in the woods” and “in the field”:

$$\mathcal{R}_{(9c)} = \{\text{woods}\} \cup \{\text{field}\} = \{\text{woods}, \text{field}\}$$

**Potentially resolving answer** A potentially resolving answer either brings a simple answer or shows that there is no answer. Thus, the  $\mathcal{R}$  obtained from a potentially resolving answer is either a singleton (if it brings a simple answer) or the empty set.

**Example 89** ( $\mathcal{R}$  corresponding to (9d), example 85)

In the case of (9d),  $\mathcal{R}_{(9d)}$  is the set that contains all the elements that have the property to be woods and to be true if and only if **Blue** doesn't build a city in the woods. Borrowing the representation of “**Blue** builds a city in the woods” from example 87, we have<sup>107</sup>:

$$\begin{aligned} \mathcal{R}_{(9d)} &= \lambda x. \text{woods}(x) \wedge \\ & \quad \neg(\exists e'. \exists z. \exists l. \text{build}(e') \wedge \text{Ag}(e', \mathbf{Blue}) \wedge \text{Th}(e', z) \wedge \text{Loc}(e', l) \wedge \text{city}(z) \wedge \text{woods}(l)) \\ &= \lambda x. \text{woods}(x) \wedge \\ & \quad (\forall e'. \forall z. \forall l. \neg \text{build}(e') \vee \neg \text{Ag}(e', \mathbf{Blue}) \vee \neg \text{Th}(e', z) \vee \neg \text{Loc}(e', l) \vee \neg \text{city}(z) \vee \neg \text{woods}(l)) \end{aligned}$$

Writing  $\mathcal{R}_{(9d)}$  in a prenex form, we obtain:

$$\begin{aligned} \mathcal{R}_{(9d)} &= \lambda x. \forall e''. \forall z'. \forall l'. \text{woods}(x) \wedge \\ & \quad (\neg \text{build}(e'') \vee \neg \text{Ag}(e'', \mathbf{Blue}) \vee \neg \text{Th}(e'', z') \vee \neg \text{Loc}(e'', l') \vee \neg \text{city}(z') \vee \neg \text{woods}(l')) \end{aligned}$$

<sup>106</sup>We leave out the representation of the future tense for the sake of simplicity here.

<sup>107</sup>See Chapter 1, section 1.2.2 for a discussion about negation for natural language semantics.

**Question answer** When the answer to the question  $Q$  is another question  $Q'$ , this answer yields a new relative response set  $\mathcal{R}_{Q'}$ , and  $Q'$  needs to be solved before a solution to  $Q$  can be found. The  $\mathcal{R}$  corresponding to the action of  $Q'$  on  $Q$  is thus  $\mathcal{R}_{Q'}$ .

**Example 90 ( $\mathcal{R}$  corresponding to (9e), example 85)**

(9e) is “when will you make a city?” directed towards **Yellow**. Following the relative response set construction method, we have:

$$\mathcal{R}_{(9e)} = \lambda x. \exists e. \text{make}(e) \wedge \text{Ag}(e, \mathbf{Yellow}) \wedge \text{Temp}(e, x)$$

Where *Temp* stands for the *Temporality* thematic role.

This overview of the different constructions of  $\mathcal{R}$  from  $\mathcal{R}_Q$  according to the type of considered answers leaves out one last type of answers: the ones expressing uncertainty or absence of knowledge. Our view of these answers in the relative response set model is that they correspond to dialogue movements rather than information-bringing utterances, meaning that their content is lying mostly outside the scope of a purely semantic approach such as the one we develop here.

#### 5.2.4 Model review and conclusions

This current point in this chapter and in this section seems to be the right place to take stock of what we now know how to model and compute and what stays left out. As stated at the beginning of this section, we have the computational means to compute the semantic representations of declarative sentences. This section contributed the computation process for polar and *wh*-questions.

As we know how to compute the semantic representations of declarative sentences, we know how to compute the semantic representations of answers that are fully articulated declarative sentences. The paper on which this section is loosely based, [Boritchev and Amblard, 2020], gives the semantic representations for “yes” and “no” to be  $\lambda P. P$  and  $\lambda P. \neg P$  respectively, following [Groenendijk et al., 1997]. Yet, this solution would not be adapted in the case of interro-negative questions.

The main contribution of this current section is the presentation of the *relative response set*, a theoretical tool that materializes the semantic glue between questions and answers in dialogue. We illustrate it using the typology of answers introduced in [Ginzburg, 2012]. The central idea behind the relative response set grows on the fairly classical view of questions as defined by their set of possible answers and extends this idea by enlarging this set of possible answers by loosening the constraints introduced by the question. This enlargement allows answers that are about the question but do not directly answer it to be taken into consideration in a way similar to all the other types of answers.

This last point gives us an opening to start the discussion about the universality of the approach. We started investigating the relative response set model aiming for a general approach to questions and answers in dialogue, in particular by keeping in mind that not all answers are simple ones. One of the major drawbacks we now see in to this model is that by chasing the universality on the answer side, we seem to have lost it on the question one – the relative response set model makes little to no sense for polar questions as we have presented them in this section. Furthermore, the relative response set model is relatively unsteady if we do not have access to a clear way of computing inferences on predicates, in particular on the ones applied to event variables. Consider the fabricated discussion shown in example 91.

**Example 91 (Illustration of the necessity for predicate inferences, fabricated)****A**<sub>1</sub> Does Charlie like tea?**B**<sub>2</sub> Charlie hates tea**B**<sub>2</sub> is a simple answer to **A**<sub>1</sub>, yet in order to see that, it is necessary to know that:

$$\forall e.(hate(e) \rightarrow \neg like(e)) \wedge (\neg like(e) \rightarrow hate(e))$$

Without this knowledge, **B**<sub>2</sub> would most likely be classified as an answer about **A**<sub>1</sub>, thus not resolving **A**<sub>1</sub>. This in turn would trigger a chain of computational complexification that is very likely to yield mistakes in the computation of the result of the negotiation phase. The wrong result would then be stored in the dialogue context, corrupting the computation for the whole dialogue.

Of course, the last paragraph presents a worst-case scenario, and most likely an operationalized version of the relative response set can find ways to bypass this type of problem. Yet, aiming for a universal treatment of questions in dialogue and pulling away from predicate-inference related troubles, we chose to stop our investigations of the relative response set here. Instead, we decided to dive deeper into the formal semantics approaches to questions, following the considerations developed at the beginning of the current section and adding insights from Inquisitive Semantics [Ciardelli et al., 2018]. The two next sections present the steps we took on that path.

### 5.3 Semantic representation of *wh*-questions

**Note.** This chapter corresponds to the main developments obtained while working in close collaboration with Maxime Amblard and Philippe de Groote.

Section 5.2 presented a model of interactions between *wh*-questions and their answers in dialogue, mainly driven by linguistic considerations such as answer typologies. The current section dives deeper into a formal semantics approach to *wh*-questions, aiming in particular for a syntax-semantics interface. To do so, we weave several formalisms and frameworks presented extensively in Chapter 1.

We present a compositional method for logically computing the semantic representations of *wh*-questions. We use compositional Neo-Davidsonian Event Semantics as presented in [Winter and Zwarts, 2011]<sup>108</sup>, with thematic roles, joined with Inquisitive Semantics [Ciardelli et al., 2018]. We implement our encoding using the Abstract Categorical Grammars framework [de Groote, 2001]. Our model results in an operationalized view of *wh*-questions and inquisitive existentials which we illustrate on several examples of *wh*-questions including *who*, *which* and *where*.

Throughout this thesis, we have extensively argued in favor of intermingling IS, NDES, and CSDS. The model we present grows on these three formalisms. We implement our encoding using the Abstract Categorical Grammars framework [de Groote, 2001].

Following classic literature on formal semantics of interrogative sentences such as [Hamblin, 1976], [Hintikka, 1983], [Groenendijk and Stokhof, 1984], we investigate questions through the semantic type of queries they raise. To do so, we use NDES supported by the set of thematic roles presented in table 5.2.

<sup>108</sup>This presentation is slightly different from the one given in [Champollion, 2017], but the intuitions stay very similar as both are presentation of Neo-Davidsonian Event Semantics with typed  $\lambda$ -calculus.



Thematic role	Definition
Participant	Entity involved in a state or event
Actor	Participant that is the instigator of an event
Cause	Actor (animate or inanimate) in an event, that initiates the event, without intentionality or consciousness, existing independently of the event
Agent	Actor in an event who initiates and carries out the event intentionally or consciously, existing independently of the event
Undergoer	Participant in a state or event that is not an instigator of the event or state
Instrument	Undergoer that is central to an event or state that is not an instigator of the event or state
Theme	Undergoer that is central to an event or state that does not have control over the way the event occurs, is not structurally changed by the event, and/or is characterized as being in a certain position or condition throughout the state
Pivot	Theme that participates in an event with another theme unequally but is central to the event
Patient	Undergoer in an event that experiences a change of state, location, or condition, that is causally involved or directly affected by other participants, and exists independently of the event
Attribute	Undergoer that is a property of an entity or entities, as opposed to the entity itself
Location	Place where an event occurs or a state is true

Table 5.2: Adapted from [Bonial et al., 2011], without thematic roles specific to events with symmetrical participants, events of perception, or events of communication, and with the addition of the `Location` role from [Bunt, 2009].

The main idea behind our construction is that it is possible to see a *wh*-extraction as an existential quantification on a variable inhabiting the thematic role corresponding to the *wh*-word used in the *wh*-question. To do so, we define a procedural calculus that produces such logical formulae. Our idea is that of a parallel between *wh*-extractions and quantification, as in both cases, the scope is central to the interpretation. We achieve a technical solution that presents a similar approach for *wh*-extractions and quantification through the use of IS, that translates the *wh*-extractions by existential quantifiers (see Chapter 2, section 2.4.1 for details).

### 5.3.1 Syntax and semantics of *wh*-interrogatives

In the categorial grammar tradition quantified noun phrases are distinguished from simple noun phrases at the syntactic level: noun phrases are assigned a simple atomic type,  $np$ , while quantified noun phrases are assigned the functional type  $(np \rightarrow s) \rightarrow s$ . This distinction reflects the fact that a quantified expression takes a scope, and allows for a smooth treatment of scope ambiguities.

Let us illustrate this approach by considering the following sentence:

- (10) *every farmer fed a donkey*

the two possible readings of which are captured by the two syntactic structures of figure 5.2, 5.3.1 and 5.3.1.

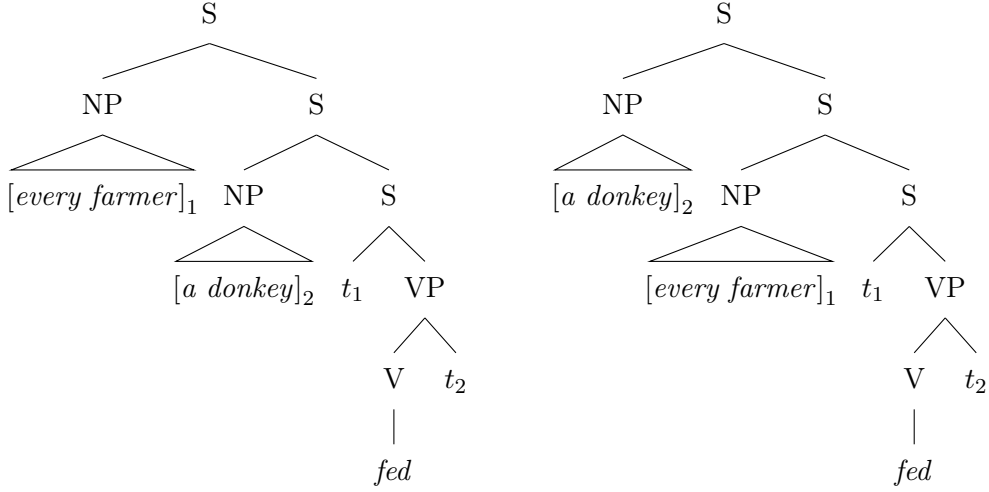


Figure 5.2: Syntactic structures of two possible readings for sentence (10).

In figure 5.2, as both noun phrases are quantified, they are raised above the syntactic structure of the original sentence; the memory of the places they used to have in the sentence is kept by the traces  $t_1$  and  $t_2$ , for  $[every\ farmer]_1$  and  $[a\ donkey]_2$ , respectively. The first part of this figure corresponds to a reading of sentence (10) where “every farmer” has a wider scope than “a donkey”. In the second part of figure 5.2, the scopes are reversed: “a donkey” has a wider scope than “every farmer”.

In order to give an abstract categorial account of sentence (10), it is possible to declare abstract constants of the following types:

$$\begin{aligned} \text{FARMER, DONKEY} &: n \\ \text{A, EVERY} &: n \rightarrow (np \rightarrow s) \rightarrow s \\ \text{FED} &: np \rightarrow np \rightarrow s \end{aligned}$$

Then, the syntactic structures presented in figure 5.2 are encoded by the  $\lambda$ -terms of figure 5.3. The @ symbols represent the application of one daughter of the tree to the other.

The categorial treatment of scope ambiguities presented here directly derives from [Montague, 1973]. It might be problematic when the targeted semantic formalism is Davidson’s event semantics. It has indeed been argued that Montague’s treatment of quantification does not combine smoothly with event semantics. The problem is that, in event semantics, a declarative sentence that is ultimately interpreted as a truth value ( $\mathbf{t}$ ) is first interpreted as a set of events ( $\mathbf{v} \rightarrow \mathbf{t}$ ). Then, switching from the latter interpretation of a sentence to the former necessitates an existential-closure operator, which may badly interact with the quantifiers that occur in the interpretation of the sentence. Fortunately, the literature provides at least two solutions to this problem. A first one is due to Champollion [Champollion, 2015], and a second one to Winter and Zwarts [Winter and Zwarts, 2011]. We follow this second solution since it is in line with the categorial tradition that we are advocating.

Winter and Zwarts’ solution consists in assigning two different syntactic types to the sentences. On the one hand, a first type ( $s_0$ ) is used for the “open” sentences, i.e., the sentences that are semantically interpreted as sets of events, and, on the other hand, a second syntactic type



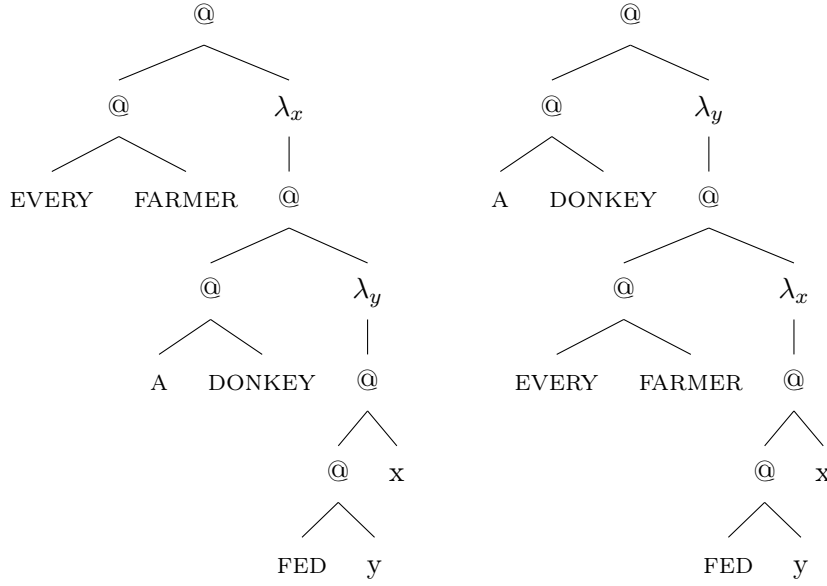


Figure 5.3:  $\lambda$ -terms corresponding to the syntactic structures presented in figure 5.2.

( $s$ ) is used for the sentences that are interpreted as truth value. Then, the existential closure operator allows values of type  $s_0$  to be coerced into values of type  $s$ . Accordingly, the abstract signature we have sketched above is transformed as follows:

$$\begin{aligned} \text{FARMER, DONKEY} &: n \\ \text{A, EVERY} &: n \rightarrow (np \rightarrow s) \rightarrow s \\ \text{FED} &: np \rightarrow np \rightarrow s_0 \\ \text{E-CLOS} &: s_0 \rightarrow s \end{aligned}$$

This ensures that the existential closure operator will always take a narrower scope with respect to the other quantifiers.

As mentioned above, there is a strong analogy between quantifier raising and *wh*-extraction. Following this analogy suggests that we should assign to a *wh*-noun phrase the type that we assign to quantified noun phrases. For example:

$$\text{WHO} : (np \rightarrow s) \rightarrow s$$

Similarly, a *wh*-determiner must be assigned the same type as a quantificational determiner (such as “a”):

$$\text{WHICH} : n \rightarrow (np \rightarrow s) \rightarrow s$$

Finally, *wh*-adverbs must be assigned the same type as the quantified adverbial modifiers, which are of type  $((s_0 \rightarrow s_0) \rightarrow s) \rightarrow s$  according to [de Groote and Winter, 2014]. Thus, we have:

$$\text{WHERE} : (s_0 \rightarrow s_0) \rightarrow s \rightarrow s$$

Putting everything together, we end up with an abstract syntax specified by the signature given in Table 5.3, where E-CLOS and Q are syntactically silent operators. The first one allows

Table 5.3: Abstract Syntax Signature

<i>Abstract Syntax</i>	
FARMER, DONKEY, MEADOW	$: n$
THE	$: n \rightarrow np$
A, SOME, EVERY, WHICH	$: n \rightarrow (np \rightarrow s) \rightarrow s$
IN	$: np \rightarrow s_0 \rightarrow s_0$
FED, DID-FEED	$: np \rightarrow np \rightarrow s_0$
WHO	$: (np \rightarrow s) \rightarrow s$
WHERE	$: ((s_0 \rightarrow s_0) \rightarrow s) \rightarrow s$
E-CLOS	$: s_0 \rightarrow s$
Q	$: s \rightarrow s$

an open sentence to be turned into a closed one (*existential closure operator*). The second allows a declarative proposition to be turned into a polar question (*questioning operator*).

In abstract categorial grammars, the language generated by an abstract signature (such as the one given in table 5.3) acts as a pivot language between surface forms and semantic interpretations (see Chapter 1, section 1.5.3 for the formal definitions). Figure 5.4 illustrates the way an abstract categorial grammar models the syntax-semantics interface.

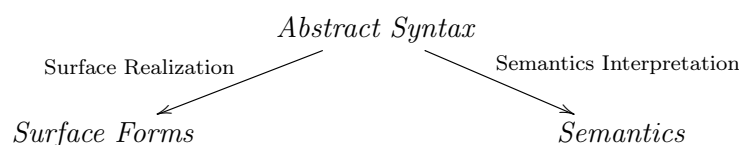


Figure 5.4: An abstract categorial grammar model of the the syntax-semantics interface.

Consequently, in order to complete the picture, we need to give the syntactic and the semantic translations of the abstract syntax specified by the signature of table 5.3. These are given in table 5.4 and table 5.5, respectively.

The surface realization lexicon, given in table 5.4, defines surface realizations of the constituents of well-formed sentences in English corresponding to the fragment that we define with our abstract categorial grammar. This lexicon gives the rules according to which the sentences are formed: for example, THE should be combined with another constituent (which as we know needs to be a noun). WHICH should be combined with two other constituents, a noun and a type-raised noun phrase, which will be applied to  $\epsilon$ , the empty word, to produce the expected result.

The semantic interpretation lexicon, given in table 5.5, defines the semantic interpretations corresponding to the elements of the abstract syntax. It gives the semantic interpretations of sentence constituents according to which their compositional combination yields the semantic representations of the sentences under consideration. For example, FARMER awaits for an argument and will result in the application of the predicate **farmer** to this argument. EVERY gives the classic semantic interpretation of the natural language universal quantifier, while A and SOME are slightly modified in order to fit within the IS formalism: as the existential introduces

Table 5.4: Surface Realization Lexicon

<i>Surface Realization</i>
FARMER := <b>farmer</b>
DONKEY := <b>donkey</b>
MEADOW := <b>meadow</b>
THE := $\lambda x. \mathbf{the} + x$
A := $\lambda xp. p(\mathbf{a} + x)$
SOME := $\lambda xp. p(\mathbf{some} + x)$
EVERY := $\lambda xp. p(\mathbf{every} + x)$
WHICH := $\lambda xy. \mathbf{which} + x + (y \epsilon)$
IN := $\lambda xy. y + \mathbf{in} + x$
FED := $\lambda xy. y + \mathbf{fed} + x$
DID-FEED := $\lambda xy. \mathbf{did} + y + \mathbf{feed} + x$
WHO := $\lambda x. \mathbf{who} + (x \epsilon)$
WHERE := $\lambda q. \mathbf{who} + (q(\lambda x. x))$
E-CLOS := $\lambda x. x$
Q := $\lambda x. x$

Table 5.5: Semantic Interpretation Lexicon

<i>Semantic Interpretation</i>
FARMER := $\lambda x. \mathbf{farmer} x$
DONKEY := $\lambda x. \mathbf{donkey} x$
MEADOW := $\lambda x. \mathbf{meadow} x$
THE := $\lambda p. \mathbf{the} (\lambda x. p x)$
A, SOME := $\lambda pq. !(\exists x. (p x) \wedge (q x))$
EVERY := $\lambda pq. \forall x. (p x) \rightarrow (q x)$
WHICH := $\lambda pq. \exists x. (p x) \wedge (q x)$
IN := $\lambda xp. \lambda e. (p e) \wedge (\mathbf{location} e x)$
FED, DID-FEED := $\lambda xy. \lambda e. (\mathbf{fed} e) \wedge (\mathbf{patient} e x) \wedge (\mathbf{agent} e y)$
WHO := $\lambda p. \exists x. p x$
WHERE := $\lambda p. \exists x. p (\lambda q. \lambda e. (q e) \wedge (\mathbf{location} e x))$
E-CLOS := $\lambda p. !(\exists e. p e)$
Q := $\lambda x. ?x$

inquisitiveness, it is necessary to apply the ! operator on top of the existential to keep only the informative part of the formula.

We can now illustrate the overall approach by treating the following example:

- (11) *where did every farmer feed a donkey*

Its two possible readings are captured by the two following  $\lambda$ -terms built upon the signature given in table 5.3:

$$(12) \quad \text{Q}(\text{WHERE}(\lambda f. \text{EVERY FARMER}(\lambda x. \text{A DONKEY}(\lambda y. \text{E-CLOS}(f(\text{DID-FEED } y x))))))$$

$$(13) \quad \text{Q}(\text{WHERE}(\lambda f. \text{A DONKEY}(\lambda x. \text{EVERY FARMER}(\lambda y. \text{E-CLOS}(f(\text{DID-FEED } x y))))))$$

Indeed, both (12) and (13) yield the same result when they are applied the translation given in table 5.4:

$$(14) \quad \textit{where} + \textit{did} + \textit{every} + \textit{farmer} + \textit{feed} + \textit{a} + \textit{donkey}$$

By contrast, when applied the translation given in table 5.5, (12) and (13) yield the two expected different readings: (5.1) and (5.2), respectively. (5.1) and (5.2) correspond, respectively, to the first and second readings given in figure 5.2.

$$\begin{aligned} ?\exists x. \forall y. (\textit{farmer } y) \rightarrow !((\exists z. (\textit{donkey } z) \wedge !((\exists e. (\textit{fed } e) \wedge (\textit{patient } e z) \\ (\textit{agent } e y) \wedge (\textit{location } e x)))))) \end{aligned} \quad (5.1)$$

$$\begin{aligned} ?\exists x. !((\exists y. (\textit{donkey } y) \wedge (\forall z. (\textit{farmer } z) \rightarrow !((\exists e. (\textit{fed } e) \wedge (\textit{patient } e y) \\ \wedge (\textit{agent } e z) \wedge (\textit{location } e x)))))) \end{aligned} \quad (5.2)$$

The fact that we obtain the same surface form but two different semantics is an encouraging indication for our model: it is fine-grained enough to capture the two different possible semantic interpretations, yet it is close enough to the linguistic reality to yield only one surface form.

### 5.3.2 Controlling *wh*-extraction and quantifier raising

The grammar we have sketched in the previous section is quite simple and has the advantage of highlighting the parallel that exists between declarative and interrogative sentences. In particular, it is based on a uniform treatment of quantification raising and *wh*-extraction. This simplicity, however, comes with drawbacks. These are threefold. Firstly, our grammar assigns the same syntactic categories to both the declarative and the interrogative forms (for instance, EVERY and WHICH are both assigned  $n \rightarrow (np \rightarrow s) \rightarrow s$ ). This gives rise to a grammar that generates ungrammatical surface forms such as the following one:

$$(15) \quad * \textit{every farmer fed which donkey in which meadow}$$

Secondly, allowing the quantifiers to take any possible scope results in fabricated (non-natural) ambiguities. For instance, a sentence such as:

$$(16) \quad \textit{a farmer fed a donkey in a meadow}$$

will give rise to six different abstract syntactic structures the semantic interpretations of which are all logically equivalent: all the quantifiers can take the higher scope, so there are three possibilities for the first one and two for the second one.

Finally, the interactions between *wh*-extraction and quantifier raising must obey some constraints. For instance, in a *wh*-question, the *wh*-quantifier must always take the wider scope.

Consider again example (11). The signature given in table 5.3 allows one to built syntactic structures, such as the (17), that do not respect the *wh*-quantifier wider scope constraint, as the wider scope is taken by the quantifier EVERY.

(17) EVERY FARMER ( $\lambda x$ . WHERE ( $\lambda f$ . A DONKEY ( $\lambda y$ . E-CLOS ( $f$  (DID-FEED  $y x$ ))))))

Consequently, our grammar might allow nonsensical semantic interpretations to be derived, as it is designed for syntactic structures in which the *wh*-quantifier takes the wider scope.

The three kinds of defects of our grammar all stem from the same fact: the signature of table 5.3 allows the derivation of too many abstract syntactic structures. Thus, we need to be able to select among the  $\lambda$ -terms that can be built upon the signature of table 5.3 the ones that correspond to legitimate abstract syntactic structures. A modular and efficient solution to this problem consists in using another abstract categorial grammar in order to *control* the *Abstract Syntax* signature, which corresponds to the grammatical architecture shown in Figure 5.5.

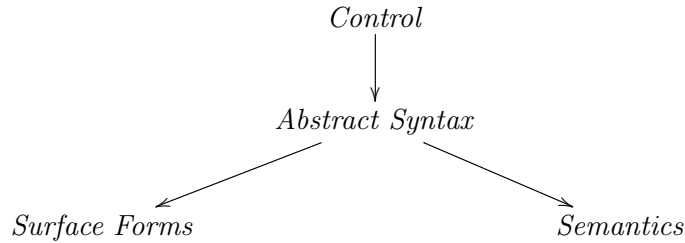


Figure 5.5: Grammatical architecture of our model.

We have developed such a control grammar, which we give *in extenso* in the annexes, B. Table 5.6 shows the excerpt that allows example (11) to unravel.

Table 5.6: Control Module

Control Signature	Control to Abstract Syntax Lexicon
$a := n \rightarrow npe$	$a := A$
$every := n \rightarrow npu$	$every := EVERY$
$donkey := n$	$donkey := DONKEY$
$farmer := n$	$farmer := FARMER$
$Inpe := npe \rightarrow np$	$Inpe := \lambda x. x$
$Inpu := npu \rightarrow np$	$Inpu := \lambda x. x$
$did-feed := np \rightarrow vpq$	$did-feed := \lambda p x f. p (\lambda y. f (DID-FEED y x))$
$did-feed_2 := vpq_{np}$	$did-feed_2 := \lambda x y f. f (DID-FEED x y)$
$where := sq_0 \rightarrow s$	$where := \lambda p. Q (WHERE (\lambda f. p (\lambda s. E-CLOS (f s))))$
$SQ4 := np \rightarrow vpq \rightarrow sq_0$	$SQ4 := \lambda p q f. p (\lambda x. q x f)$
$SQ14 := npu \rightarrow vpq_{np} \rightarrow sq_{0npe}$	$SQ14 := \lambda p q x f. p (\lambda y. q (x y) f)$
$QRq13 := npe \rightarrow sq_{0npe} \rightarrow sq_0$	$QRq13 := \lambda p q f. p (\lambda x. q x f)$

In the *Control Module*, *npe* and *npu* are, respectively, the type of existentially and universally quantified noun phrases. *Inpe* and *Inpu* are projector operators that transform existentially and universally quantified noun phrases into noun phrases. *vpq* stands for a verb phrase that is lacking a question, and *vpq<sub>np</sub>* specifies that the question is about the noun phrase. The interpretation with quantifiers *in situ* uses *SQ4*. Finally, *QRq13* is a quantifier raiser on a question: it is an

operator that takes an existentially quantified noun phrase and a non-closed sentence that is lacking a question that is about an existentially quantified noun phrase, built with `SQ14`, and finally yields a non-closed sentence that is lacking a question. The only operators in the *Control Module* that takes arguments of an *s* type are the ones for *wh*-words: in table 5.6, `where`. This ensures that a *wh*-word can only be applied as the last operation. Finally, the *Control Module* provides the answers to the drawbacks listed above: the types ensure that the *wh*-operators can only be applied once; the way the quantifier raising is done restrains the ambiguities.

### 5.3.3 Discussion and shortcomings

In the previous sections, we have presented a categorial formalisation of the syntax and semantics of *wh*-interrogatives. In its current state, our model doesn't give a direct solution for the representation of *wh*-questions that contain the *wh*-words “what”, “whose”, “why”, and “how”.

“what” appears to be close in behavior either to “who” and “whom”, or to “which”. The difference between “what” and “which” seems to come from pragmatic considerations: the interpretation of “what” hugely depends on the context in which this interrogative word is used, while “which” is restrained in its interpretation by the definition of the set from which the choice of the response is made. “what did the farmer do” is an example that we cannot account for in the current state of our model: the issue raised by this question concerns the action, therefore the event, in our Neo-Davidsonian view. In order to be able to model this example, we would need a quantification on the predicates, so a higher-order logic.

Regarding “how”, the difficulty is quite close to the one encountered with “what”: the meaning of the *wh*-word “how” depends on the expression “how” is paired with. Consider the difference between “how old” and “how far”: in the first case, the interrogated thematic role could be something time or age-related; in the second, the thematic role is `Location`.

“whose” raises the problem of modeling the possessive relation. The semantics of possessives is a subject on its own, see for example [Peters and Westerståhl, 2013] for investigations on the semantics of English possessives. In our case, modeling “whose” would require an extension of our grammar with the addition of the preposition “of” used as a possessive, therefore as a modifier of a noun. This is the solution we chose below, in section 5.4.

Finally, the difficulty with “why?” comes from its very large scope. Indeed, “why” can take a whole sentence as its argument (ex: “why did a farmer feed every donkey?”). In particular, it means that “why” introduces a quantification on the event predicate, which requires a higher-order logic treatment.

We can now compare our model to existing works by discussing related approaches, in particular related to the linguistic and mathematical tools that we use. The first point of comparison is extensively presented throughout this thesis: *Quantificational event semantics* gives an original take on Neo-Davidsonian event semantics. Quantificational event semantics is compositional and natively computational. It is operationalized through the Lambda Calculator<sup>109</sup> [Champollion et al., 2007], an interactive graphical application for semantic representation derivation computation following syntactic trees. Our method takes advantage of Champollion's compositional presentation of event semantics while following a different path in terms of sentence closure, to be able to account for interrogative sentences.

[Ciardelli et al., 2017] introduces typed Inquisitive Semantics, a compositional take on Inquisitive Semantics, that addresses quantification, negation, some *wh*-questions, notions of entailment and coordination. [Champollion et al., 2015] and [Champollion et al., 2017] use typed

<sup>109</sup><http://lambdacalculator.com/>, visited on 18/06/2021.

Inquisitive Semantics to model questions and presuppositions. Our approach can be viewed as very similar to a combination of typed Inquisitive Semantics and quantification event semantics, which, to our knowledge, had not yet been done.

Finally, as also extensively presented throughout this thesis, [de Groote, 2006, Lebedeva, 2012] presents a dynamic approach to discourse modeling. Based on Montague semantics, it is compositional and computational through the usage of simple type theory [Church, 1940] and  $\lambda$ -calculus. It is operationalized through the ACG toolkit [de Groote, 2001, Pogodalla, 2016]. Our model is directly in line with this work.

### 5.3.4 Model review and conclusions

In the current section, we present a compositional method for the computation of logical representations of *wh*-questions through the use of thematic roles, based on NDES and IS and implemented through the ACG framework. We undertake the proof that our approach is operational and computationally sound and showcase our method on several variations on classical examples.

The approach presented in this section aims at staying as general as possible with regard to language specificities. We focused the presentation on English because of the simplicity of a *wh*-words-based presentation of non-closed questions, but we aim to produce a system that would be easily extendable to other languages such as French and Russian, using for example insights from [Zamaraeva, 2021].

The grammar presented in this section and in the annexes is a toy grammar for a fragment of English. It can easily be extended, for example in order to include syntactic tags such as the ones from the Stanford parser<sup>110</sup> [Klein and Manning, 2003, Klein et al., 2003], currently available for 6 languages: Arabic, Chinese, English, French, German, and Spanish.

A next step for our work on this model is to add dynamicity. A way to achieve this goal is through the integration of CSDS [de Groote, 2006, Lebedeva, 2012] along with Inquisitive Dynamic Epistemic Logic [Ciardelli and Roelofsen, 2015]. To do so, we need to zoom out from the *wh*-questions and look at the slightly bigger picture of the dialogue, taking the dialogue context into account. The next section focuses on the dynamic aspects of the semantics of utterances in dialogue, in particular through their interactions with the dialogue context.

## 5.4 Dynamic interactions and the dialogue context

**Note.** The following section corresponds to the main developments presented in [Boritchev and de Groote, 2020].

This section introduces a formal model of dialogue based on insights and ideas from KoS, developed by Jonathan Ginzburg in [Ginzburg, 2012]. This model, which is logic-based, takes advantage of IS [Ciardelli et al., 2018], which allows the modeling of both declarative and interrogative sentences in a uniform way. It appeals to ideas derived from classical epistemic logic in order to model the knowledge states of the dialogue participants, and includes a context-updating mechanism based on CSDS [de Groote, 2006, Lebedeva, 2012].

As we have stated throughout this thesis, dialogues are built in a dynamic way. An utterance follows another and may contain references to concepts and language constructions introduced by previous utterances, but also by the context of the conversation. This dialogue context is

---

<sup>110</sup><https://stanfordnlp.github.io/CoreNLP/>

constantly updated as the dialogue unrolls, both for each dialogue participant, privately, and in a public way, building the *common ground* [Stalnaker, 2002]<sup>111</sup>, composed of information that is available to everyone equally (participants and possible audience). Consider the following piece of dialogue (part of the example we work with in section 5.4.5):

- (18) [CONTEXT: Albert and Bernard would like to know when is Cheryl’s birthday. She gives them some clues that might help them guess the date.]
- a. Cheryl (to Albert): *Can you figure it out now?*
  - b. Albert (to Cheryl): *I don’t know when your birthday is, but I know Bernard doesn’t know, either.*

This simple excerpt illustrates several features that are characteristic of dialogue. It stresses the importance of the context, and in particular, of keeping track of the issues that are being raised, of what questions are under discussion. This allows, for instance, the pronominal anaphoric “it” in 18a to be resolved. It also demonstrates that a dialogue is not only made of declarative sentences, but also of interrogative ones (direct, as in 18a, or embedded under a propositional attitude verb, as in 18b). Consequently, the development of a logic-based formal model of dialogue requires a logic that can express the semantic content of both declaratives and interrogatives. Inquisitive logic [Ciardelli et al., 2018] is such a logic, see Chapter 1, section 1.7 for a presentation.

Example 18 also demonstrates the need for a dialogue model to integrate epistemic modalities. In 18b, Albert mentions his (private) knowledge about Bernard’s (private) knowledge. Therefore, we need to model the knowledge states of the dialogue participants; see section 5.4.3. The whole dialogue is then modeled using a *dialogue gameboard*, see section 5.4.1 for a presentation of the gameboard and section 5.4.4 for the formal model. We showcase the way our model works on an example in English, *Cheryl’s birthday problem*, a logical puzzle that went viral on the internet of few years ago<sup>112</sup>, see section 5.4.5.

### 5.4.1 Dialogue gameboards

Dialogue semantics is radically context-dependent. Following [Ginzburg, 2012], we model the dialogues and dialogue context in particular using *dialogue gameboards* (DGB). We use one DGB per participant in order to model their private contexts, plus one DGB for the public context. A dialogue gameboard is composed of different fields:

$$\left[ \begin{array}{ll} \textit{speaker} : & \textit{Individual} \\ \textit{addressee} : & \textit{Individual} \\ \textit{FACTS} : & \textit{set of propositions} \\ \textit{QUD} : & \textit{partially ordered set of questions} \end{array} \right]$$

This is a simplified version of definition 13, given in Chapter 2, section 2.2.3. Several fields are used to store information about the indexicals, typically, the speaker and the addressee. *FACTS* is used to store the propositions that have been agreed on by the dialogue participants in the case of *public dialogue gameboards* (the ones that model a shared view of the dialogue), and propositions that are personal to the participant in the case of *private dialogue gameboards*.

<sup>111</sup>See Chapter 2, section 2.2.2 for a presentation of the common ground.

<sup>112</sup><https://www.nytimes.com/2015/04/15/science/a-math-problem-from-singapore-goes-viral-when-is-cheryls-birthday.html>



*QUD*, which stands for *questions under discussion*, stores the issues that have to be solved by the dialogue participants. These issues are raised by questions asked by the dialogue participants, but also by other types of utterances, as any proposition has to be discussed before being accepted by all the participants. The *QUD* is a partially ordered set where the order is used to decide which issue has to be solved first if several issues are raised at the same time.

A simple dialogue gameboard representation of 18b would then be the following:

$$\left[ \begin{array}{l} \textit{speaker} \\ \textit{addressee} \\ \textit{FACTS} \\ \textit{QUD} \end{array} \right. = \begin{array}{l} \textit{Albert} \\ \textit{Cheryl} \\ \{ \textit{Albert doesn't know when Cheryl's birthday is,} \\ \textit{Albert knows Bernard doesn't know either} \} \\ \{ \textit{When is Cheryl's birthday?} \} \end{array} \right.$$

Consider figure 5.6, which represents a dynamic view of a dialogue divided in *negotiation phases*. As presented in section 5.1, a negotiation phase corresponds to the discussion by the participants of one issue; it begins with the introduction of this issue and ends when an agreement has been reached (while this agreement can be to drop the issue, to disagree). The result of the negotiation phase is then stored in the dialogue context and can be referenced in the utterances build inside future negotiation phases. In terms of dialogue gameboards, a negotiation phase begins when a new question is added to the *QUD* and ends when it has been solved. In this section, we focus on modeling the dialogue interactions at the level of a negotiation phase, elaborating on the presentation given in section 5.1. We do not discuss the way negotiation phases articulate with one another.

Now that we have outlined our usage and adaptation of DGBs, we can move on to the next formal tool we include in our model. The next section shows an adaptation of Inquisitive Semantics [Ciardelli et al., 2018] in a compositional way, inspired by Montague semantics [Montague, 1970].

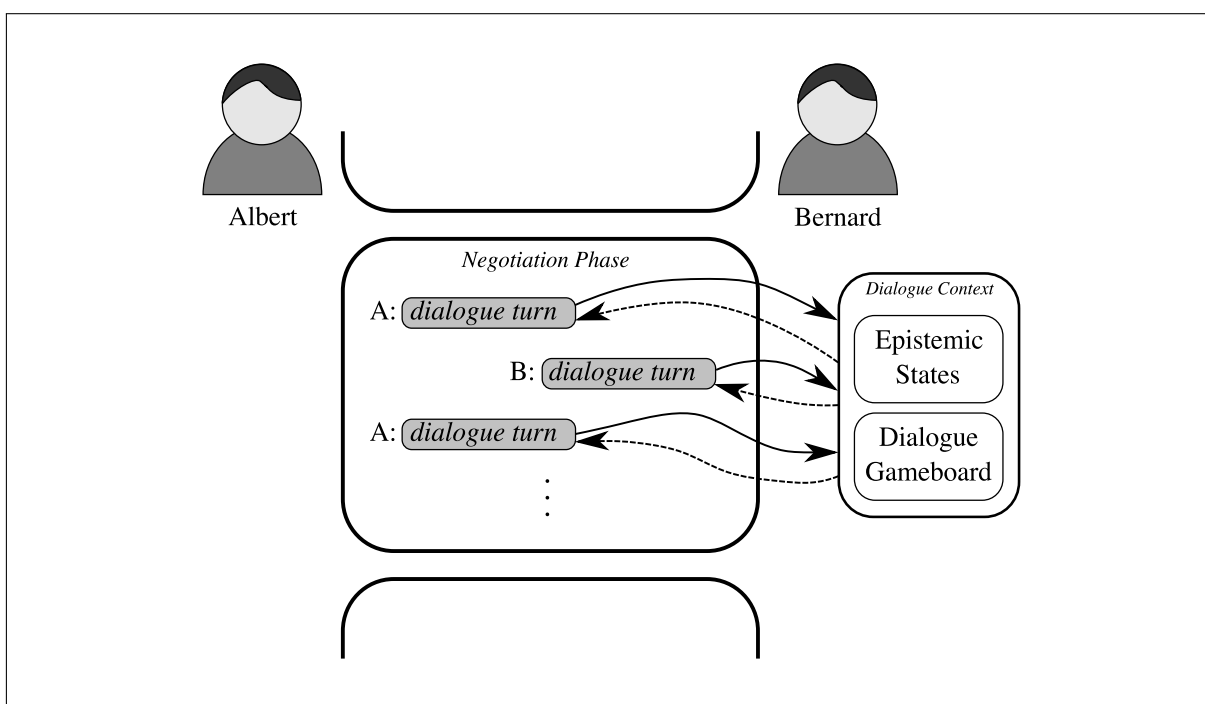
### 5.4.2 Formal settings

Inquisitive Semantics (IS) has extensively been discussed throughout this thesis: first, we presented it as a formal model of discourse, in Chapter 1, section 1.7; then, we illustrated IS's take on questions and in particular inclusive and exclusive disjunction in Chapter 2, section 2.4.1. The following section provides a Montague-like compositional semantics to questions through IS. To this end, we consider the question *when Cheryl's birthday is* (as it occurs in the sentence *Albert does not know when Cheryl birthday is*). The abstract syntax of the sentence is specified by means of the term:

$$\text{WHEN } (\lambda x^{NP}. \text{IS } x \text{ (POSSESSIVE CHERYL BIRTHDAY)}) \quad (5.4)$$

This term is read as follows: first, the possessive relation links Cheryl and her birthday through application. Then,  $x$  and Cheryl's birthday are linked through the application of "is", roughly corresponding to an equality between the two. As we actually do not know when Cheryl's birthday is and this date is the object of the query in the question *when Cheryl's birthday is*, we leave the  $x$  to be filled with the answer. To do so, we abstract the  $x$ , while specifying the type it is supposed to have ( $NP$ , for noun phrase<sup>113</sup>). Finally, the wh-word "when" closes the representation of the question.

<sup>113</sup>See Chapter 1, section 1.2 for a presentation of usual syntactic categories and types.



We use inquisitive semantics to model both interrogative and declarative sentences (see the **grey rectangles** around the utterances). **Dotted lines** represent dynamic phenomena between dialogue context and utterances, as dynamicity allows to reference previously stored information. **Full lines** represent the dialogue context's updates, both in terms of epistemic states and dialogue gameboard, as each new utterance brings new information about the participants' epistemic states, but also about the *FACTS* and the *QUD*.

Figure 5.6: Subdivision of dialogue in negotiation phases, adapted from [Boritchev and Amblard, 2021], [Boritchev and de Groote, 2020].

The term 5.4 is built upon the following signature:

$$\begin{aligned}
 \text{CHERYL} &: NP \\
 \text{BIRTHDAY} &: N \\
 \text{POSSESSIVE} &: NP \rightarrow N \rightarrow NP \\
 \text{IS} &: NP \rightarrow NP \rightarrow S \\
 \text{WHEN} &: (NP \rightarrow S) \rightarrow S
 \end{aligned}$$

CHERYL is a proper noun, its syntactic category is the one of a noun phrase, *NP*. BIRTHDAY is a noun, its syntactic category is *N*. The POSSESSIVE links a nouns phrase with a noun by a relation of possession and results in a noun phrase (such as *Cheryl's birthday* or *the unicorn's tea*). IS, as used in our example, corresponds to a relation of equality between two noun phrases and results in a sentence, as in *Cheryl's birthday is Wednesday*. If we were considering "is" in a sentence such as "The unicorn is green", the syntactic type of "is" would have been different, as it should acknowledge the characterization of the *NP*, instead of the equality between two *NPs*. Finally, WHEN is the wh-word that marks the query: in order to produce a proper sentence, it awaits a *NP*. As WHEN is also responsible for the closure of the sentence, an argument of the

type  $NP \rightarrow S$  is needed to produce the final sentence.

We define  $\mathbf{p}$  to be the type of inquisitive propositions, i.e.,  $\mathbf{p} = (\mathbf{s} \rightarrow \mathbf{t}) \rightarrow \mathbf{t}$ , where  $s$  is the type of possible worlds and  $\mathbf{s} \rightarrow \mathbf{t}$  is the type of sets of possible worlds. As the discussion of possible worlds is out of the scope of this thesis, we highly recommend to the interested reader to refer to [Ciardelli et al., 2018] for explanations and justifications. Then, the semantic interpretation of the syntactic categories is as follows:

$$\begin{aligned} \llbracket NP \rrbracket &:= (\mathbf{e} \rightarrow \mathbf{p}) \rightarrow \mathbf{p} \\ \llbracket N \rrbracket &:= \mathbf{e} \rightarrow \mathbf{p} \\ \llbracket S \rrbracket &:= \mathbf{p} \end{aligned}$$

In order to express the semantic interpretation of (5.4), we use the following non-logical constants:

$$\begin{aligned} \mathbf{cheryl} &: \mathbf{e} \\ \mathbf{birthday} &: \mathbf{e} \rightarrow \mathbf{s} \rightarrow \mathbf{t} \\ \mathbf{of} &: \mathbf{e} \rightarrow (\mathbf{e} \rightarrow \mathbf{s} \rightarrow \mathbf{t}) \rightarrow \mathbf{t} \end{aligned}$$

**cheryl** is the name of a person, it is represented by a constant of type  $e$ . **birthday** is a noun, its semantic type corresponds to the semantic interpretation of nouns, as  $\mathbf{e} \rightarrow \mathbf{p} = \mathbf{e} \rightarrow \mathbf{s} \rightarrow \mathbf{t}$  by definition of  $\mathbf{p}$ . **of** links a constant of type  $e$  and a noun and yields a truth value.

Following the inquisitive interpretation of an atomic proposition, we raise the types of the relation symbols:

$$\begin{aligned} \mathbf{birthday}_i &:= \lambda x^e. \mathcal{P}(\mathbf{birthday} x) \\ \mathbf{of}_i &:= \lambda x^e y^e. \mathcal{P}(\mathbf{of} x y) \end{aligned}$$

Where  $\mathcal{P}(X)$  denotes the powerset (set of all subsets) of the set  $X$ . We also raise the equality relation between entities:

$$(x =_i y) := \mathcal{P}(\lambda w^s. x = y) \quad (5.5)$$

In 5.5,  $w$  is a possible world, and as the equality is supposed to be the same in all the possible worlds, the abstraction on  $w$  is used for purely syntactical purposes in the type-raising.

Finally, we provide the lexical entries with the following interpretations:

$$\begin{aligned} \llbracket \mathbf{CHERYL} \rrbracket &:= \lambda p^{\mathbf{e} \rightarrow \mathbf{p}}. p \mathbf{cheryl} \\ \llbracket \mathbf{BIRTHDAY}_i \rrbracket &:= \mathbf{birthday}_i \\ \llbracket \mathbf{POSSESSIVE} \rrbracket &:= \lambda p^{(\mathbf{e} \rightarrow \mathbf{p}) \rightarrow \mathbf{p}} q^{\mathbf{e} \rightarrow \mathbf{p}} r^{\mathbf{e} \rightarrow \mathbf{p}}. p (\lambda x^e. !(\exists_i y^e. (q y) \wedge_i (\mathbf{of}_i x y) \wedge_i (r y))) \\ \llbracket \mathbf{IS} \rrbracket &:= \lambda p^{(\mathbf{e} \rightarrow \mathbf{p}) \rightarrow \mathbf{p}} q^{(\mathbf{e} \rightarrow \mathbf{p}) \rightarrow \mathbf{p}}. q (\lambda x^e. p (\lambda y^e. x =_i y)) \\ \llbracket \mathbf{WHEN} \rrbracket &:= \lambda p^{((\mathbf{e} \rightarrow \mathbf{p}) \rightarrow \mathbf{p}) \rightarrow \mathbf{p}}. \exists_i x^e. p (\lambda q^{\mathbf{e} \rightarrow \mathbf{p}}. q x) \end{aligned}$$

The operators written with the index  $i$  along with  $!$  are the operators of IS, see Chapter 1, section 1.7 for the formal definitions. The semantic representation of **CHERYL** is a type-raised version of the corresponding constant. The semantic representation of the possessive awaits for three arguments: a noun phrase  $p$ , a noun  $q$ , and  $r$ , a hidden variable in the semantic representation of the noun phrase, that has no lexical realization but appears in the computation. The semantic representation of **IS** corresponds to a type raised version of the equality as shown in 5.5. The semantic representation of **WHEN** introduces the inquisitive existential quantification on the queried variable ( $x$ ).

Then, applying the above semantic recipes to term (5.4) yields the following results:

$$\exists_i x^e . !(\exists_i y^e . (\mathbf{birthday}_i y) \wedge_i (\mathbf{of}_i \mathbf{cheryl} y) \wedge_i (y =_i x)) \quad (5.6)$$

The term under the scope of ! is read as follows: there is an entity  $y$ , such that  $y$  is a birthday, it is in the possession of Cheryl, and there is an equality between  $y$  and  $x$ .  $x$  is linked to the existential quantifier outside the scope of !. As the ! cancels the inquisitive component of the proposition under its scope, the inquisitive part of 5.6 comes from the existential quantification on  $x$ : the query of this question is about  $x$ .

### 5.4.3 Dialogue participant’s epistemic state

As explained in the beginning of section 5.4, the short exchange between Cheryl and Albert in example 18 exemplifies the epistemic nature of a (cooperative) dialogue, and demonstrates the need to represent the knowledge states of the dialogue participants. Consequently, we must add to the dialogue context some information that models the private knowledge of each agent (i.e., each dialogue participant)<sup>114</sup>.

Following [Ciardelli and Roelofsen, 2015], we associate to each agent  $\mathbf{a}$  and each possible world  $w$  an inquisitive proposition  $\Sigma_{\mathbf{a},w}$  that models the epistemic and inquisitive state of agent  $\mathbf{a}$  at world  $w$ . In type-theoretic terms, this can be modeled by a function  $\Sigma$  of type  $\mathbf{e} \rightarrow \mathbf{s} \rightarrow \mathbf{p}$ . The epistemic modality associated to agent  $\mathbf{a}$  is then defined as follows in equation 5.7, where  $\mathbf{K} \mathbf{a} \varphi$  stands for “ $\mathbf{a}$  knows  $\varphi$ ”.

$$\mathbf{K} \mathbf{a} \varphi := \lambda q^{\mathbf{s} \rightarrow \mathbf{t}} . \forall w^{\mathbf{s}} . (q w) \rightarrow (\varphi (\bigcup (\Sigma \mathbf{a} w))) \quad (5.7)$$

Where  $\bigcup S = \lambda x . \exists a . (S a) \wedge (a x)$ .  $\mathbf{K} \mathbf{a} \varphi$  is defined as a set of sets of possible worlds and contains a picture of  $\mathbf{a}$ ’s knowledge of  $\varphi$  across all possible worlds.

Let us now continue the example started in the previous section by showing how to interpret the sentence *Albert does not know when Cheryl birthday is*. Its abstract syntax is given by the term:

$$\text{NOT} (\text{KNOW} (\text{WHEN} (\lambda x^{NP} . \text{IS } x (\text{POSSESSIVE CHERYL BIRTHDAY})))) \text{ALBERT} \quad (5.8)$$

where in addition to the already defined abstract syntactic constants, we have:

$$\begin{aligned} \text{ALBERT} &: NP \\ \text{KNOW} &: S \rightarrow NP \rightarrow S \\ \text{NOT} &: (NP \rightarrow S) \rightarrow NP \rightarrow S \end{aligned}$$

The syntactic constant for Albert follows the same constraints as the one for Cheryl. KNOW awaits for two arguments, the first one being the subject of the verb (the person who knows) and the second the object of knowledge. The type of NOT follows the classical definition, the negation is applied to a verb phrase ( $VP$ ) and yields a  $VP$  as well. As  $VP = NP \rightarrow S$ , we obtain the type given above.

<sup>114</sup>In [Ginzburg, 2012], the dialogue context includes, in addition to the common dialogue gameboard, private dialogue gameboards, one for each agent. Our approach is slightly different: what we model is not quite the private knowledge of each agent but rather what is commonly known about this private knowledge.

The semantic interpretations of these new lexical entries is then as follows:

$$\begin{aligned} \llbracket \text{ALBERT} \rrbracket &:= \lambda p^{\mathbf{e} \rightarrow \mathbf{P}}. p \text{ albert} \\ \llbracket \text{KNOW} \rrbracket &:= \lambda p^{\mathbf{P}} q^{(\mathbf{e} \rightarrow \mathbf{P}) \rightarrow \mathbf{P}}. q (\lambda x^{\mathbf{e}}. \mathbf{K} x p) \\ \llbracket \text{NOT} \rrbracket &:= \lambda p^{((\mathbf{e} \rightarrow \mathbf{P}) \rightarrow \mathbf{P}) \rightarrow \mathbf{P}} q^{(\mathbf{e} \rightarrow \mathbf{P}) \rightarrow \mathbf{P}}. q (\lambda x^{\mathbf{e}}. \neg_i (p (\lambda r^{\mathbf{e} \rightarrow \mathbf{P}}. r x))) \end{aligned}$$

With these interpretations, we compute the expected interpretation of (5.8):

$$\neg_i (\mathbf{K} \text{ albert} (\exists_i x^{\mathbf{e}}. !(\exists_i y^{\mathbf{e}}. (\text{birthday}_i y) \wedge_i (\text{of}_i \text{ cheryl} y) \wedge_i (x =_i y))))$$

This term states the inquisitive negation of the fact that Albert knows the proposition given in equation 5.6, corresponding to the semantic representation of the question “when Cheryl’s birthday is”.

Now that we have presented the computation process for the utterances involving epistemic modalities, we need to address the issues of dynamicity in dialogue, specifically through context updating mechanisms.

#### 5.4.4 Dialogue dynamics and context updating

As we have seen, inquisitive logic allows one to assign a formal semantics to each dialogue turn in a compositional way akin to Montague’s [Ciardelli et al., 2017]. The next step is to provide our model with some dynamics that will allow a dialogue turn to update the current dialogue context. For this purpose, we adapt CSDS, introduced in [de Groote, 2006] and further developed in [Lebedeva, 2012]. This approach has several advantages. It allows several dynamic phenomena to be integrated in a same framework (typically, discourse dynamic, as in [Groenendijk and Stokhof, 1991, Dotlačil and Roelofsen, 2019], and epistemic dynamic as in [Van Ditmarsch et al., 2007, Ciardelli et al., 2018]). It also allows for a treatment of dynamics at a subsentential level (as in [Muskens, 1996])<sup>115</sup>.

The first question to settle is how to model dialogue contexts. We have seen that a typical dialogue gameboard consists of the *speaker*, the *addressee*, the *FACTS*, and the *QUD*. We have also seen that both the *FACTS* and the *QUD* may be encoded as a single inquisitive proposition. In addition, a dialogue context must also contain information about the private knowledge of the dialogue participants. Accordingly, we define a *dialogue context* as following:

**Definition 34 (Dialogue context, [Boritchev and de Groote, 2020])**

We represent a dialogue context by a 4-tuple  $(s, a, Q, K)$  where:

- $s$ , which is of type  $\mathbf{e}$ , is the speaker;
- $a$ , which is of type  $\mathbf{e}$ , is the addressee;
- $Q$ , which is of type  $\mathbf{P}$ , is an inquisitive proposition that models both the *FACTS* and the *QUD*;
- $K$ , which is of type  $\mathbf{e} \rightarrow \mathbf{s} \rightarrow \mathbf{P}$ , is the function that associates to each agent their epistemic state at a given possible world.

---

<sup>115</sup>For the sake of conciseness and simplicity, in this section, we give a simplified version that does not allow for anaphora resolution. This simplification dispenses one from modeling the so-called right context using a continuation. Taking anaphora resolution into account is feasible but involves a lot of technical details that are orthogonal to our main concern.

$K$  takes an agent and a possible world as argument and returns an inquisitive proposition corresponding to the sought epistemic state.

Let  $\mathbf{d} = \mathbf{e} \times \mathbf{e} \times \mathbf{p} \times (\mathbf{e} \rightarrow \mathbf{s} \rightarrow \mathbf{p})$  be the type of dialogue contexts, corresponding to definition 34. We posit the existence of four context accessing functions:

$$\begin{aligned} \mathbf{speaker} &: \mathbf{d} \rightarrow \mathbf{e} \\ \mathbf{addressee} &: \mathbf{d} \rightarrow \mathbf{e} \\ \mathbf{qud} &: \mathbf{d} \rightarrow \mathbf{p} \\ \Sigma &: \mathbf{d} \rightarrow \mathbf{e} \rightarrow \mathbf{s} \rightarrow \mathbf{p} \end{aligned}$$

which correspond to *projection operators*: **speaker** retrieves the speaker from the dialogue context, **addresse** does the same for the addressee and **qud** for the inquisitive proposition corresponding both to FACTS and the QUD.  $\Sigma$  retrieves the mapping of the function  $K$ .

Given the notion of a dynamic context, a *dynamic proposition* is defined to be an inquisitive proposition depending on such a context. Hence, we define  $\mathbf{P} = \mathbf{d} \rightarrow \mathbf{p}$  to be the type of dynamic propositions. The interpretation of the logical connectives and quantifiers must then be changed in order to accommodate dynamic propositions. First, let us recall the interpretations in a static setting, given in Chapter 1, section 1.7:

$$\begin{aligned} \mathbf{R}_i t_1 \dots t_n &:= \mathcal{P}(\mathbf{R} t_1 \dots t_n) \\ &\text{where } \mathcal{P} a = \lambda b^{\mathbf{s} \rightarrow \mathbf{t}}. \forall w^{\mathbf{s}}. (b w) \rightarrow (a w) \\ \varphi \wedge_i \psi &:= \lambda a^{\mathbf{s} \rightarrow \mathbf{t}}. (\varphi a) \wedge (\psi a) \\ \varphi \vee_i \psi &:= \lambda a^{\mathbf{s} \rightarrow \mathbf{t}}. (\varphi a) \vee (\psi a) \\ \varphi \rightarrow_i \psi &:= \lambda a^{\mathbf{s} \rightarrow \mathbf{t}}. \forall b^{\mathbf{s} \rightarrow \mathbf{t}}. (\forall w^{\mathbf{s}}. (b w) \rightarrow (a w)) \rightarrow ((\varphi b) \rightarrow (\psi b)) \\ \neg_i \varphi &:= \lambda a^{\mathbf{s} \rightarrow \mathbf{t}}. \forall w^{\mathbf{s}}. (a w) \rightarrow \neg(\exists b^{\mathbf{s} \rightarrow \mathbf{t}}. (\varphi b) \wedge (b w)) \\ \forall_i x^{\mathbf{e}}. \varphi x &:= \lambda a^{\mathbf{s} \rightarrow \mathbf{t}}. \forall x^{\mathbf{e}}. \varphi x a \\ \exists_i x^{\mathbf{e}}. \varphi x &:= \lambda a^{\mathbf{s} \rightarrow \mathbf{t}}. \exists x^{\mathbf{e}}. \varphi x a \\ !\varphi &:= \neg_i \neg_i \varphi \\ ?\varphi &:= \varphi \vee_i \neg_i \varphi \end{aligned}$$

The new, dynamic interpretation is the following:

$$\begin{aligned} \mathbf{R}_d t_1 \dots t_n &:= \lambda c^{\mathbf{d}}. (\mathbf{R}_i t_1 \dots t_n) \wedge_i !(\mathbf{qud} c) \\ \varphi \wedge_d \psi &:= \lambda c^{\mathbf{d}}. (\varphi c) \wedge_i (\psi c) \\ \varphi \vee_d \psi &:= \lambda c^{\mathbf{d}}. (\varphi c) \vee_i (\psi c) \\ \varphi \rightarrow_d \psi &:= \lambda c^{\mathbf{d}}. (\varphi c) \rightarrow_i (\psi c) \\ \neg_d \varphi &:= \lambda c^{\mathbf{d}}. \neg_i (\varphi c) \\ \forall_d x^{\mathbf{e}}. \varphi x &:= \lambda c^{\mathbf{d}}. \forall_i x^{\mathbf{e}}. \varphi x c \\ \exists_d x^{\mathbf{e}}. \varphi x &:= \lambda c^{\mathbf{d}}. \exists_i x^{\mathbf{e}}. \varphi x c \\ !_d \varphi &:= \lambda c^{\mathbf{d}}. !(\varphi c) \\ ?_d \varphi &:= \lambda c^{\mathbf{d}}. ?(\varphi c) \\ \mathbf{K}_d \mathbf{a} \varphi &:= \lambda c^{\mathbf{d}}. \lambda q^{\mathbf{s} \rightarrow \mathbf{t}}. (\forall w^{\mathbf{s}}. (q w) \rightarrow (\varphi c (\bigcup (\Sigma c \mathbf{a} w)))) \wedge !(\mathbf{qud} c) q \end{aligned}$$

This calls for few comments. A first remark is that  $!(\mathbf{qud} c) = \text{FACTS}$  and  $?(\mathbf{qud} c) = \text{QUD}$  in the typical DGB view. Then, it is important to note that the interpretation of an atomic

proposition is now sensitive to the context because it is intersected with the current *FACTS*. Thus, if the context establishes that Cheryl's birthday is either May 15 or June 17, the question "when is Cheryl's birthday" will be equivalent to the question "is Cheryl's birthday May 15 or June 17". The interpretation of the epistemic modality is also sensitive to the context.

Using the dynamic logic, it is now possible to provide an interpretation to the sentence "I don't know when your birthday is". To this end, we interpret the syntactic categories dynamically:

$$\begin{aligned} \llbracket NP \rrbracket &:= (\mathbf{e} \rightarrow \mathbf{P}) \rightarrow \mathbf{P} \\ \llbracket N \rrbracket &:= \mathbf{e} \rightarrow \mathbf{P} \\ \llbracket S \rrbracket &:= \mathbf{P} \end{aligned}$$

The lexical entries are kept unchanged except for the fact that the atomic propositions and the logical connectives are interpreted dynamically. For instance, we now have:

$$\begin{aligned} \llbracket \text{BIRTHDAY} \rrbracket &:= \mathbf{birthday}_d \\ \llbracket \text{WHEN} \rrbracket &:= \lambda p^{(\mathbf{e} \rightarrow \mathbf{P}) \rightarrow \mathbf{P}}. \exists_d x^{\mathbf{e}}. p(\lambda q^{\mathbf{e} \rightarrow \mathbf{P}}. q x) \end{aligned}$$

Then, we can extend our semantic lexicon with the semantic interpretations of deictics "I" and "you":

$$\begin{aligned} \llbracket \text{I} \rrbracket &:= \lambda p^{\mathbf{e} \rightarrow \mathbf{P}}. \lambda c^d. p(\mathbf{speaker} c) c \\ \llbracket \text{YOU} \rrbracket &:= \lambda p^{\mathbf{e} \rightarrow \mathbf{P}}. \lambda c^d. p(\mathbf{addressee} c) c \end{aligned}$$

Their semantic interpretations correspond to the retrieving of the speaker and the addressee, respectively, from the context. The semantic representation of "I" is type-raised. It awaits for a dialogue context  $c$ . It applies the **speaker** accessing function to  $c$ , in order to retrieve the current speaker. The semantic representation of "you" proceeds similarly to access the current addressee.

We now need to model the way an utterance acts on the dialogue context. To this end, we define the following updating functions:

$$\begin{aligned} \mathbf{sets} &:= \lambda c^d x^{\mathbf{e}}. (x, \mathbf{addressee} c, \mathbf{qud} c, \Sigma c) \\ \mathbf{seta} &:= \lambda c^d x^{\mathbf{e}}. (\mathbf{speaker} c, x, \mathbf{qud} c, \Sigma c) \\ \mathbf{upd} &:= \lambda c^d a^{\mathbf{P}}. (\mathbf{speaker} c, \mathbf{addressee} c, (\mathbf{qud} c) \wedge_i (a c), \lambda x^{\mathbf{e}} w^{\mathbf{s}}. (\Sigma c x w) \wedge (a c)) \end{aligned}$$

The first two set the speaker and the addressee, respectively. The third updates both the inquisitive proposition modeling *FACTS* and *QUD* and the mapping of the epistemic function  $K$ . The **sets** updating function sets the speaker by injecting the variable  $x$  in the speaker position in the dialogue context. **seta** proceeds similarly, injecting  $x$  in the addressee position in the dialogue context. **upd** updates the two last fields: it adds the inquisitive dynamic proposition  $a$  to the *QUD&FACTS* field while simultaneously updating the epistemic state of the agents for a given possible world  $w$ .

We then define a *dialogue turn* to be a triple  $(s, a, \varphi)$ , of type  $\mathbf{e} \times \mathbf{e} \times \mathbf{P}$ , where  $s$  is the speaker,  $a$  is the addressee, and  $\varphi$  is a dynamic proposition that expresses the semantics of the dialogue turn. Finally, we define the action on a dialogue context  $C$  of such a dialogue turn as follows:

$$C \circ (s, a, \varphi) = \mathbf{upd}(\mathbf{seta}(\mathbf{sets} c s) a) \varphi$$

A dialogue turn sets the speaker and the addressee. Then it updates the *QUD&FACTS* along with the epistemic state of the agents.

Now that we have formally defined all the ingredients of our model, we can showcase it on a complete example. The next section unravels the model of toy dialogue, using the computations we presented throughout the current section.

### 5.4.5 A complete example: Cheryl’s birthday

We illustrate our model by applying it to the logical puzzle known as “*When is Cheryl’s Birthday*”. Here is the wording of the problem as it appeared on the New York Times website in April 2015.<sup>116</sup>

- (19) [CONTEXT: Albert and Bernard just met Cheryl. “When’s your birthday?” Albert asked Cheryl. Cheryl thought a second and said, “I’m not going to tell you, but I’ll give you some clues.” She wrote down a list of 10 dates:

May 15,	May 16,	May 19,
June 17,	June 18	
July 14,	July 16	
August 14,	August 15,	August 17

“My birthday is one of these,” she said. Then Cheryl whispered in Albert’s ear the month — and only the month — of her birthday. To Bernard, she whispered the day, and only the day.]

- a. Cheryl (to Albert): *Can you figure it out now?*
- b. Albert: *I don’t know when your birthday is, but I know Bernard doesn’t know, either.*
- c. Bernard: *I didn’t know originally, but now I do.*
- d. Albert: *Well, now I know, too!*

When is Cheryl’s birthday?

To solve the problem, the first step is to formalize the initial dialogue context. To this end, a solution could be to define a first-order object language. This language would include atomic propositions such as **May15**, **Jun17**, **Jul14**, *etc.* (with the obvious intended meanings). Then, we would have to posit meaning postulates such as  $\neg_i(\mathbf{May15} \wedge_i \mathbf{Jun17})$ ,  $\neg_i(\mathbf{May15} \wedge_i \mathbf{Jul14})$ , *etc.* By following such an approach, we would model the *QUD* as a formula expressing the inquisitive disjunction of the possible birthdate:

$$\mathbf{May15} \vee_i \mathbf{Jun17} \vee_i \mathbf{Jul14} \vee_i \text{etc}$$

As our objective is to make our explanation clear by not overcharging them with too much syntactic details, we prefer to leave the object language implicit and reason in semantic terms with possible worlds. The set of possible world is defined in such a way that each world corresponds to a possible birthdate:

$$W = \{w_{5.15}, w_{5.16}, w_{5.19}, w_{6.17}, w_{6.18}, w_{7.14}, w_{7.16}, w_{8.14}, w_{8.15}, w_{8.17}\}$$

Then the initial *QUD*, in its semantic version, corresponds to the following inquisitive proposition:

$$Q_0 = \left[ \begin{array}{l} \{w_{5.15}\}, \{w_{5.16}\}, \{w_{5.19}\}, \{w_{6.17}\}, \{w_{6.18}\}, \\ \{w_{7.14}\}, \{w_{7.16}\}, \{w_{8.14}\}, \{w_{8.15}\}, \{w_{8.17}\} \end{array} \right]$$

<sup>116</sup><https://www.nytimes.com/2015/04/15/science/a-math-problem-from-singapore-goes-viral-when-is-cheryls-birthday.html>



Where  $[a, b, c, \dots]$  denotes the downward-closure of the set  $\{a, b, c, \dots\}$ , following the convention introduced in Chapter 1, section 1.7. The initial  $QUD$  is the downward-closure of the set of singletons, each corresponding to a possible birthdate.

We can now focus on the modeling of the knowledge of the agents. Consider, for instance, the possible world  $w_{5.15}$ . In this world, the month of Cheryl's birthdate is May, and Albert knows it. Albert is therefore in an inquisitive state where he wonders what is the day of Cheryl's birthdate, knowing that it is either the 15th, the 16th, or the 19th. This inquisitive state is represented by the following proposition:

$$[\{w_{5.15}\}, \{w_{5.16}\}, \{w_{5.19}\}]$$

Continuing this line of reasoning, we obtain that Albert's knowledge is modeled by the following map, as the reasoning developed for the possible world  $w_{5.15}$  is the same for all the worlds in which the month of the birthdate is May, and similarly for the other months.

$$K_0 \mathbf{albert} = \begin{cases} w_{5.15} \mid w_{5.16} \mid w_{5.19} \mapsto [\{w_{5.15}\}, \{w_{5.16}\}, \{w_{5.19}\}] \\ w_{6.17} \mid w_{6.18} \mapsto [\{w_{6.17}\}, \{w_{6.18}\}] \\ w_{7.14} \mid w_{7.16} \mapsto [\{\{w_{7.14}\}, \{w_{7.16}\}\}] \\ w_{8.14} \mid w_{8.15} \mid w_{8.17} \mapsto [\{w_{8.14}\}, \{w_{8.15}\}, \{w_{8.17}\}] \end{cases}$$

Similarly for Bernard, replacing the months by the dates in the reasoning:

$$K_0 \mathbf{bernard} = \begin{cases} w_{7.14} \mid w_{8.14} \mapsto [\{w_{7.14}\}, \{w_{8.14}\}] \\ w_{5.15} \mid w_{8.15} \mapsto [\{w_{5.15}\}, \{w_{8.15}\}] \\ w_{5.16} \mid w_{7.16} \mapsto [\{w_{5.16}\}, \{w_{7.16}\}] \\ w_{6.17} \mid w_{8.17} \mapsto [\{w_{6.17}\}, \{w_{8.17}\}] \\ w_{6.18} \mapsto [\{w_{6.18}\}] \\ w_{5.19} \mapsto [\{w_{5.19}\}] \end{cases}$$

As for Cheryl, her knowledge (which is irrelevant for the example) corresponds to the map that assigns to each world  $w$  the proposition  $[\{w\}]$ , because she knows when her birthday is.

The initial dialogue context is therefore  $C_0 = (Q_0, K_0)$ .<sup>117</sup> Let us now consider the dialogue turns. The first one, (2-a), simply restates the  $QUD$  and does not affect the context. We therefore have  $C_1 = C_0$ . The second turn, (2-b), is more interesting. It is interpreted as the following dynamic proposition:

$$\neg_d(K_d \mathbf{albert} \varphi) \wedge (K_d \mathbf{albert} \neg_d(K_d \mathbf{bernard} \varphi)) \quad (5.10)$$

$\varphi$  corresponds to a dynamic proposition that amounts to the  $QUD$  when evaluated with respect to the current context. The first conjunct, when evaluated with respect to the current context, yields  $\mathcal{P}(W)$ , i.e., the always true proposition. The second conjunct is more interesting. When evaluated with respect to  $C_1$ ,  $(K_d \mathbf{bernard} \varphi)$  is interpreted as  $[\{w_{6.18}, w_{5.19}\}]$ . Hence,  $\neg_d(K_d \mathbf{bernard} \varphi)$  is interpreted as  $[\{w_{5.15}, w_{5.16}, w_{6.17}, w_{7.14}, w_{7.16}, w_{8.14}, w_{8.15}, w_{8.17}\}]$ . And finally,  $K_d \mathbf{albert} \neg_d(K_d \mathbf{bernard} \varphi)$  is interpreted as  $[\{w_{7.14}, w_{7.16}, w_{8.14}, w_{8.15}, w_{8.17}\}]$ . Then, updating the context  $C_1$  yields a context  $C_2$  with the  $QUD$   $Q_2$  and the function  $K_2$  where:

$$Q_2 = [\{w_{7.14}\}, \{w_{7.16}\}, \{w_{8.14}\}, \{w_{8.15}\}, \{w_{8.17}\}]$$

<sup>117</sup>We leave implicit the speaker and the addressee.

$$\begin{aligned}
K_2 \text{ albert} &= \begin{cases} w_{7.14} | w_{7.16} & \mapsto [\{w_{7.14}\}, \{w_{7.16}\}] \\ w_{8.14} | w_{8.15} | w_{8.17} & \mapsto [\{w_{8.14}\}, \{w_{8.15}\}, \{w_{8.17}\}] \\ - & \mapsto [\emptyset] \end{cases} \\
K_2 \text{ bernard} &= \begin{cases} w_{7.14} | w_{8.14} & \mapsto [\{w_{7.14}\}, \{w_{8.14}\}] \\ w_{5.15} | w_{8.15} & \mapsto [\{w_{8.15}\}] \\ w_{5.16} | w_{7.16} & \mapsto [\{w_{7.16}\}] \\ w_{6.17} | w_{8.17} & \mapsto [\{w_{8.17}\}] \\ - & \mapsto [\emptyset] \end{cases}
\end{aligned}$$

By applying the same analysis to the other dialogue turns, one obtains successively the following *QUDs*:  $Q_3 = [\{w_{7.16}\}, \{w_{8.15}\}, \{w_{8.17}\}]$  and  $Q_4 = [\{w_{8.15}\}]$ , which settles the problem: Cheryl’s birthday is August 15th.

#### 5.4.6 Model review and conclusions

We can now review this model and compare it to related approaches. [Van Ditmarsch et al., 2007] presents Dynamic Epistemic Logic (DEL). In DEL, situations are described through sets of agents, each with individual available states of information. Then, as agents perform actions, DEL gives a way to describe the changes in the state of available information, for each agent.

Growing on DEL with an inquisitive take, [Ciardelli and Roelofsen, 2015] introduces Inquisitive Dynamic Epistemic Logic (IDEL), a framework designed to provide tools that can be used to model the information exchange between a set of agents as a dynamic process through raising and solving of issues. The approach taken in the paper chooses a bi-categorical presentation of Inquisitive Semantics, with a strict separation between interrogatives and declarative sentences. The authors reference [Ciardelli and Roelofsen, 2015] for a meaning-preserving translation between this presentation and the one we use in the current section, where interrogatives and declarative sentences are modeled as the same type of objects. In IDEL, issues are raised when the agents ask questions and resolved when they make assertions. This is quite orthogonal to the vision of dialogue defended by Ginzburg in KoS framework [Ginzburg, 2016] and that we follow here, where every speech act gives rise to a QUD, which corresponds to an issue.

IDEL is designed “under the assumption that an agent’s information is always truthful”. Though the example we show here does not illustrate this, our model is designed keeping the objective of working with real-life data in mind and therefore, taking settings where disagreements can and do occur into account. Participants may reject an asserted fact. The negotiation phases model adds a protective additional step in the computation of the dialogue representations that directly bypasses this issue. [Ciardelli and Roelofsen, 2015] suggests to try using weaker epistemic modalities such as belief and allowing disagreement to occur in order to address this difficulty.

The last section of [Ciardelli and Roelofsen, 2015] draws a comparison between IDEL and Dynamic Epistemic Logic with Questions (DELQ), as presented in [Bentham and Minică, 2012]. DELQ is based on epistemic models enriched with a set of issues, one per agent. Then, dynamics is added through several actions, of which we focus on two: public announcement “that  $\phi$  is the case” and public asking “whether  $\phi$  is the case”. Thus, in DELQ all the sentences are considered to be declarative, none are treated as syntactically interrogative or semantically inquisitive [Ciardelli and Roelofsen, 2015]. The difference between questions and assertions is drawn through dynamic actions, at the speech act level. In IDEL, the difference between questions and assertions exists at the syntactic level through the form of interrogative sentences. In

our approach, the difference between questions and assertions is acknowledged at the syntactic level but is smoothed in the semantics, as we represent issues and propositions as the same type of objects.

[Ciardelli and Roelofsen, 2015] concludes on the need to investigate a dynamic epistemic version of [Ciardelli et al., 2018], the version of Inquisitive Semantics we presented in section 5.4.2. The current section presents our take on this investigation. We do not claim here that our model works better than IDEL as our idea comes from a different perspective: starting from dialogue studies and taking an orientation towards real-life data modeling.

Our approach grows from linguistic considerations of interrogative and declarative sentences as speech acts. We take roots in [Ginzburg, 2012] but also in the syntactic parses of the speech acts in order to build our representations. The model presented in the current section addresses phenomena related to context-managing but also to dialogue management, through the way utterances influence public knowledge of private contexts. We combine several frameworks in order to model complex dialogical interactions in a logically sound way. Solving *Cheryl's birthday* puzzle gives us a proof of concept for the possibilities of logical reasoning through dialogue modeling. Inquisitive Semantics provides a uniform way of modeling interrogative and declarative sentences, which we think to be of the greatest importance when dealing with dialogue modeling, especially in a real-life data perspective. Next, our model needs to be scaled up in order to be applied to bigger and more complex dialogues. We hope to achieve that through the articulation of negotiation phases. For now, we bypassed several linguistic and logical problems related to tense and modality; future work should take these into account. Another interesting research direction would be to compare the way our model behaves on English with other, especially non-Indo-European, languages.

## 5.5 Conclusion of the chapter

The work presented in this thesis aims for the elaboration and construction of a model of dialogue in a dynamic framework. The current chapter presents the major steps we took towards achieving this goal. We started by considering the overall task of modeling long interactions and came up with a mean of subdivision through the idea of *negotiation phases*. Inside negotiation phases, we focused on questions, answers, and mechanisms for question-and-answer articulation. This led us to the definition of the *relative response set*, a mathematical tool that acts as a materialization of the pragmatic glue that exists between a question and its answer. Focusing on the *wh*-questions and using the answers typology introduced in Chapter 2, section 2.3.1, we analysed the possible usage cases of the relative response set. We concluded section 5.2 by stating the need for a more formal approach with an explicit syntax-semantics interface.

Section 5.3 answers this need and presents a syntax-semantics interface for part of the *wh*-questions. We show that our approach unfolds fully for *wh*-words such as “when”, “where”, “who”, “whom”, “which”. This model also accounts for polar questions and declarative sentences. Then, we discuss the complications raised by “what”, “whose”, “why”, and “how”. In some cases, these difficulties can be bypassed with solutions such as the one presented in section 5.4 for the possessive in the phrase “Cheryl’s birthday”, that could be envisioned to account for questions featuring the *wh*-word “whose”. On the contrary, questions such as “why are you doing this?” or “what is happening?” bring forward complex considerations. The major difficulty with both these questions is linked to the fact that the scope of the answer is extremely wide: one can imagine a huge variety of possible answers, of different semantic and even syntactic types, that could all fit these questions (“Because I can”; “They started it”; “Unicorns like tea”; “Love.”, etc.).

After showing the possibility for a syntax-semantics interface to be written for dialogue utterances, we zoom out to a larger view of dialogue, approaching the negotiation phases from an external point of view. We dig into the interactions between utterances and context in dialogue and use the *Cheryl's birthday* puzzle to develop and showcase our model. Though we base our model on a toy example, we develop it in a systematic and computational approach through the ACGs framework and the aggregation of several well-studied and well-established formalisms and frameworks.

The current chapter is the last chapter of this thesis. There is still a lot of work to do to achieve our goal of a dynamic dialogue model, but we believe that our investigations have resulted in a number of formal and language-based considerations that can pave the way to further works on these matters. We conclude this chapter by drawing a commented outline of the work that has yet to be done.

1. Extend the model presented in section 5.3 to the wh-words “what”, “whose”, “why”, and “how”.
2. Check that the new version of the model accounts for disjunctive questions.
3. Confront the model with the account of answers presented in section 5.2.
4. Adapt the model for Clarification Requests.
5. Check that the model can now account for Non-Sentential Utterances.
6. Test the model on real-life negotiation phases from the DinG corpus.
7. Adapt the context-related mechanisms from section 5.4 to the new version of the model.
8. Operationalize this final version of the model so that it is able to process complete multi-logues from DinG.

As discussed in section 5.3.3, achieving (1) would demand the use of a higher-order logic as basis for our formalisms. The problem this raises is not theoretic; the main issue resides in the fact that the ACG toolkit, in its current implementation, does not account for higher-order logic. This means that it will be possible to obtain at least a proof of concept by doing the computation by hand before an extension of the ACG toolkit is implemented.

As presented in Chapter 2, section 2.3.1, disjunctive questions present several alternatives among which the answer should be selected. The reason why (2) might not be straightforward is twofold: first, IS makes a clear distinction between the semantics of inclusive and exclusive disjunctive questions (see Chapter 2, section 2.4.1), and we need to choose where to stand in regards of this distinction, because of the second reason. Indeed, the annotation experiments we conducted on the DinG corpus, presented in Chapter 4, section 4.3, show that it is not always easy to separate disjunctive questions from yes/no questions when working with transcribed oral data. In these conditions, it is not clear how wise would keeping a further specification of disjunctive questions be.

In its current state, our model is able to compute semantic representations of declarative sentences, which accounts for part of the answers in dialogue. Section 5.2 develops a description of different types of answer-related phenomena in dialogue from a formal point of view. In order to achieve (3), we need to test updated versions of the model while being particularly careful with the notions surrounding underspecified answers (such as *aboutness*-related phenomena, see

section 5.2.3). If the updated model can account for these fine-grained phenomena, it will be possible to integrate to it notions of coherence between questions and answers.

Keeping in mind our will to account for real-life interactions, it is fundamental for our model to produce logically valid semantic representations of Clarification Requests (CRs). (4) and (5) raise the same issue: how can we produce semantic representations that are logically valid and respect the type-constraints for utterances that are not complete and grammatically valid sentences? Unfortunately, the answer might be that we cannot. Working with CRs first restrains the problem to a category of NSUs that has been extensively studied and for which a clear typology has been established.

Once points (1-5) are more or less successfully addressed, it will be necessary to confront our updated model with real-life data. This is where the negotiation phases will play their role: for computational reasons, we do not want to test our model directly on long interactions. However, we still want to test the model on coherent units of dialogue, so that it is possible for us to (at least) solve inter-utterance references. Thus, to achieve (6), it is first necessary for us to be able to automatically slice multilogues from DinG in negotiation phases, at least approximately. Tools to this end are currently under development as part of the SLODiM<sup>118</sup> project.

Addressing points (1-6) is bound to change our model significantly. Thus, it will be important at this point to check whether the context-handling mechanisms we detailed in section 5.4 are still valid, and if not, what adaptations are necessary to account for the different changes. This is the objective of (7).

Finally, we hope that all these efforts will allow us to reach a point when (8) will be the only concern left. Then, it seems reasonable to assume that by this point, we will need to integrate our model in a complete pipeline that would feature at least a dialogue parser and an automatic tagger for questions and answers types. Chapter 4, section 4.3 lets us envision part of the issues the development of the latter can encounter. We leave the description of the other difficulties that we will encounter at this moment in our work to our future happy selves that will have reached it.

---

<sup>118</sup><https://slodim.loria.fr/>

# Conclusion

This thesis shows the successive steps we took towards the creation of a formal semantics model of dialogue in a dynamic framework. First, we present an overview of formal semantics of discourse in Chapter 1. We define the linguistic and logic phenomena we focus on throughout this thesis and compare the existing formalisms and frameworks for formal semantics through the scope of these phenomena, in a dialogue modeling perspective. We continue with an exposition of theories that focus on dialogue from different points of view in Chapter 2, ranging from the most linguistic ones to the most computational ones. In this same chapter, we show that in order to study dialogue, it is important to focus on question-answer mechanisms in dialogue. Chapter 2 also introduces typologies of questions and answers.

The second part of this thesis concerns real-life data. We start by displaying our first main contribution: the DinG corpus. Chapter 3 extensively describes the constitution process before comparing DinG to existing corpora of real-life dialogue data of English and French. Then, in Chapter 4, we dive into questions and answers annotation schema for dialogue, in order to achieve a better understanding and knowledge of questions and answers mechanisms in dialogue. We present several annotation schemas, starting from linguistics motivated one and simplifying it to shift to a more computational perspective. We apply our annotation schema for questions on data from DinG.

The third and final part of this thesis focuses on semantic processing of dialogue. Chapter 5 presents three steps we took towards the elaboration and construction of a model of dialogue in a dynamic framework. We gave the general outline of the model we want to obtain in its different aspects: the subdivision of dialogues into negotiation phases, the semantic links between questions and their answers, the semantic representation of questions and answers, the interaction between negotiation phases and the dialogue context.

Writing down the different parts of this thesis has helped us to grasp the full extent of the task at hand. While we do not present a fully functional pipeline that takes as input an oral dialogue and outputs a semantic representation with fully annotated discrepancies between questions and answers, we feel like we have taken significant steps towards this somewhat idealistic goal. During the past years, we have worked simultaneously in two perspectives: a very theoretical one, formal, computational, and anchored in logic; and an experimental one, based on real-life data and observations. We have tried, as much as possible, to build a dialogue between the two, and use the advances in one to inspire the other.

In the past few years, the Natural Language Processing (NLP) community has seen a growth in voices questioning the links between NLP and linguistic research and knowledge. We hope that our work has, at least partly, achieved a satisfying balance in this regard.

## Contributions

The following section states and comments the main contributions of this thesis, in the order in which they appear in the document.

As stated in the introduction of this conclusion, we conducted simultaneously studies in two directions: a formal one and an experimental one. At the early stages of our work, we figured that in order to confront the formal models with real-life data, this real-life data had to be chosen wisely, in a way that the piling-up of real-life phenomena would still allow us to identify which one is causing trouble to the model at the moment. We were also conscious of the fact that, while a lot of data was available on English, even wide-spread languages such as French suffered from a huge lack of freely available transcribed oral data. As we wanted to use the existing research and the local (French) context, restrain the scope of the study in order to not encompass too many different phenomena so we could focus on questions and answers articulation in real-life dialogue, we decided to create Dialogues in Games (DinG): a corpus of transcriptions of real-life, oral, spontaneous multi-party dialogues between French-speaking players of *Catan*.

- The first contribution of this thesis is the **Dialogues in Games (DinG) corpus**, the complete process of its constitution through the recording, transcription and annotation.

The production of a corpus of real-life data with high-quality transcriptions, in particular in the context of multi-party dialogue, is an extremely challenging task. It is costly moneywise but also timewise, as the transcription and segmentation of one minute of recording take approximately two hours for an expert transcriber. We established a recording protocol, in particular with regards to ethics and recording of personal data, a complete and detailed transcription guide, along with an annotation guide.

The annotation of DinG brings us to another contribution of this thesis. In the past years, mainly thanks to collaborations with the international students of the NLP master of the University of Lorraine, we were able to conduct detailed cross-lingual studies of questions and answers in dialogues. This research has resulted in publications such as [Amblard et al., 2019] and [Cruz Blandon et al., 2019]. In the context of this thesis, these collaborations have provided us with a precious cross-lingual perspective on dialogues, that guided us throughout our annotation-related work.

- The second contribution of this thesis is the **elaboration and testing of several cross-lingual annotation schemas for questions and answers in dialogue**, first discussed on English dialogues and then tested by native speakers on corpora in French, Italian, Spanish, Dutch, and Mandarin Chinese.

We applied a simplified annotation schema for questions that we elaborated on the complete set of questions extracted from DinG. This gave us knowledge about the types of questions encountered in DinG’s somewhat simplified context of interaction.

Indeed, we wanted to get better knowledge about the zoology of interactions encountered in real-life dialogue such as the ones from DinG, in order to pursue the second direction of our work: the formal semantics modeling of dialogue. The work we conducted in this direction has resulted in the production of several dialogue models that combine Neo-Davidsonian Event Semantics (as presented in [Champollion, 2017]) with Inquisitive Semantics [Ciardelli et al., 2017] in a compositional and dynamic way. They are based on a syntax-semantics interface within the framework of Abstract Categorical Grammars (ACG) [de Groote, 2001]. The semantic interpretations of declarative and interrogative sentences are computed via an embedding of inquisitive logic in the simply-typed  $\lambda$ -calculus, implemented in the ACG toolkit [Pogodalla, 2016].



- 
- The third contribution of this thesis is the **creation of formal semantic models of dialogue in a dynamic framework**.

We offer an approach to dialogue as a concatenation of negotiation phases governed by a dialogue context. We constitute a compositional process that allows us to compute the semantic representation of utterances from transcriptions of these utterances, via a syntax-semantics interface.

The road we took towards the creation of a formal semantics model of dialogue in a dynamic framework is far from being the most straightforward one, and it is difficult, at this point, to say whether the choices we made were the most efficient. We tried as much as possible to keep a balance between precision and clarity required by our formal approaches and definitions and the necessity of being flexible motivated by our studies of real-life data. The work we conducted in the past years has given us a better understanding of dialogue, both from the formal semantics and the real-life data-driven points of view. It has also made us discover the spectrum existing in between. Because of this, we can now better outline what is left to be done to continue the work presented in this thesis.

## Perspectives

The work presented in this thesis can of course be extended and continued in various directions. The following paragraphs give our main ideas as to where these extensions should start and towards what they should be directed.

**DinG and corpus studies** Chapter 3 shows the work we conducted to build the DinG corpus. We can now present a complete, fully transcribed corpus, but by no means should this sign the end of DinG’s scientific journey. We hope that the perspectives of DinG will be twofold: first, through its extension, and second, through a deepening of the studies that take it as object.

For now, DinG is composed of transcriptions of dialogues between French-speaking players of Catan. This opens at least two straightforward directions for extension: either adding transcriptions of dialogues between players of Catan speaking other languages, or enrich DinG with transcriptions of dialogues between French-speakers playing a game different from Catan. The university game night context in which we conducted our recordings offered us the theoretical possibility for both extensions, and we would of course welcome contributions from other sources, provided the ethical considerations and transcription guide developed for DinG are respected.

We feel like we barely started studying DinG. We have obtained quite a lot of statistical data from automatic analysis of the transcriptions, for example about vocabulary usage, but we didn’t yet have the chance to interpret it and decide on the next phenomena that should be investigated. We will of course continue our study of questions and answers in dialogue in DinG, but we are also opened to collaborations with other researchers on different research questions, and we hope that data from DinG will prove itself useful in various contexts. In particular, we would like to conduct a comparative study with the STAC project, [Asher et al., 2016]<sup>119</sup>, containing a corpus of chat logs from dialogues between English-speaking players of an online version of Catan. As the conversation topic across the two corpora is the same, the cross-language comparison should be meaningful.

---

<sup>119</sup><https://www.irit.fr/STAC/index.html>



**Crosslingual annotations** Chapter 4 presents several questions and answers annotation schema for dialogue along with cross-lingual tests we conducted on English, French, Italian, Spanish, Dutch, and Mandarin Chinese. The first sequel we envision for this work is an extension of the tests to other languages along with an increase of the set of annotators available for each language, for us to be able to evaluate the annotations for each of the considered languages.

Another research direction that appears as a natural continuation of our work lies in the automation of the annotation. We already started working in this direction by using simple rule-based algorithms to facilitate our annotation of questions from DinG, but of course, the goal for which we aim is a fully integrated neural net-based annotation pipeline. The main difficulty we envision for now comes from the fact that the corpus is still relatively small, which can entail an alteration of the quality of automatic annotations.

Finally, we would like to consider applying other annotation schemas to DinG, in particular, the one developed for STAC, but also the very detailed one provided by DIT++ [Bunt, 2009]. We have started working towards a computer-assisted annotation of DinG through the DIT++ schema.

**Dialogue modeling** Chapter 5 lays down the steps we took towards the creation of a model of dialogue in a dynamic framework. We have extensively described the future work in this direction in the conclusion of Chapter 5. To summarize our thoughts here, we envision three main steps in the future development of our model. First, it should be extended in order to account at least for all the linguistic and logic phenomena that we presented in this thesis, throughout Chapters 1 and 2.

Then, we need to conduct thorough tests of our model on DinG. By probing the incremental way in which humans collaboratively explore options, clarify and revise suggestions, and so come to agreements, we hope to shed direct light on the dynamics involved in the process of human understanding. In particular, we want to explore the logical underpinnings for questions and answers in dialogue and the overall dynamic of the conversational exchange.

Next, we would like to integrate a stabilized version of our model into a complete dialogue processing pipeline. This should be part of a much bigger dialogue processing project. We hope that our logical considerations will be a valuable input in the vast subject of dialogue modeling, in particular in contexts where data is scarce and it is still difficult to find and train large statistical models. The logic-based models we have developed throughout this thesis might also prove to be useful for research involving natural language reasoning and inferences. As pointed out by Farah Benamara, we would need to conduct a thorough comparison between a deep-learning model augmented with our logical semantic representations and one without such features, on different types of tasks for dialogue modeling and analysis, following investigations such as the ones presented by [Veldhoen and Zuidema, 2017] and [Badene et al., 2019].

# Appendixes



A

DinG – complete statistical data

## DinG – complete statistical data

N° Rec., speaker	# speech turns	# words	# words / speech turn	Ratio stopwords	# turns / min	# quest.	# quest. / turn	Length (min)
<b>DinG1-1</b>								
Player 1	426	2275	5,34	0,34	7,04	62	0,15	
Player 2	574	3000	5,23	0,32	9,48	85	0,15	
Player 3	720	5570	7,74	0,36	11,89	107	0,15	
Player 4	279	1180	4,23	0,3	4,61	19	0,07	
Other	60	93	1,55	0,13	0,99	0	0,0	
Total	2059	12118			34,02	273		60,53
<b>DinG1-2</b>								
Player 1	423	2058	4,87	0,32	9,66	56	0,13	
Player 2	259	995	3,84	0,31	5,91	16	0,06	
Player 3	376	1862	4,95	0,31	8,58	70	0,19	
Player 4	400	2390	5,97	0,36	9,13	90	0,23	
Other	55	70	1,27	0,07	1,26	1	0,02	
Total	1513	7375			34,54	233		43,8
<b>DinG1 Total</b>	3572	19493				506		104,33
<b>DinG2-1</b>								
Player 1	452	2236	4,95	0,29	6,49	50	0,11	
Player 2	839	5563	6,63	0,3	12,04	100	0,12	
Player 3	529	3026	5,72	0,3	7,59	30	0,06	
Player 4	429	2126	4,96	0,29	6,16	37	0,09	
Other	130	237	1,82	0,18	1,87	4	0,03	
Total	2379	13188			34,15	221		69,66
<b>DinG2-2</b>								
Player 1	121	552	4,56	0,26	7,27	6	0,05	
Player 2	192	1051	5,47	0,26	11,53	29	0,15	
Player 3	102	505	4,95	0,28	6,13	22	0,22	
Player 4	142	794	5,59	0,33	8,59	12	0,08	
Other	33	59	1,79	0,22	1,98	0	0	
Total	590	2961			35,44	69		16,65
<b>DinG2 Total</b>	2969	16149				290		86,31

Table A.1: Quantitative data, DinG corpus, DinG1 and DinG2.

N° Rec., speaker	# speech turns	# words	# words / speech turn	Ratio stopwords	# turns / min	# quest.	# quest. / turn	Length (min)
<b>DinG3-1</b>								
Player 1	373	1477	3,96	0,32	8,12	25	0,07	
Player 2	307	1487	4,84	0,31	6,68	42	0,14	
Player 3	454	2411	5,31	0,31	9,88	28	0,06	
Player 4	247	1201	4,86	0,3	5,38	17	0,07	
Other	96	329	3,43	0,31	2,09	1	0,01	
Total	1477	6905			30,06	113		45,94
<b>DinG3-2</b>								
Player 1	59	263	4,46	0,33	7,61	4	0,07	
Player 2	49	252	5,14	0,28	6,32	2	0,04	
Player 3	78	431	5,53	0,27	10,06	5	0,06	
Player 4	47	245	5,21	0,34	6,06	2	0,04	
Other	6	6	1	0	0,77	0	0	
Total	239	1197			30,04	13		7,76
<b>DinG3 Total</b>	1716	8102				126		53,7
<b>DinG4</b>								
Player 1	550	2772	4,26	0,3	8,56	79	0,12	
Player 2	766	2988	3,9	0,3	10,09	106	0,14	
Player 3	918	6408	6,98	0,35	12,09	104	0,11	
Player 4	568	2536	4,46	0,3	7,48	42	0,07	
Other	183	346	1,89	0,18	2,41	2	0,01	
<b>DinG4 Total</b>	2985	15050			40,63	333		75,93
<b>DinG5-1</b>								
Player 1	114	503	4,41	0,31	13,29	9	0,08	
Player 2	100	539	5,39	0,39	11,66	9	0,09	
Player 3	37	153	4,14	0,32	4,31	4	0,11	
Player 4	122	618	5,07	0,33	14,22	18	0,15	
Other	5	40	4,0	0,4	1,17	0	0,0	
Total	378	1853			44,65	40		8,58
<b>DinG5-2</b>								
Player 1	789	4433	5,62	0,37	11,30	97	0,12	
Player 2	539	2503	4,64	0,33	7,72	61	0,11	
Player 3	406	2014	4,96	0,37	5,81	42	0,10	
Player 4	872	4361	5,0	0,34	12,49	120	0,14	
Other	28	116	4,14	0,32	0,40	2	0,07	
Total	2634	13427			37,72	322		69,83
<b>DinG5 Total</b>	3012	15280				362		78,41

Table A.2: Quantitative data, DinG corpus, DinG3, DinG4 and DinG5.

Appendix A. *DinG* – complete statistical data

N° Rec., speaker	# speech turns	# words	# words / speech turn	Ratio stopwords	# turns / min	# quest.	# quest. / turn	Length (min)
<b>DinG6</b>								
Player 1	881	4452	5,05	0,30	10,49	84	0,10	
Player 2	900	4182	4,65	0,28	10,71	57	0,06	
Player 3	1225	8498	6,94	0,35	14,58	124	0,10	
Other	124	168	1,35	0,09	1,48	0	0,0	
<b>DinG6 Total</b>	3130	17300			37,25	265		84,02
<b>DinG7-1</b>								
Player 1	444	2180	4,91	0,29	8,38	39	0,09	
Player 2	435	2499	5,74	0,33	8,21	37	0,09	
Player 3	437	2145	4,91	0,32	8,24	53	0,12	
Player 4	426	2490	5,85	0,3	8,04	71	0,17	
Other	92	120	1,3	0,12	1,74	0	0,0	
Total	1834	9434			34,59	200		53,02
<b>DinG7-2</b>								
Player 1	373	2037	5,46	0,31	8,61	45	0,12	
Player 2	336	1673	4,98	0,29	7,76	38	0,11	
Player 3	323	1300	4,02	0,33	7,46	37	0,12	
Player 4	354	1285	3,63	0,28	8,17	20	0,06	
Other	73	78	1,07	0,01	1,69	0	0,0	
Total	1459	6373			33,68	140		43,32
<b>DinG7 Total</b>	3293	15807				340		96,34
<b>DinG8-1</b>								
Player 1	110	438	3,98	0,27	5,52	9	0,08	
Player 2	196	1300	6,63	0,37	9,84	16	0,08	
Player 3	240	1150	4,79	0,3	12,05	33	0,14	
Player 4	206	1060	5,15	0,3	10,34	41	0,20	
Other	35	78	2,23	0,27	1,76	0	0,0	
Total	787	4026			39,51	99		19,92
<b>DinG8-2</b>								
Player 1	115	563	4,9	0,31	5,75	9	0,08	
Player 2	176	1181	6,71	0,37	8,80	18	0,10	
Player 3	208	1262	6,07	0,33	10,40	38	0,18	
Player 4	246	1606	6,53	0,31	12,30	32	0,13	
Other	95	194	2,04	0,24	4,75	0	0,0	
Total	840	4806			42,0	97		20,0
<b>DinG8 Total</b>	1627	8832				196		39,92

Table A.3: Quantitative data, DinG corpus, DinG6, DinG7 and DinG8.

N° Rec., speaker	# speech turns	# words	# words / speech turn	Ratio stopwords	# turns / min	# quest.	# quest. / turn	Length (min)
<b>DinG9</b>								
Player 1	88	298	3,39	0,21	2,11	11	0,13	
Player 2	190	788	4,15	0,22	4,55	18	0,10	
Player 3	115	483	4,2	0,33	2,76	24	0,21	
Other	193	953	4,94	0,36	4,63	7	0,04	
Other 2	209	1080	5,17	0,36	5,01	9	0,04	
<b>DinG9 Total</b>	795	3602			14,43	69		41,71
<b>DinG10</b>								
Player 1	134	457	3,41	0,27	3,26	13	0,10	
Player 2	121	436	3,6	0,31	2,94	25	0,21	
Player 3	97	198	2,04	0,15	2,36	0	0,0	
Other	124	304	2,45	0,26	3,01	3	0,02	
<b>DinG10 Total</b>	476	1395			11,57	41		41,13
<b>All DinGs</b>	23575	121010				2528		701,80

Table A.4: Quantitative data, DinG corpus, DinG9 and DinG10, total.





# B

## An Abstract Categorical Grammar for *wh*-questions

## An Abstract Categorical Grammar for *wh*-questions

```
signature Abstract_syntax =

  n, np, s0, s : type ;

  farmer, donkey, meadow : n ;
  the : n -> np ;
  a, some, every : n -> (np -> s) -> s ;
  in : np -> s0 -> s0 ;
  fed, did_feed : np -> np -> s0 ;

  who : (np -> s) -> s ;
  which : n -> (np -> s) -> s ;
  where : ((s0 -> s0) -> s) -> s ;

  Eclos : s0 -> s ;

  Q : s -> s ;

end

signature Surface_form =

  o : type ;
  string = o -> o: type ;

  e = lambda x. x : string ;
  infix + = lambda x y z. x (y z) : string -> string -> string ;

  a, did, donkey, every, farmer, fed, feed, in, meadow, some, the,
  where, which, who : string ;

end

lexicon surface_form (Abstract_syntax) : Surface_form =

  n, np, s0, s := string ;

  farmer := farmer ;
  donkey := donkey ;
  meadow := meadow ;
  the := lambda x. the + x ;
  a := lambda x p. p (a + x) ;
  some := lambda x p. p (some + x) ;
  every := lambda x p. p (every + x) ;
  in := lambda x y. y + in + x ;
  fed := lambda x y. y + fed + x ;
```

---

```

did_feed := lambda x y. did + y + feed + x ;

who := lambda x. who + (x e) ;
which := lambda x y. which + x + (y e) ;
where := lambda q. where + (q (lambda x. x)) ;

Eclos := lambda x. x ;

Q := lambda x. x ;

end

signature Semantics =

  e, v, t : type ;

  farmer, donkey, meadow : e -> t ;
  the : (e -> t) -> e ;
  fed : v -> t ;
  agent, patient, location : v -> e -> t ;

  infix [Right] & : t -> t -> t ;
  infix [Right] > : t -> t -> t ;
  binder All : (e => t) -> t ;
  binder Ex : (e => t) -> t ;
  binder Exv : (v => t) -> t ;
  prefix ? : t -> t ;
  prefix ! : t -> t ;

end

lexicon semantics (Abstract_syntax) : Semantics =

  n := e -> t ;
  np := e ;
  s0 := v => t ;
  s := t ;

  farmer := farmer ;
  donkey := donkey ;
  meadow := meadow ;
  the := lambda x. the x ;
  a, some := lambda p q. ! (Ex x. (p x) & (q x)) ;
  every := lambda p q. All x. (p x) > (q x) ;
  in := lambda x p. Lambda e. (p e) & (location e x) ;
  fed := lambda x y. Lambda e. (fed e) & (patient e x) & (agent e y) ;
  did_feed := lambda x y. Lambda e. (fed e) & (patient e x)
  & (agent e y) ;

```

`who := lambda p. Ex x. p x ;`  
`which := lambda p q. Ex x. (p x) & (q x) ;`  
`where := lambda p. Ex x. p (lambda q. Lambda e. (q e)`  
`& (location e x)) ;`

`Eclos := lambda p. ! (Exv e. p e) ;`

`Q := lambda x. ? x ;`

end

signature Control =

`n, np, npe, npu, vp, s : type ;`

`the : n -> np ;`  
`a, some : n -> npe ;`  
`every : n -> npu ;`  
`Inpe : npe -> np ;`  
`Inpu : npu -> np ;`  
`farmer, donkey, meadow : n ;`  
`in : np -> vp -> vp ;`  
`fed : np -> vp ;`

`S : np -> vp -> s ;`

(\* np extraction \*)

`vp_np, vp_npe, vp_npu, s_npe, s_npu : type ;`

`fed1 : vp_np ;`  
`in1 : vp -> vp_np ;`

`in2 : np -> vp_np -> vp_np ;`  
`in3 : npe -> vp_np -> vp_npu ;`  
`in4 : npu -> vp_np -> vp_npe ;`

`S1 : npe -> vp_np -> s_npu ;`  
`S2 : npu -> vp_np -> s_npe ;`

`QR1 : npe -> vp_npe -> vp ;`  
`QR2 : npu -> vp_npu -> vp ;`  
`QR3 : npe -> s_npe -> s ;`  
`QR4 : npu -> s_npu -> s ;`

(\* double np extraction \*)

---

vp\_np\_np,  
s\_np\_npe, s\_np\_npu, s\_npe\_np, s\_npu\_np : type ;

in5 : vp\_np -> vp\_np\_np ;

S3 : npe -> vp\_np\_np -> s\_np\_npu ;  
S4 : npu -> vp\_np\_np -> s\_np\_npe ;  
S5 : npe -> vp\_np\_np -> s\_npu\_np ;  
S6 : npu -> vp\_np\_np -> s\_npe\_np ;

QR5 : npe -> s\_npe\_np -> s\_npe ;  
QR6 : npe -> s\_npe\_np -> s\_npu ;  
QR7 : npu -> s\_npu\_np -> s\_npe ;  
QR8 : npu -> s\_npu\_np -> s\_npu ;  
QR9 : npu -> s\_np\_npu -> s\_npe ;  
QR10 : npe -> s\_np\_npe -> s\_npu ;

(\* interrogative vp \*)

whnp\_s, whnp, vpq : type ;  
who : whnp\_s ;  
which : n -> whnp ;  
Inwh : whnp -> whnp\_s ;

inq : np -> vpq -> vpq ;  
did\_feed : np -> vpq ;

SQ1 : np -> vpq -> s ;  
SQ2 : whnp\_s -> vpq -> s ;

(\* object wh-movement \*)

vpq\_whnp, s\_whnp : type ;

did\_feed1 : vpq\_whnp ;

inq1 : np -> vpq\_whnp -> vpq\_whnp ;  
SQ3 : np -> vpq\_whnp -> s\_whnp ;  
WH1 : whnp -> s\_whnp -> s ;

(\* wh-adverb \*)

sq0 : type ;

where : sq0 -> s ;

SQ4 : np -> vpq -> sq0 ;

(\* np extraction – interrogative case \*)

vpq\_np, vpq\_npe, vpq\_npu, sq\_npe, sq\_npu : type ;

did\_feed2 : vpq\_np ;  
inq2 : vpq -> vpq\_np ;

inq3 : np -> vpq\_np -> vpq\_np ;  
inq4 : npe -> vpq\_np -> vpq\_npu ;  
inq5 : npu -> vpq\_np -> vpq\_npe ;

SQ5 : npe -> vpq\_np -> sq\_npu ;  
SQ6 : npu -> vpq\_np -> sq\_npe ;

QRq1 : npe -> vpq\_npe -> vpq ;  
QRq2 : npu -> vpq\_npu -> vpq ;  
QRq3 : npe -> sq\_npe -> s ;  
QRq4 : npu -> sq\_npu -> s ;

vpq\_np\_np, sq\_np\_npe, sq\_np\_npu, sq\_npe\_np, sq\_npu\_np : type ;

inq6 : vpq\_np -> vpq\_np\_np ;

SQ7 : npe -> vpq\_np\_np -> sq\_np\_npu ;  
SQ8 : npu -> vpq\_np\_np -> sq\_np\_npe ;  
SQ9 : npe -> vpq\_np\_np -> sq\_npu\_np ;  
SQ10 : npu -> vpq\_np\_np -> sq\_npe\_np ;

QRq5 : npe -> sq\_npe\_np -> sq\_npe ;  
QRq6 : npe -> sq\_npe\_np -> sq\_npu ;  
QRq7 : npu -> sq\_npu\_np -> sq\_npe ;  
QRq8 : npu -> sq\_npu\_np -> sq\_npu ;  
QRq9 : npu -> sq\_np\_npu -> sq\_npe ;  
QRq10 : npe -> sq\_np\_npe -> sq\_npu ;

vpq\_whnp\_np, vpq\_whnp\_npe, vpq\_whnp\_npu,  
s\_whnp\_npe, s\_whnp\_npu : type ;

inq7 : vpq\_whnp -> vpq\_whnp\_np ;

SQ11 : npe -> vpq\_whnp\_np -> s\_whnp\_npu ;  
SQ12 : npu -> vpq\_whnp\_np -> s\_whnp\_npe ;  
QRq11 : npu -> s\_whnp\_npu -> s\_whnp ;  
QRq12 : npe -> s\_whnp\_npe -> s\_whnp ;

sq0\_npu, sq0\_npe : type ;

SQ13 : npe -> vpq\_np -> sq0\_npu ;

---

```

SQ14  : npu -> vpq_np -> sq0_npe ;
QRq13 : npe -> sq0_npe -> sq0 ;
QRq14 : npu -> sq0_npu -> sq0 ;

```

```
end
```

```
lexicon abstract_syntax (Control) : Abstract_syntax =
```

```

n := n ;
np, npe, npu := (np -> s) -> s ;
vp := np -> (s0 -> s) -> s ;
s := s ;

the := lambda x k. k (the x) ;
a := a ;
some := some ;
every := every ;
Inpe, Inpu := lambda x. x ;
farmer := farmer ;
donkey := donkey ;
meadow := meadow ;
in := lambda p q x f. p (lambda y. q x (lambda s. f (in y s))) ;
fed := lambda p x f. p (lambda y. f (fed y x)) ;

S := lambda p q. p (lambda x. q x Eclos) ;

vp_np, vp_npe, vp_npu := np -> np -> (s0 -> s) -> s ;
s_npe, s_npu := np -> s ;

fed1 := lambda x y f. f (fed x y) ;
in1 := lambda p x y f. p y (lambda s. f (in x s)) ;

in2, in3, in4 := lambda p q x y f. p (lambda z. q x y
(lambda s. f (in z s))) ;

S1, S2 := lambda p q x. p (lambda y. q x y Eclos) ;

QR1, QR2 := lambda p q x f. p (lambda y. q y x f) ;
QR3, QR4 := lambda p q. p (lambda x. q x) ;

vp_np_np := np -> np -> np -> (s0 -> s) -> s ;
s_np_npe, s_np_npu, s_npe_np, s_npu_np := np -> np -> s ;

in5 := lambda p x y z f. p x z (lambda s. f (in y s)) ;

S3, S4, S5, S6 := lambda p q x y. p (lambda z. q x y z Eclos) ;

```



QR5, QR6, QR7, QR8 := lambda p q x. p (lambda y. q y x) ;  
 QR9, QR10 := lambda p q x. p (lambda y. q x y) ;  
  
 whnp\_s, whnp := (np -> s) -> s ;  
 vpq := np -> (s0 -> s) -> s ;  
  
 who := who ;  
 which := which ;  
 Inwh := lambda x. x ;  
 inq := lambda p q x f. p (lambda y. q x (lambda s. f (in y s))) ;  
 did\_feed := lambda p x f. p (lambda y. f (did\_feed y x)) ;  
  
 SQ1, SQ2 := lambda p q. Q (p (lambda x. q x Eclos)) ;  
  
 vpq\_whnp := np -> np -> (s0 -> s) -> s ;  
 s\_whnp := np -> s ;  
  
 did\_feed1 := lambda x y f. f (did\_feed x y) ;  
  
 inq1 := lambda p q x y f. p (lambda z. q x y (lambda s. f (in z s))) ;  
 SQ3 := lambda p q x. (p (lambda y. q x y Eclos)) ;  
 WHI := lambda p q. Q (p q) ;  
  
 sq0 := (s0 -> s) -> s ;  
  
 where := lambda p. Q (where (lambda f. p (lambda s. Eclos (f s)))) ;  
  
 SQ4 := lambda p q f. (p (lambda x. q x f)) ;  
  
 vpq\_np, vpq\_npe, vpq\_npu := np -> np -> (s0 -> s) -> s ;  
 sq\_npe, sq\_npu := np -> s ;  
  
 did\_feed2 := lambda x y f. f (did\_feed x y) ;  
 inq2 := lambda p x y f. p y (lambda s. f (in x s)) ;  
  
 inq3, inq4, inq5 := lambda p q x y f. p (lambda z. q x y  
 (lambda s. f (in z s))) ;  
  
 SQ5, SQ6 := lambda p q x. p (lambda y. q x y Eclos) ;  
  
 QRq1, QRq2 := lambda p q x f. p (lambda y. q y x f) ;  
 QRq3, QRq4 := lambda p q. Q (p (lambda x. q x)) ;  
  
 vpq\_np\_np := np -> np -> np -> (s0 -> s) -> s ;  
 sq\_np\_npe, sq\_np\_npu, sq\_npe\_np, sq\_npu\_np := np -> np -> s ;  
  
 inq6 := lambda p x y z f. p x z (lambda s. f (in y s)) ;

---

SQ7, SQ8, SQ9, SQ10 := lambda p q x y. p (lambda z. q x y z Eclos) ;

QRq5, QRq6, QRq7, QRq8 := lambda p q x. p (lambda y. q y x) ;  
QRq9, QRq10 := lambda p q x. p (lambda y. q x y) ;

vpq\_whnp\_np, vpq\_whnp\_npe, vpq\_whnp\_npu := np -> np -> np ->  
(s0 -> s) -> s ;  
s\_whnp\_npe, s\_whnp\_npu := np -> np -> s ;

inq7 := lambda p x y z f. p y z (lambda s. f (in x s)) ;

SQ11, SQ12 := lambda p q x y. p (lambda z. q x y z Eclos) ;  
QRq11, QRq12 := lambda p q x. (p (lambda y. q y x)) ;

sq0\_npu, sq0\_npe := np -> (s0 -> s) -> s ;

SQ13, SQ14 := lambda p q x f. p (lambda y. q x y f) ;

QRq13, QRq14 := lambda p q f. p (lambda x. q x f) ;

end



# Index

- Aboutness, 63  
ACG – Abstract Categorical Grammars, 39  
Alternative question, 61  
AMR – Abstract Meaning Representation, 31  
Anaphora, 24
- C-ORAL-ROM corpus, 90  
Catan, 99  
CFPP2000 – Spoken Parisian French Corpus, 89  
CLAPI – Corpus of Spoken Language in Interaction, 89  
Closed question, 60  
Common ground, 55  
Compositionality, 21  
Context, 23  
CR – Clarification Requests, 65  
CRFP – Reference Corpus of Spoken French, 91  
CSDS – Continuation Style Dynamic Semantics, 28, 34
- DAMSL – Dialog Act Markup in Several Layers, 54  
Dialogue, 25  
DinG – Dialogues in Games, 97  
DIT++: updated Dynamic Interpretation Theory, 54  
Donkey-sentences, 24  
DRT – Discourse Representation Theory, 26  
Dynamicity, 25
- ESLO – corpus of the sociolinguistic study in Orleans project, 94  
Event semantics, 41
- FCS – File Change Semantics, 26  
FQB – French QuestionBank, 94  
French interrogatives, 77  
Gricean maxims, 52
- Illocutionary force, 53  
Interrogatives, 25  
IS – Inquisitive Semantics, 30, 46, 68, 70  
KoS, 28, 56, 72
- Locutionary act, 53
- MRS – Minimal Recursion Semantics, 30  
MS – Montague Semantics, 26
- NDES – Neo-Davidsonian Event Semantics, 29, 42  
Negation, 23  
Negotiation phases, 166
- Possible worlds, 68  
Potentially resolving answer, 64
- Quantification, 23  
Quantificational event semantics, 41, 42
- Rhapsodie, 93
- S-BNC – Spoken British National Corpus, 86  
SCoSE – Saarbrücken Corpus of Spoken English, 86  
SDRT – Segmented Discourse Representation Theory, 27, 58, 71  
Semantics, 19  
Simple answer, 62  
SLAM – Schizophrenia and Language, Analysis and Modeling, 96  
Speech act, 53  
STAC – Strategic Conversation, 87  
Strongly exhaustive answer, 63  
Syntax, 17  
Syntax-semantics interface, 21
- TCOF – Treatment of Oral Corpora in French, 92  
Thematic roles, 42, 74

TTR – Type Theory with Records, 28

TUFS – Tokyo University of Foreign Studies  
corpus, 92

Wh-questions, 61, 69, 155

Yes/no-question, 60

# Bibliography

- [Aarts et al., 2014] Aarts, B., Chalker, S., and Weiner, E. (2014). *The Oxford dictionary of English grammar*. Oxford University Press.
- [Alama and Korbmacher, 2021] Alama, J. and Korbmacher, J. (2021). The Lambda Calculus. In Zalta, E. N., editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, Summer 2021 edition.
- [Amblard, 2007] Amblard, M. (2007). *Semantic representations and generative grammar: the Minimalist Categorical Grammars*. Theses, Université Sciences et Technologies - Bordeaux I.
- [Amblard et al., 2019] Amblard, M., Boritchev, M., Carletti, M., Dieudonat, L., and Tsai, Y. (2019). A taxonomy of real-life questions and answers in dialogue. In *SemDial 2019-LondonLogue-23rd Workshop on the semantics and pragmatics of dialogue*.
- [Amblard et al., 2015] Amblard, M., Fort, K., Demily, C., Franck, N., and Musiol, M. (2015). Analyse lexicale outillée de la parole transcrite de patients schizophrènes. *Traitement Automatique des Langues*, 55(3):91 – 115.
- [Amblard et al., 2014] Amblard, M., Fort, K., Musiol, M., and Rebuschi, M. (2014). L'impossibilité de l'anonymat dans le cadre de l'analyse du discours. In *Journée ATALA éthique et TAL*, Paris, France.
- [André and Canut, 2010] André, V. and Canut, E. (2010). Mise à disposition de corpus oraux interactifs: le projet TCOF (Traitement de Corpus Oraux en Français). *Pratiques. Linguistique, littérature, didactique*, (147-148):35–51.
- [Asher et al., 2016] Asher, N., Hunter, J., Morey, M., Benamara, F., and Afantenos, S. (2016). Discourse Structure and Dialogue Acts in Multiparty Dialogue: the STAC Corpus. In *10th International Conference on Language Resources and Evaluation (LREC 2016)*, pages 2721–2727, Portoroz, Slovenia.
- [Asher and Lascarides, 2003] Asher, N. and Lascarides, A. (2003). *Logics of conversation*. Cambridge University Press.
- [Asher and Pogodalla, 2010] Asher, N. and Pogodalla, S. (2010). SDRT and continuation semantics. In *JSAI International Symposium on Artificial Intelligence*, pages 3–15. Springer.
- [ATILF, 2020] ATILF (2020). TCOF : traitement de corpus oraux en français. ORTOLANG (Open Resources and TOols for LANGuage) –[www.ortolang.fr](http://www.ortolang.fr).
- [Austin, 1975] Austin, J. L. (1975). *How to do things with words*. Oxford university press.

- [Babonnaud, 2019] Babonnaud, W. (2019). A Topos-Based Approach to Building Language Ontologies. In Bernardi, R., Kobele, G., and Pogodalla, S., editors, *Formal Grammar. 24th International Conference, FG 2019*, volume 11668 of *Lecture notes in Computer Science*, pages 18–34, Riga, Latvia.
- [Badene et al., 2019] Badene, S., Thompson, K., Lorré, J.-P., and Asher, N. (2019). Weak supervision for learning discourse structure. In *EMNLP*.
- [Balthasar and Bert, 2005] Balthasar, L. and Bert, M. (2005). La plateforme « Corpus de langues parlées en interaction » (CLAPI). Historique, état des lieux, perspectives. *Lidil. Revue de linguistique et de didactique des langues*, (31):13–33.
- [Banarescu et al., 2013] Banarescu, L., Bonial, C., Cai, S., Georgescu, M., Griffitt, K., Hermjakob, U., Knight, K., Koehn, P., Palmer, M., and Schneider, N. (2013). Abstract meaning representation for sembanking. In *Proceedings of the 7th linguistic annotation workshop and interoperability with discourse*, pages 178–186.
- [Belnap, 1982] Belnap, N. D. (1982). Questions and answers in Montague grammar. In *Processes, beliefs, and questions*, pages 165–198. Springer.
- [Bender and Lascarides, 2019] Bender, E. M. and Lascarides, A. (2019). Linguistic Fundamentals for Natural Language Processing II: 100 Essentials from Semantics and Pragmatics. *Synthesis Lectures on Human Language Technologies*, 12(3):1–268.
- [Benthem and Minică, 2012] Benthem, J. V. and Minică, Ş. (2012). Toward a dynamic logic of questions. *Journal of Philosophical Logic*, 41(4):633–669.
- [Benzitoun et al., 2016] Benzitoun, C., Debaisieux, J.-M., and Deulofeu, H.-J. (2016). Le projet ORFÉO : un corpus d’étude pour le français contemporain. *Corpus*, (15).
- [Bernard and Champollion, 2018] Bernard, T. and Champollion, L. (2018). Negative events in compositional semantics. In *Semantics and Linguistic Theory*, volume 28, pages 512–532.
- [Bert et al., 2010] Bert, M., Bruxelles, S., Etienne, C., Jouin-Chardon, E., Lascar, J., Mondada, L., Teston, S., and Traverso, V. (2010). Grands corpus et linguistique outillée pour l’étude du français en interaction (plateforme CLAPI et corpus CIEL). *Pratiques. Linguistique, littérature, didactique*, (147-148):17–34.
- [Betarte and Tasistro, 1998] Betarte, G. and Tasistro, A. (1998). *Extension of Martin-Löf’s type theory with record types and subtyping*. Oxford Univ. Press.
- [Blanche-Benveniste and Jeanjean, 1987] Blanche-Benveniste, C. and Jeanjean, C. (1987). *Le français parlé : transcription et édition*. Didier érudition.
- [Bonial et al., 2011] Bonial, C., Corvey, W., Palmer, M., Petukhova, V. V., and Bunt, H. (2011). A hierarchical unification of LIRICS and VerbNet semantic roles. In *2011 IEEE Fifth International Conference on Semantic Computing*, pages 483–489. IEEE.
- [Bonial et al., 2020] Bonial, C., Donatelli, L., Abrams, M., Lukin, S., Tratz, S., Marge, M., Artstein, R., Traum, D., and Voss, C. (2020). Dialogue-AMR: Abstract Meaning Representation for Dialogue. In *Proceedings of The 12th Language Resources and Evaluation Conference*, pages 684–695.

- 
- [Bonial et al., 2019] Bonial, C., Donatelli, L., Lukin, S., Tratz, S., Artstein, R., Traum, D., and Voss, C. (2019). Augmenting abstract meaning representation for human-robot dialogue. In *Proceedings of the First International Workshop on Designing Meaning Representations*, pages 199–210.
- [Boritchev and Amblard, 2019] Boritchev, M. and Amblard, M. (2019). A compositional view of questions. In *WiNLP - Widening NLP - ACL Workshop*, Florence, Italy.
- [Boritchev and Amblard, 2020] Boritchev, M. and Amblard, M. (2020). There is as yet Insufficient Data for a Meaningful Answer. In *SemDial - WatchDial The 24th Workshop on the Semantics and Pragmatics of Dialogue*, Brandeis, United States.
- [Boritchev and Amblard, 2021] Boritchev, M. and Amblard, M. (2021). Picturing questions and answers—a formal approach to slam. In *(In) coherence of Discourse*, pages 65–89. Springer.
- [Boritchev and de Groote, 2020] Boritchev, M. and de Groote, P. (2020). On dialogue modeling: a dynamic epistemic inquisitive approach. In *LENLS17 : Logic & Engineering of Natural Language Semantics*, Online, Japan.
- [Bos, 2011] Bos, J. (2011). A survey of computational semantics: Representation, inference and knowledge in wide-coverage text understanding. *Language and Linguistics Compass*, 5(6):336–366.
- [Bos et al., 1994] Bos, J., Mastenbroek, E., MacGlashan, S., Millies, S., and Pinkal, M. (1994). A compositional DRS-based formalism for NLP applications.
- [Branca-Rosoff et al., 2009] Branca-Rosoff, S., Fleury, S., Lefevre, F., and Pires, M. (2009). Discours sur la ville. Corpus de Français Parlé Parisien des années 2000.
- [Breitholtz, 2014] Breitholtz, E. (2014). *Enthymemes in Dialogue: A micro-rhetorical approach*. PhD thesis, Göteborgs universitet. Humanistiska fakulteten University of Gothenburg. Faculty of Arts.
- [Breitholtz, 2020] Breitholtz, E. (2020). *Enthymemes in Dialogue*. Brill.
- [Breitholtz et al., 2021] Breitholtz, E., Cooper, R., Howes, C., and Lavelle, M. (2021). Reasoning in multiparty dialogue involving patients with schizophrenia. In *(In) coherence of Discourse*, pages 43–63. Springer.
- [Bunt, 1994] Bunt, H. (1994). Context and dialogue control. *Think Quarterly*, 3(1):19–31.
- [Bunt, 2009] Bunt, H. (2009). The DIT++ taxonomy for functional dialogue markup. In *AAMAS 2009 Workshop, Towards a Standard Markup Language for Embodied Dialogue Acts*, pages 13–24.
- [Bunt et al., 2012] Bunt, H., Alexandersson, J., Choe, J.-W., Fang, A. C., Hasida, K., Petukhova, V., Popescu-Belis, A., and Traum, D. R. (2012). Iso 24617-2: A semantically-based standard for dialogue annotation. In *LREC*, pages 430–437.
- [Burnard, 2000] Burnard, L. (2000). Reference guide for the british national corpus (world edition).
- [Canavan and Zipperlen, 1996] Canavan, A. and Zipperlen, G. (1996). CALLFRIEND Spanish-Non-Caribbean Dialect LDC96S58.



- [Carcassi and Szymanik, 2021] Carcassi, F. and Szymanik, J. (2021). ‘Most’vs ‘More Than Half’: An Alternatives Explanation. *Proceedings of the Society for Computation in Linguistics*, 4(1):334–343.
- [Carnie, 2012] Carnie, A. (2012). *Syntax: A generative introduction*, volume 18. John Wiley & Sons.
- [Champollion, 2011] Champollion, L. (2011). Quantification and negation in event semantics.
- [Champollion, 2015] Champollion, L. (2015). The interaction of compositional semantics and event semantics. *Linguistics and Philosophy*, 38(1):31–66.
- [Champollion, 2017] Champollion, L. (2017). *Parts of a whole: Distributivity as a bridge between aspect and measurement*, volume 66. Oxford University Press.
- [Champollion et al., 2019] Champollion, L., Bumford, D., and Henderson, R. (2019). Donkeys under discussion. *Semantics and Pragmatics*, 12:1.
- [Champollion et al., 2015] Champollion, L., Ciardelli, I., and Roelofsen, F. (2015). Some questions in typed inquisitive semantics. In *Talk at workshop on Questions in Logic and Semantics, Amsterdam*.
- [Champollion et al., 2017] Champollion, L., Ciardelli, I., and Roelofsen, F. (2017). On questions and presuppositions in typed inquisitive semantics. In *Handout for the talk given at the 2nd workshop on Inquisitiveness Below and Beyond the Sentence Boundary (InqBnB)*.
- [Champollion et al., 2007] Champollion, L., Tauberer, J., and Romero, M. (2007). The Penn Lambda Calculator: Pedagogical software for natural language semantics.
- [Chui and Lai, 2008] Chui, K. and Lai, H.-I. (2008). The NCCU corpus of spoken chinese: Mandarin, Hakka, and Southern Min. *Taiwan Journal of Linguistics*, 6(2).
- [Church, 1940] Church, A. (1940). A formulation of the simple theory of types. *The journal of symbolic logic*, 5(02):56–68.
- [Ciardelli et al., 2018] Ciardelli, I., Groenendijk, J., and Roelofsen, F. (2018). *Inquisitive semantics*. Oxford University Press.
- [Ciardelli et al., 2017] Ciardelli, I., Roelofsen, F., and Theiler, N. (2017). Composing alternatives. *Linguistics and Philosophy*, 40(1):1–36.
- [Ciardelli and Roelofsen, 2015] Ciardelli, I. A. and Roelofsen, F. (2015). Inquisitive dynamic epistemic logic. *Synthese*, 192(6):1643–1687.
- [Cohen, 1960] Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and psychological measurement*, 20(1):37–46.
- [Cooper, 2005a] Cooper, R. (2005a). Austinian truth, attitudes and type theory. *Research on Language and Computation*, 3(2-3):333–362.
- [Cooper, 2005b] Cooper, R. (2005b). Records and record types in semantic theory. *Journal of Logic and Computation*, 15(2):99–112.

- 
- [Cooper, 2008] Cooper, R. (2008). Type theory with records and unification-based grammar. *Logics for Linguistic Structures*, 201(9).
- [Cooper and Ginzburg, 2012] Cooper, R. and Ginzburg, J. (2012). Negative inquisitiveness and alternatives-based negation. In *Logic, Language and Meaning*, pages 32–41. Springer.
- [Cooper and Larsson, 1999] Cooper, R. and Larsson, S. (1999). Dialogue moves and information states. In *Third International Workshop on Computational Semantics (IWCS3)*.
- [Copestake, 2007] Copestake, A. (2007). Semantic composition with (robust) minimal recursion semantics. In *ACL 2007 Workshop on Deep Linguistic Processing*, pages 73–80.
- [Copestake et al., 2005] Copestake, A., Flickinger, D., Pollard, C., and Sag, I. A. (2005). Minimal recursion semantics: An introduction. *Research on Language and Computation*, 3(2-3):281–332.
- [Core and Allen, 1997] Core, M. G. and Allen, J. (1997). Coding dialogs with the DAMSL annotation scheme. In *AAAI fall symposium on communicative action in humans and machines*, volume 56, pages 28–35. Boston, MA.
- [Cresti et al., 2004] Cresti, E., do Nascimento, F. B., Moreno-Sandoval, A., Veronis, J., Martin, P., and Choukri, K. (2004). The C-ORAL-ROM CORPUS. A Multilingual Resource of Spontaneous Speech for Romance Languages. In *LREC*. Citeseer.
- [Crocco and Savy, 2003] Crocco, C. and Savy, R. (2003). API-Archivio di Parlato Italiano (final DVD of the API Project).
- [Cruz Blandon et al., 2019] Cruz Blandon, M. A., Minnema, G., Nourbakhsh, A., Boritchev, M., and Amblard, M. (2019). Toward Dialogue Modeling: A Semantic Annotation Scheme for Questions and Answers. In *Proceedings of the 13th Linguistic Annotation Workshop (LAW XIII)*.
- [Damonte et al., 2017] Damonte, M., Cohen, S. B., and Satta, G. (2017). An incremental parser for abstract meaning representation. In *Proceedings of EACL*.
- [Davidson, 1967] Davidson, D. (1967). The logical form of action sentences.
- [de Groote, 2001] de Groote, P. (2001). Towards abstract categorial grammars. In *Proceedings of the 39th Annual Meeting on Association for Computational Linguistics*, pages 252–259. Association for Computational Linguistics.
- [de Groote, 2006] de Groote, P. (2006). Towards a montagovian account of dynamics. *Proceedings of semantics and linguistic theory XVI*.
- [de Groote and Winter, 2014] de Groote, P. and Winter, Y. (2014). A type-logical account of quantification in event semantics. In Murata, T., Mineshima, K., and Bekki, D., editors, *LENLS, JURISIN, and GABA, Kanagawa, Japan, Revised Selected Papers*, volume 9067 of *Lecture Notes in Computer Science*, pages 53–65. Springer.
- [De Saussure, 1972] De Saussure, F. (1972). Cours de linguistique générale (1916). *Edition critique préparée par T. de mauro. paris: payothèque*.
- [Delic, 2004] Delic, É. (2004). Présentation du corpus de référence du français parlé. *Recherches sur le français parlé*, 18:11–42.

- [Demner-Fushman and Lin, 2007] Demner-Fushman, D. and Lin, J. (2007). Answering clinical questions with knowledge-based and statistical techniques. *Computational Linguistics*, 33(1):63–103.
- [Dotlačil and Roelofsen, 2019] Dotlačil, J. and Roelofsen, F. (2019). Dynamic inquisitive semantics: anaphora and questions. In *Proceedings of Sinn und Bedeutung*, volume 23, pages 365–382.
- [Dowty et al., 2012] Dowty, D. R., Wall, R., and Peters, S. (2012). *Introduction to Montague semantics*, volume 11. Springer Science & Business Media.
- [Egg, 1998] Egg, M. (1998). Wh-questions in underspecified minimal recursion semantics. *Journal of Semantics*, 15(1):37–82.
- [Eshkol-Taravella et al., 2011] Eshkol-Taravella, I., Baude, O., Maurel, D., Hriba, L., Dugua, C., and Tellier, I. (2011). Un grand corpus oral « disponible »: le corpus d’Orléans 1 1968-2012. *Traitement automatique des langues*, 53(2):17–46.
- [Fernández et al., 2007] Fernández, R., Ginzburg, J., and Lappin, S. (2007). Classifying non-sentential utterances in dialogue: A machine learning approach. *Computational Linguistics*, 33(3):397–427.
- [Freed, 1994] Freed, A. F. (1994). The form and function of questions in informal dyadic conversation. *Journal of Pragmatics*, 21(6):621 – 644.
- [Frege, 1892] Frege, G. (1892). Über sinn und bedeutung. *Zeitschrift für Philosophie und philosophische Kritik*, 100:25–50.
- [Gallin, 1975] Gallin, D. (1975). *Intensional and Higher-Order Modal Logic*, volume 19 of *Mathematics Studies*. North-Holland.
- [Gamut, 1991] Gamut, L. (1991). *Logic, language and meaning* vol. 1.
- [Gawne and McCulloch, 2019] Gawne, L. and McCulloch, G. (2019). Emoji as digital gestures. *Language@ Internet*, 17(2).
- [Geach, 1962] Geach, P. T. (1962). *Reference and generality*. Cornell University Press.
- [Geurts, 2002] Geurts, B. (2002). Donkey business. *Linguistics and philosophy*, 25(2):129–156.
- [Ginzburg, 1996] Ginzburg, J. (1996). Dynamics and the semantics of dialogue. In Seligman, J. and Westerstahl, D., editors, *Logic, language and computation*. CSLI Publications, Stanford.
- [Ginzburg, 2012] Ginzburg, J. (2012). *The interactive stance*. Oxford University Press.
- [Ginzburg, 2016] Ginzburg, J. (2016). Semantics of dialogue. *The Cambridge Handbook of Formal Semantics*.
- [Ginzburg and Purver, 2012] Ginzburg, J. and Purver, M. (2012). Quantification, the reprise content hypothesis, and type theory. *From Quantification to Conversation. Festschrift for Robin Cooper on the occasion of his 65th birthday*. College Publications, London.
- [Ginzburg and Sag, 2000] Ginzburg, J. and Sag, I. (2000). *Interrogative investigations*. Stanford: CSLI publications.

- 
- [Girard, 1987] Girard, J.-Y. (1987). Linear logic. *Theoretical computer science*, 50(1):1–101.
- [Green, 2020] Green, M. (2020). Speech Acts. In Zalta, E. N., editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, Winter 2020 edition.
- [Grice, 1975] Grice, H. P. (1975). Logic and conversation. *Syntax and Semantics*, page 3:41?58.
- [Groenendijk, 2008] Groenendijk, J. (2008). Inquisitive semantics and dialogue pragmatics. *Rutgers lecture notes, available via [www.illc.uva.nl/inquisitive-semantics](http://www.illc.uva.nl/inquisitive-semantics)*.
- [Groenendijk and Stokhof, 1991] Groenendijk, J. and Stokhof, M. (1991). Dynamic predicate logic. *Linguistics and philosophy*, pages 39–100.
- [Groenendijk et al., 1997] Groenendijk, J., Stokhof, M., Van Benthem, J., and ter Meulen, A. (1997). Handbook of logic and language.
- [Groenendijk and Stokhof, 1984] Groenendijk, J. A. G. and Stokhof, M. J. B. (1984). *Studies on the Semantics of Questions and the Pragmatics of Answers*. PhD thesis, Univ. Amsterdam.
- [Grouin et al., 2015] Grouin, C., Griffon, N., and Névéal, A. (2015). Is it possible to recover personal health information from an automatically de-identified corpus of French EHRs? In *Proceedings of the Sixth International Workshop on Health Text Mining and Information Analysis*, pages 31–39.
- [Guillaume, 2021] Guillaume, B. (2021). Graph matching and graph rewriting: Grew tools for corpus exploration, maintenance and conversion. In *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: System Demonstrations*, pages 168–175.
- [Hackl, 2009] Hackl, M. (2009). On the grammar and processing of proportional quantifiers: most versus more than half. *Natural language semantics*, 17(1):63–98.
- [Hamblin, 1970] Hamblin, C. L. (1970). *Fallacies*. Methuen, London.
- [Hamblin, 1976] Hamblin, C. L. (1976). Questions in Montague English. In *Montague grammar*, pages 247–259. Elsevier.
- [Harris, 1954] Harris, Z. S. (1954). Distributional structure. *Word*, 10(2-3):146–162.
- [Heim, 1982] Heim, I. (1982). The semantics of definite and indefinite noun phrases.
- [Heim, 2012] Heim, I. (2012). *File change semantics and the familiarity theory of definiteness*. de Gruyter.
- [Hintikka, 1983] Hintikka, J. (1983). New foundations for a theory of questions and answers. In *Questions and answers*, pages 159–190. Springer.
- [Holle and Rein, 2013] Holle, H. and Rein, R. (2013). The modified Cohen’s kappa: Calculating interrater agreement for segmentation and annotation. *Understanding Body Movement: A Guide to Empirical Research on Nonverbal Behaviour, H. Lausberg, Ed. Frankfurt am Main: Peter Lang Verlag*, pages 261–277.
- [Holle and Rein, 2015] Holle, H. and Rein, R. (2015). EasyDIAG: A tool for easy determination of interrater agreement. *Behavior research methods*, 47(3):837–847.

- [Hough et al., 2015] Hough, J., Kennington, C., Schlangen, D., and Ginzburg, J. (2015). Incremental semantics for dialogue processing: Requirements and a comparison of two approaches. In *Proceedings of the 11th International Conference on Computational Semantics (IWCS) 2015*.
- [Howes et al., 2014] Howes, C., Purver, M., and McCabe, R. (2014). Linguistic indicators of severity and progress in online text-based therapy for depression. *ACL 2014*, 7.
- [Hüning and Schlücker, 2015] Hüning, M. and Schlücker, B. (2015). Multi-word expressions. In *Volume 1 Word-Formation*, pages 450–467. De Gruyter Mouton.
- [Janssen and Zimmermann, 2021] Janssen, T. M. V. and Zimmermann, T. E. (2021). Montague Semantics. In Zalta, E. N., editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, Summer 2021 edition.
- [Judge et al., 2006] Judge, J., Cahill, A., and Van Genabith, J. (2006). Questionbank: Creating a corpus of parse-annotated questions. In *Proceedings of the 21st International Conference on Computational Linguistics and 44th Annual Meeting of the Association for Computational Linguistics*, pages 497–504.
- [Jurafsky and Martin, 2000] Jurafsky, D. and Martin, J. H. (2000). *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*. Prentice Hall PTR, Upper Saddle River, NJ, USA, 1st edition.
- [Jurafsky and Martin, 2015] Jurafsky, D. and Martin, J. H. (2015). *Speech and Language Processing*.
- [Kamp, 1981] Kamp, H. (1981). A theory of truth and semantic representation. *Formal semantics – the essential readings*, pages 189–222.
- [Kamp and Reyle, 2013] Kamp, H. and Reyle, U. (2013). *From discourse to logic: Introduction to modeltheoretic semantics of natural language, formal logic and discourse representation theory*, volume 42. Springer Science & Business Media.
- [Kanazawa, 1994] Kanazawa, M. (1994). Weak vs. strong readings of donkey sentences and monotonicity inference in a dynamic setting. *Linguistics and philosophy*, 17(2):109–158.
- [Karttunen, 1976] Karttunen, L. (1976). Discourse referents. In *Notes from the linguistic underground*, pages 363–385. Brill.
- [Karttunen, 1977] Karttunen, L. (1977). Syntax and semantics of questions. *Linguistics and philosophy*, 1(1):3–44.
- [Kingsbury and Palmer, 2002] Kingsbury, P. R. and Palmer, M. (2002). From TreeBank to PropBank. In *LREC*, pages 1989–1993. Citeseer.
- [Klein and Manning, 2003] Klein, D. and Manning, C. D. (2003). Accurate unlexicalized parsing. In *Proceedings of the 41st annual meeting of the association for computational linguistics*, pages 423–430.
- [Klein et al., 2003] Klein, D., Manning, C. D., et al. (2003). Fast exact inference with a factored model for natural language parsing. *Advances in neural information processing systems*, pages 3–10.

- 
- [Krifka, 2001] Krifka, M. (2001). Quantifying into question acts. *Natural language semantics*, 9(1):1–40.
- [Lacheret-Dujour et al., 2019] Lacheret-Dujour, A., Kahane, S., and Pietrandrea, P. (2019). *Rhapsodie: A prosodic and syntactic treebank for spoken French*, volume 89. John Benjamins Publishing Company.
- [Landis and Koch, 1977] Landis, J. R. and Koch, G. G. (1977). The measurement of observer agreement for categorical data. *biometrics*, pages 159–174.
- [Lappin and Fox, 2015] Lappin, S. and Fox, C. (2015). *The handbook of contemporary semantic theory*. John Wiley & Sons.
- [Lascarides and Asher, 2003] Lascarides, A. and Asher, N. (2003). Imperatives in dialogue. *Pragmatics and Beyond New Series*, pages 1–24.
- [Lebedeva, 2012] Lebedeva, E. (2012). *Expressing discourse dynamics through continuations*. PhD thesis.
- [Leidner and Plachouras, 2017] Leidner, J. L. and Plachouras, V. (2017). Ethical by design: Ethics best practices for natural language processing. In *Proceedings of the First ACL Workshop on Ethics in Natural Language Processing*, pages 30–40.
- [Lihoreau and Rebuschi, 2014] Lihoreau, F. and Rebuschi, M. (2014). *Epistemology, Context, and Formalism*. Springer.
- [Liu and Chen, 2019] Liu, Z. and Chen, N. (2019). Reading turn by turn: Hierarchical attention architecture for spoken dialogue comprehension. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 5460–5466.
- [Love et al., 2017] Love, R., Dembry, C., Hardie, A., Brezina, V., and McEnery, T. (2017). The Spoken BNC2014: Designing and building a spoken corpus of everyday conversations. *International Journal of Corpus Linguistics*, 22(3):319–344.
- [Maraev et al., 2018] Maraev, V., Ginzburg, J., Larsson, S., Tian, Y., and Bernardy, J.-P. (2018). Towards KoS/TTR-based proof-theoretic dialogue management. *Proceedings of SemDial*.
- [Maršik, 2016] Maršik, J. (2016). *Effects and handlers in natural language*. PhD thesis, Université de Lorraine.
- [McCulloch, 2019] McCulloch, G. (2019). *Because internet: Understanding how language is changing*. Random House.
- [Montague, 1970] Montague, R. (1970). Universal grammar. *Theoria*, 36(3):373–398.
- [Montague, 1973] Montague, R. (1973). The proper treatment of quantification in ordinary English. In *Approaches to natural language*, pages 221–242. Springer.
- [Montague and Thomason, 1975] Montague, R. and Thomason, R. H. (1975). Formal philosophy. Selected papers of Richard Montague. *Erkenntnis*, 9(2).
- [Muskens, 1996] Muskens, R. (1996). Combining Montague semantics and discourse representation. *Linguistics and philosophy*, 19(2):143–186.

- [Nasreen et al., 2019] Nasreen, S., Purver, M., and Hough, J. (2019). A corpus study on questions, responses and misunderstanding signals in conversations with Alzheimer’s patients. In *Proceedings of the 23rd Workshop on the Semantics and Pragmatics of Dialogue-Full Papers. SEMDIAL, London, United Kingdom (Sep 2019)*, <http://semdial.org/anthology/Z19-Nasreensemdial>, volume 13.
- [Norrick, 2017] Norrick, N. (2017). SCoSE part 1: Complete conversations. English Linguistics, Department of English at Saarland University.
- [Oostdijk, 2001] Oostdijk, N. (2001). The design of the Spoken Dutch Corpus. *Language and Computers*, 36:105–112.
- [Pariße and Liégeois, 2020] Pariße, C. and Liégeois, L. (2020). Utiliser les outils CORLI de conversion TEI pour l’analyse de corpus de langage oral. In *6e conférence conjointe Journées d’Études sur la Parole (JEP, 33e édition), Traitement Automatique des Langues Naturelles (TALN, 27e édition), Rencontre des Étudiants Chercheurs en Informatique pour le Traitement Automatique des Langues (RÉCITAL, 22e édition). Volume 4: Démonstrations et résumés d’articles internationaux*, pages 64–65. ATALA; AFCP.
- [Parsons, 1990] Parsons, T. (1990). *Events in the Semantics of English*, volume 5. MIT press Cambridge, MA.
- [Partee, 1984] Partee, B. (1984). Compositionality. *Varieties of formal semantics*, 3:281–311.
- [Peldszus et al., 2012] Peldszus, A., Buß, O., Baumann, T., and Schlangen, D. (2012). Joint satisfaction of syntactic and pragmatic constraints improves incremental spoken language understanding.
- [Peters and Westerstähl, 2013] Peters, S. and Westerstähl, D. (2013). The semantics of possessives. *Language*, pages 713–759.
- [Pogodalla, 2016] Pogodalla, S. (2016). ACGTK: un outil de développement et de test pour les grammaires catégorielles abstraites.
- [Pollard and Sag, 1994] Pollard, C. and Sag, I. A. (1994). *Head-driven phrase structure grammar*. University of Chicago Press.
- [Purver, 2006] Purver, M. (2006). Clarie: Handling clarification requests in a dialogue system. *Research on Language and Computation*, 4(2-3):259–288.
- [Purver et al., 2003] Purver, M., Ginzburg, J., and Healey, P. (2003). On the means for clarification in dialogue. In *Current and new directions in discourse and dialogue*, pages 235–255. Springer.
- [Purver, 2004] Purver, M. R. J. (2004). *The theory and use of clarification requests in dialogue*. PhD thesis, University of London.
- [Qian et al., 2017] Qian, J., Du, H., Hou, J., Chen, L., Jung, T., Li, X.-Y., Wang, Y., and Deng, Y. (2017). Voicemask: Anonymize and sanitize voice input on mobile devices. *arXiv preprint arXiv:1711.11460*.
- [Qian, 2014] Qian, S. (2014). *Accessibility of Referents in Discourse Semantics*. PhD thesis, Université de Lorraine.

- 
- [Rebuschi et al., 2013] Rebuschi, M., Amblard, M., and Musiol, M. (2013). Schizophrénie, logique et compréhension en première personne. *L'Évolution Psychiatrique*, 78(1):127–141.
- [Riegel et al., 1994] Riegel, M., Pellat, J.-C., and Rioul, R. (1994). Grammaire méthodique du français. *Linguistique nouvelle*.
- [Roulet et al., 1987] Roulet, E., Auchlin, A., Moeschler, J., and Rubattel, C. (1987). *L'articulation du discours en français contemporain*, volume 11. Peter Lang GmbH, Internationaler Verlag Der Wissenschaften.
- [Roussarie, 2017] Roussarie, L. (2017). *Sémantique formelle: Volume 1: Introduction à la grammaire de Montague*. Language Science Press.
- [Schegloff et al., 1974] Schegloff, E., Jefferson, G., and Sacks, H. (1974). A simplest systematics for the organization of turn-taking for conversation. *Language*, 50(4):696–735.
- [Schlangen, 2020] Schlangen, D. (2020). An outline of a model of situated discourse representation and processing. In *Proceedings of semdial 2020 (WatchDial)*, Brandeis University / Internet.
- [Seddah and Candito, 2016] Seddah, D. and Candito, M. (2016). Hard time parsing questions: Building a questionbank for French. In *Tenth International Conference on Language Resources and Evaluation (LREC 2016)*.
- [Senft, 2009] Senft, G. (2009). Phatic communion. In Senft, G., Östman, J.-O., and Verschueren, J., editors, *Culture and language use*, pages 226–233. Amsterdam/Philadelphia.
- [Siyanova-Chanturia et al., 2017] Siyanova-Chanturia, A., Conklin, K., Caffarra, S., Kaan, E., and van Heuven, W. J. (2017). Representation and processing of multi-word expressions in the brain. *Brain and language*, 175:111–122.
- [Stalnaker, 2002] Stalnaker, R. (2002). Common ground. *Linguistics and philosophy*, 25(5-6):701–721.
- [Stolcke et al., 2000] Stolcke, A., Ries, K., Coccaro, N., Shriberg, E., Bates, R., Jurafsky, D., Taylor, P., Martin, R., Ess-Dykema, C. V., and Meteer, M. (2000). Dialogue act modeling for automatic tagging and recognition of conversational speech. *Computational linguistics*, 26(3):339–373.
- [Taulé et al., 2015] Taulé, M., Martí, M. A., Bies, A., Nofre, M., Garí, A., Song, Z., Strassel, S., and Ellis, J. (2015). Spanish treebank annotation of informal non-standard web text. In *International Conference on Web Engineering*, pages 15–27. Springer.
- [Traum et al., 1999] Traum, D., Bos, J., Cooper, R., Larsson, S., Lewin, I., Matheson, C., and Poesio, M. (1999). A model of dialogue moves and information state revision. Technical report. Task Oriented Instructional Dialogue project, Deliverable D2.1.
- [Van Ditmarsch et al., 2007] Van Ditmarsch, H., van Der Hoek, W., and Kooi, B. (2007). *Dynamic epistemic logic*, volume 337. Springer Science & Business Media.
- [Veldhoen and Zuidema, 2017] Veldhoen, S. and Zuidema, W. (2017). Can neural networks learn logical reasoning? *CLASP Papers in Computational Linguistics*, page 34.



- [Westerståhl, 2019] Westerståhl, D. (2019). Generalized Quantifiers. In Zalta, E. N., editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, Winter 2019 edition.
- [Winter and Zwarts, 2011] Winter, Y. and Zwarts, J. (2011). Event semantics and Abstract Categorical Grammar. In Kanazawa, M., Kornai, A., Kracht, M., and Seki, H., editors, *Proceedings of Mathematics of Language, MOL12*, volume 6878 of *Lecture Notes in Artificial Intelligence, LNAI*, pages 174–191, Berlin. Springer-Verlag.
- [Wittenburg et al., 2006] Wittenburg, P., Brugman, H., Russel, A., Klassmann, A., and Sloetjes, H. (2006). ELAN: a professional framework for multimodality research. In *5th International Conference on Language Resources and Evaluation (LREC 2006)*, pages 1556–1559.
- [Yao and Zhang, 2010] Yao, X. and Zhang, Y. (2010). Question generation with minimal recursion semantics. In *Proceedings of QG2010: The Third Workshop on Question Generation*, pages 68–75. Citeseer.
- [Zamaraeva, 2021] Zamaraeva, O. (2021). *Assembling Syntax: Modeling Constituent Questions in a Grammar Engineering Framework*. PhD thesis, University of Washington.

## Résumé

L'étude formelle du discours soulève de nombreuses interrogations liées à la nature et à la définition des phrases et de la manière dont une suite de phrases s'articule pour former un discours cohérent. Le langage est intrinsèquement dynamique : dans sa sémantique en contexte (par exemple, l'utilisation de références) et dans l'interaction (par exemple, les liens entre les actes de dialogue). Le passage du discours au dialogue donne lieu à des questions plus spécifiques en particulier liées à la relation entre les questions et les réponses. Afin d'aborder ces thématiques, nous nous concentrons sur la sémantique des questions.

Il existe de nombreux formalismes et cadres de travail pour la sémantique formelle des phrases déclaratives et du discours ; le dialogue, pour sa part, est largement étudié d'un point de vue linguistique et traitement automatique des langues. L'objectif de notre travail est d'utiliser les théories classiques de sémantique formelle dans un cadre orienté vers le dialogue réel. Cette thèse présente une formalisation sémantique du dialogue dans une théorie dynamique des types simples.

Nous produisons des modèles du dialogue, et en particulier de l'articulation des questions et des réponses, en mêlant la Neo-Davidsonian Event Semantics (NDES, telle que présentée dans (Champollion, 2017)) à la Inquisitive Semantics (IS, Ciardelli et al, 2018) de manière compositionnelle et dynamique à travers l'usage de la Continuation Style Dynamic Semantics (CSDS, de Groote, 2006, étendue dans Lebedeva, 2012). Notre modèle est ancré dans une implémentation d'interface syntaxe-sémantique appelée Abstract Categorical Grammars (ACG, de Groote, 2001).

Une autre façon d'aborder la sémantique du dialogue est de s'intéresser aux données réelles, ce qui permet de mettre en perspective nos idées formelles et de les confronter aux observations. Pour ce faire, nous avons constitué un corpus, appelé Dialogues in Games (DinG), composé de transcriptions d'enregistrements de personnes jouant au jeu de société Catane (en français). Nos études, centrées sur les questions et les réponses dans des données orales multilingues (anglais, français, néerlandais, espagnol mexicain, italien du nord et mandarin), a donné lieu à plusieurs schémas d'annotation, dont une partie a été appliquée à DinG.

**Mots-clés:** sémantique formelle, linguistique computationnelle, dialogue, questions

## Abstract

Formal studies of discourse raise numerous interrogations on the nature and the definition of the way consecutive sentences coherently combine with one another. Language is intrinsically dynamic: in its semantics in context (e.g. use of references) and in the interaction (e.g. connections between dialogue acts). The shift from discourse to dialogue brings forward even more specific issues among which the ones related to questions and answers articulation. In order to address these issues, we start by focusing on questions from a semantic point of view.

There are numerous existing formalisms and frameworks for formal semantics of declarative sentences and discourse; dialogue is broadly studied from a linguistic and Natural Language Processing point of view. The goal of our work is to bring classical formal semantics theories to use in a setting oriented towards real-life dialogue. This thesis presents a formal semantics account of dialogue in a type-theoretical dynamic framework.

We produce a model of dialogue and in particular of the articulation of questions and answers by mingling Neo-Davidsonian Event Semantics (NDES, as presented in Champollion, 2017) with Inquisitive Semantics (IS, Ciardelli et al, 2017) in a compositional and dynamic way through the use of Continuation Style Dynamic Semantics (CSDS, de Groote, 2006, extended in Lebedeva, 2012). Our model is rooted in a syntax-semantics interface implementation called Abstract Categorical Grammars (ACG, de Groote, 2001).

Another way to address semantics of dialogue is through real-life data by putting our formal ideas in perspective by confronting them with field observation. To do so, we constituted a corpus, called Dialogues in Games (DinG), composed of transcriptions of recordings of people playing the board game Catan (in French). Our study focuses on real-life questions and answers in multilingual (English, French, Dutch, Mexican Spanish, Northern Italian, and Mandarin Chinese) oral data has resulted in several annotation schema, part of which was applied to DinG.

**Keywords:** formal semantics, computational linguistics, dialogue, questions

