

# **Schema Architecture for Language-Vision InterActions**

A Computational Cognitive Neuroscience Model of  
Language Use

by

Victor Barrès

under the direction of  
Dr. Michael A. Arbib



A Dissertation Presented to the

FACULTY OF USC GRADUATE SCHOOL

of the

UNIVERSITY OF SOUTHERN CALIFORNIA

In Partial Fulfillment of the Requirements for the Degree of

**DOCTOR OF PHILOSOPHY  
(NEUROSCIENCE)**

August 2017

## Acknowledgements

The final result of my graduate work is surely a testament of a time spent exploring regions of human knowledge that to this day continue to fill me with wonder and in which I feel as in a new home.

But if I had to answer the question of what my proudest achievement is, it would not be what I have produced. Rather it would be the fact that I managed to learn about and build connections between fields of research that were largely unknown to me when I first stepped into the graduate school arena: computational neuroscience, psycholinguistics, cognitive science, computer science, and vision science.

This achievement I owe it to first and foremost to my advisor Dr. Michael Arbib. Never have I met a person so eager to constantly tackle new intellectual challenges and dedicated to push his students to explore the roads less traveled. Venturing so often off the beaten path has more than once been a source of worries, but I have always felt reassured, knowing that I had someone with a sure scientific compass guiding me.

Working at the crossroads of multiple field, one has to be humble and ask for help engaging with new disciplines. I owe it to my committee to have provided me with this help: the chair Dr. Lauren Itti, Dr. Aziz Zadeh, Dr. Elsi Kaiser, and Dr. Toben Mintz. They have all, at one point, seen me, a student with no particular background in the field they are expert in, come to them. Each of them accepted to patiently teach and guide me through the difficult problems among which they navigate daily. My committee meetings, those opportunities to be challenged by such a diverse set of mentors, have been among the most fruitful intellectual experiences I have had.

I want to thank Dr. Luc Steels for making it possible for me to visit his Fluid Construction Grammar team and learn from their many years of experience in the field of computational construction grammar, in particular through Dr. Michael Spranger and Dr. Remi Van Trijp. Dr. Peter Dominey similarly invited me to visit his laboratory and I cannot thank him enough for all the time we have spent discussing the computational construction grammar and its relations to neural organization and processes.

I am of course greatly indebted to Dr. Jinyong Lee who not only jump started the work on Template Construction Grammar but also took the time to patiently teach me about his research, allowing me to continue what he had started.

My lab mates Dr. James Bonaiuto, Dr. Brad Gasser, Dr. Nader Noori, Rob Schuler, and Dr. Vesna Gamez-Djokic have provided with tremendous help and constructive criticisms. In particular, I cannot express enough how grateful I am that I have been able to work alongside a friend, Brad, for all these years.

Thanks to Arthur III Bray-Simons for stumbling one day into the lab. His relentless enthusiasm, curiosity, and capacity to master new topics, made the work on Synthetic ERP possible.

I owe to my friend Dr. David Colliaux to have been introduced to cognitive science and to my friend Dr. Geoffroy Hermann to have decided to try and work with Dr. Arbib. Their help, guidance, and friendship has been invaluable in the years that it took to achieve this work.

Thanks to my friend Dr. Priyanka Biswas for patiently answering my questions about the field of linguistics.

I have to end by thanking my family for their unconditional support. But in particular, I want to dedicate this work to my wife Lauren, for always believing in me.

To Lauren and Madeleine.

# Contents

<b>1 Computational Cognitive Neuroscience of Language</b>	<b>9</b>
1.1 A Brain that Makes Meaningful Use of Language in Context: a Dynamic Distributed System	9
1.1.1 Psycholinguistics and the Visual World Paradigm: Incremental Constraint Satisfaction over Multiple Knowledge Sources . . . . .	9
1.1.2 Neurolinguistics: Studying an Adaptive and Organized Distributed System and its Degradations . . . . .	11
1.2 NeuroCognitive Modeling: A (very) Brief Overview . . . . .	13
1.2.1 Computational Modeling of Neuro-Cognitive Systems: A Look from the Language Processing Perspective . . . . .	13
1.2.2 Vision-for-Action: Integration of Bottom-Up and Top-Down Visual Attention Signal .	15
1.3 Grammatical Processes Beyond Syntax: Computational Construction Grammar . . . . .	20
1.3.1 Cognitive Linguistics . . . . .	20
1.3.2 Embodied Construction Grammar . . . . .	20
1.3.3 Fluid Construction Grammar . . . . .	22
1.3.4 Dynamic Construction Grammar . . . . .	24
1.4 Schema Theory: System Level Cognitive Modeling Framework and Dynamic Cooperative Computation. . . . .	25
1.4.1 Schema Theory: A Brain Theory Framework to Model Organisms' behaviors within Action-Perception Cycles. . . . .	25
1.4.2 VISIONS: A Schema Theoretic Model of Scene Interpretation. . . . .	28
1.4.3 Dynamic Cooperative Computation: A Core Principle of Cognitive Modeling . . . . .	31
1.4.4 Schema Theory: From VISIONS to COAST . . . . .	32
1.5 From Gaze to Speech: the Template Construction Grammar Research Program . . . . .	34
1.6 Outline of the Thesis . . . . .	34
<b>2 SALVIA: An Implemented Schema-Theoretic Framework for Investigating the Linkage of Vision and Language Production</b>	<b>37</b>
2.1 Introduction . . . . .	37
2.2 The Schema Architecture Language-Vision InterAction (SALVIA) Cognitive Model . . . . .	39
2.2.1 From VISIONS to SALVIA: Towards a Schema Theory Approach to Language Processing	39
2.2.2 SALVIA: Schema Architecture . . . . .	41
2.3 From Perceptual to Linguistic Meaning via Conceptualization . . . . .	45
2.3.1 Conceptual Knowledge . . . . .	50
2.3.2 Long Term Memories: Various Types of Knowledge . . . . .	52
2.4 Visual Attention . . . . .	53
2.4.1 Subscenes: A Cognitive-Level Scene Structuring Principle . . . . .	53
2.4.2 Subscene Recognition . . . . .	53
2.4.3 Attentional Parsing: Building and Navigating Hierarchical Scene Representations .	54
2.5 Grammatical Processing . . . . .	54
2.5.1 Template Construction Grammar (TCG) . . . . .	54
2.5.2 Grammatical WM . . . . .	57

2.5.3	From Construction Assemblages to Utterances: Phonological WM and Utterance Production . . . . .	59
2.5.4	Linguistic WM: Hierarchy but no Tree . . . . .	60
2.5.5	Good Enough Production of Utterances: Speaker and Task Relevant Parameters . . . . .	60
2.6	Language-Vision Interactions: Visual Guidance vs. Verbal Guidance . . . . .	62
2.7	Language Production Schema System: Modeling Language Use . . . . .	66
2.7.1	Incrementality at its Core . . . . .	66
2.7.2	Seeing-for-Saying . . . . .	66
2.8	Preliminary Conclusions . . . . .	69
<b>3</b>	<b>From Gaze Patterns to Utterances: Simulating the Dynamics of Visual Scene Description</b>	<b>70</b>
3.1	Introduction . . . . .	70
3.2	Input-Outputs . . . . .	70
3.2.1	Inputs . . . . .	70
3.2.2	Outputs . . . . .	73
3.3	Parameter Space . . . . .	74
3.4	From Incremental Semantic Representation to Utterances . . . . .	75
3.4.1	Sanity Check . . . . .	75
3.4.2	From Conceptual to Computational Example . . . . .	75
3.4.3	Increasing the Complexity of the Message: Impact of Time Pressure . . . . .	77
3.5	Scene Description: Interaction between Visual Attention and Language Processes . . . . .	78
3.5.1	SALVIA: States . . . . .	78
3.5.2	General Example: From Eye Movements to Utterance Production . . . . .	81
3.5.3	Saliency, Perceptual Guidance and Information Structure . . . . .	82
3.6	Simulating Key Visual World Paradigm Psycholinguistic Results . . . . .	86
3.6.1	Time Pressure and Utterance Fragmentation . . . . .	86
3.6.2	Perceptual Guidance, Verbal Guidance, and Cognitive Thresholds . . . . .	90
3.6.3	Saliency, Perceptual Guidance and Information Structure: Impact of Saliency on the Use of Active vs. Passive Construction. . . . .	94
3.7	Conclusion . . . . .	94
3.7.1	Summary of Results: SALVIA as a Cognitive Model of Language Production . . . . .	94
3.7.2	Summary of Results: SALVIA as a Psycholinguistic Model . . . . .	95
3.7.3	Representation and Semantics of Complex Visual Scenes . . . . .	96
<b>4</b>	<b>Template Construction Grammar (TCG): Formalism for Dynamic Grammatical Processing of Incrementally Built Semantic Representations.</b>	<b>99</b>
4.1	Introduction . . . . .	99
4.2	System-Level View . . . . .	100
4.2.1	A Schema-Theoretic Model of Language Production . . . . .	100
4.2.2	Schema Theory and Cooperative Computation: What You Need To Know . . . . .	101
4.3	Incremental and Dynamic Semantic Representation (SemRep) . . . . .	101
4.4	Grammatical Processing . . . . .	102
4.4.1	Template Construction Grammar: Language Representations as Templates of Meaning-Form Mapping . . . . .	102
4.4.2	Language Schemas . . . . .	105
4.4.3	Dynamic Grammatical Processing of Incrementally Built Semantic Representations . .	106
4.4.4	Construction Schema Instantiation: SemMatch . . . . .	106
4.4.5	Cooperative Computation (C2): Match . . . . .	110
4.4.6	Generating Form . . . . .	118
4.4.7	Linguistic WM . . . . .	121
4.5	Good Enough Production of Utterances: Speaker and Task Relevant Parameters . . . . .	121
4.6	Conclusion . . . . .	123

<b>5 SALVIA: Toward a Neuro-Cognitive Model of Normal and Agrammatic Language Comprehension</b>	<b>124</b>
5.1 SALVIA as Neurally Informed Model of Comprehension . . . . .	124
5.1.1 Comprehension Patterns of Agrammatic Aphasics . . . . .	125
5.1.2 Light and Heavy Semantics . . . . .	126
5.2 Dynamic Interactions of World Knowledge and Linguistic Information during Language Comprehension . . . . .	127
5.2.1 Lessons from Neurolinguistics: A Two-Route Model for the Processing of Linguistic Inputs . . . . .	127
5.3 Semantic WM: Incremental and Dynamic Semantic Representation (SemRep) . . . . .	129
5.4 From Form to Meaning via Grammar: The Grammatical Route (GR) (Step1) . . . . .	131
5.4.1 Template Construction Grammar as a Schema-Theoretic Model of Grammatical Processing for Incremental Language Comprehension . . . . .	131
5.4.2 A Look at a More Complex Example (Limited Prediction) . . . . .	131
5.5 From Form to Meaning via World Knowledge: the World (event) Knowledge Route (WKR) (Step2) . . . . .	133
5.5.1 From Incremental Linguistic Input to Incrementally Built Semantic Representations: Case of World Knowledge Route Only . . . . .	133
5.6 Multi-Route Concurrent Incremental Processing of Verbal Input: C2 Between Routes (Step3) . . . . .	135
5.6.1 Semantic WM as a Locus of Cooperative Computation . . . . .	135
5.7 Conceptual Account of Agrammatic Comprehension Performances in Sentence-Picture Matching Tasks . . . . .	137
5.8 Conclusion . . . . .	141
<b>6 TCG-SALVIA: Formalism for Incremental and Dynamic Grammatical Processing of Utterances.</b>	<b>143</b>
6.1 Introduction . . . . .	143
6.2 Grammatical Route: Cooperative-Computation Driven Generation of Dynamic Construction-Based Form-Meaning Mappings . . . . .	144
6.2.1 Template Construction Grammar as a Schema-Theoretic Model of Grammatical Processing for Incremental Language Comprehension . . . . .	144
6.2.2 Phonological WM . . . . .	144
6.2.3 Grammatical Knowledge Representation . . . . .	144
6.2.4 Dynamic Grammatical Processing: from Incremental Linguistic Input to Incrementally Built Semantic Representations . . . . .	145
6.2.5 Generating Meaning . . . . .	150
6.3 World (event) Knowledge Route (WKR): Generating Dynamic Frame-Based, Pragmatically Motivated, Form-Meaning Mapping . . . . .	150
6.3.1 Phonological WM: Interaction with World Knowledge . . . . .	151
6.3.2 WK Event Frame Schemas . . . . .	151
6.3.3 WK Frame Retrieval . . . . .	151
6.3.4 WK Processing . . . . .	151
6.3.5 Generating Meaning . . . . .	151
6.4 Dynamic Interactions Between Grammatical And World Knowledge Route Routes as Cooperative Computation Between Concurrent Processes . . . . .	153
6.4.1 Good Enough Comprehension of Utterances: Speaker and Task Relevant Parameters .	153
6.4.2 Semantic WM C2 Dynamics: The Key Role of Route Weights . . . . .	153
6.4.3 Dynamic Coordination of Multiple Information Sources: Sub-Multigraph Isomorphisms	154
6.5 Input-Output . . . . .	156
6.6 From Conceptual To Computational Examples . . . . .	156
6.6.1 Sanity Check . . . . .	156
6.6.2 Simulation 1: Grammatical Route Only . . . . .	157
6.6.3 Simulation 2: World Knowledge Route Only (+ Lexical Constructions) . . . . .	158
6.6.4 Simulation 3: Cooperation Between Routes . . . . .	160

6.6.5	Simulation 4: Competition Between Routes . . . . .	162
6.7	SALVIA: Simulating the Agrammatic Aphasics Comprehension Performances . . . . .	164
6.7.1	Agrammatic Comprehension: A Novel Interpretation . . . . .	164
6.8	Discussion . . . . .	166
6.8.1	Towards a Full SALVIA Model: Linking Vision, Production, and Comprehension . . . . .	166
6.8.2	Model Comparisons . . . . .	169
6.8.3	Future challenges . . . . .	171
<b>7</b>	<b>Synthetic Event-Related Potentials: A Computational Bridge Between Neurolinguistic Models and Experiments</b>	<b>173</b>
7.1	Linking Computational Models to Brain Data . . . . .	173
7.2	Background . . . . .	174
7.2.1	fMRI and Synthetic Brain Imaging . . . . .	174
7.2.2	Event-Related Potentials: A privileged window into how the brain processes language	174
7.2.3	Synthetic ERP in comparison to Dynamic Causal Modeling and other ERP modeling approaches . . . . .	175
7.3	“The 2002 Model”: Friederici’s 2002 Model of Auditory Sentence Processing . . . . .	180
7.3.1	A basic feedforward model of the timing and localization of processes involved in integrating the auditory form of a word into the comprehension of a sentence . . . . .	180
7.3.2	Data on functional anatomy . . . . .	182
7.4	The Two Phases of Synthetic ERP: A Preliminary Computational Framework . . . . .	184
7.4.1	A preliminary framework . . . . .	184
7.4.2	The challenge of timing data for the 2002 model . . . . .	185
7.5	Phase 2 in Detail: From Areas of Cortical Activity to ERPs . . . . .	187
7.5.1	Forward modeling in the Synthetic ERP framework . . . . .	187
7.5.2	Head model . . . . .	187
7.5.3	Forward model and lead field computation . . . . .	189
7.5.4	Processing model . . . . .	189
7.6	Simulation Results . . . . .	192
7.6.1	Mapping the 2002 model onto cortical geometry . . . . .	192
7.6.2	Processing model: Brain modules and activity timing . . . . .	193
7.6.3	Simulation results 1: Scalp potential topographic maps . . . . .	194
7.6.4	Simulation results 2: Synthetic ERPs . . . . .	196
7.7	From Preliminary Results to Emerging Challenges . . . . .	200
7.7.1	Source dipole modeling and cortical geometry . . . . .	200
7.7.2	Activation modeling . . . . .	200
7.7.3	Quantitative ERP data extraction from literature . . . . .	200
<b>8</b>	<b>Toward a Neurocomputational Model: The Challenge of Brain Anchoring.</b>	<b>202</b>
8.1	An Attempt . . . . .	202
8.2	Quantitatively Linking Phenomenological Models and Computational Models: Toward Synthetic Brain Imaging Approach to Neuro-Computational Modeling of Language Processes . .	209
8.2.1	Testing a Neural Network Model Against ERP Data Using a Forward Approach: Modeling Requirements . . . . .	209
8.2.2	Models of Language Processing . . . . .	213
8.3	The Elusive Role of Broca’s Area in Comprehension and Production . . . . .	215
8.3.1	Definition . . . . .	215
8.3.2	Syntax . . . . .	216
8.3.3	Verbal Working Memory and Articulatory Rehearsal . . . . .	217
8.3.4	Cognitive Control . . . . .	218
8.3.5	Multi-Stream Integration . . . . .	219
8.3.6	Interaction Between Production and Comprehension in Broca’s Area . . . . .	219
8.4	The Neuroscience of Semantic Processing: A Frontier . . . . .	220
8.4.1	Overview of the Problem . . . . .	220

8.4.2	Neural Organization of the Semantic System . . . . .	221
8.4.3	Embodiment, Disembodiment, Weak-embodiment . . . . .	224
8.4.4	Access vs. Composition: Problems of Linguistic Semantic Processing . . . . .	226
8.4.5	Computational Theories . . . . .	226
8.4.6	Conclusion . . . . .	227
<b>A</b>	<b>Cognitive Architecture Schema Theory (COAST): Formalism and Implementation</b>	<b>248</b>
A.1	Overview . . . . .	248
A.2	Model as System-of-Systems (SoS) . . . . .	248
A.3	Long Term Memory System . . . . .	249
A.3.1	Schema . . . . .	249
A.3.2	Long Term Memory as Schema network . . . . .	250
A.3.3	Schema instance & Schema instantiation . . . . .	250
A.4	Working Memory & Cooperative Computation . . . . .	250
A.4.1	Overview . . . . .	250
A.4.2	C2_links . . . . .	251
A.4.3	WM State . . . . .	252
A.4.4	WM processes . . . . .	252
A.4.5	Cooperative Computation Dynamics . . . . .	252
A.4.6	Assemblage . . . . .	253
<b>B</b>	<b>SALVIA Production Simulations</b>	<b>254</b>
B.1	Parameter Space and Default Values . . . . .	254
B.2	SemRep Incremental Input . . . . .	256
B.3	Complex SemRep Incremental Input . . . . .	257
B.4	Simulations . . . . .	257
B.4.1	Simulation 1 . . . . .	257
B.4.2	Simulation 2 . . . . .	263
B.4.3	Simulation 3 . . . . .	274
B.4.4	Simulation 4 . . . . .	281
B.4.5	Simulation 5 . . . . .	287
B.5	Scene Builder . . . . .	293
<b>C</b>	<b>TCG Production Theory</b>	<b>294</b>
C.1	TCG Production Processing Algorithms . . . . .	294
C.1.1	SemMatch and Construction Invocation . . . . .	294
C.1.2	Match . . . . .	296
C.1.3	Cooperative Computation Dynamics . . . . .	298
C.1.4	Construction Schema Instance Assemblage . . . . .	298
C.1.5	Incremental Semantic Representation Format (ISRF) . . . . .	298
<b>D</b>	<b>TCG-SALVIA Comprehension</b>	<b>302</b>
D.1	Left-Corner Parser for Context Free Grammars . . . . .	302
D.1.1	Context Free Grammar and Chart Notations . . . . .	302
D.1.2	Left-Corner Relation . . . . .	302
D.1.3	Left-Corner Parser . . . . .	302
<b>E</b>	<b>SALVIA Comprehension Simulations</b>	<b>304</b>
E.1	Simulation 0: Grammatical Route Only . . . . .	304
E.1.1	WM states . . . . .	304
E.1.2	Simulation Summaries . . . . .	307
E.2	Simulation 1: Grammatical Route Only . . . . .	308
E.2.1	WM states . . . . .	309
E.2.2	Simulation Summaries . . . . .	311
E.3	Simulation 2: World Knowledge Route Only (+ Lexical Constructions) . . . . .	311

E.3.1	WM states . . . . .	313
E.3.2	Simulation Summaries . . . . .	316
E.4	Simulation 3: Cooperation Between Routes . . . . .	317
E.4.1	WM states . . . . .	317
E.4.2	Simulation Summaries . . . . .	319
E.5	Simulation 4: Competition Between Routes . . . . .	321
E.5.1	WM states . . . . .	321
E.5.2	Simulation Summaries . . . . .	324
<b>F</b>	<b>Synthetic ERP: Mathematical, Physical and Computational Foundations</b>	<b>327</b>
F.1	General physical Formulation of the Forward Problem . . . . .	327
F.2	Dipole Modeling of Current Sources . . . . .	329
F.3	Conductor Modeling: the Head Model . . . . .	333
F.4	Algebraic Formulation . . . . .	336
F.5	Numerical Method: Boundary Element Method (BEM) . . . . .	338
<b>G</b>	<b>Abstract Constructions</b>	<b>341</b>

# Chapter 1

## Computational Cognitive Neuroscience of Language

*“Overcoming the formalist point of view also led to a change in the concept of structure. Structures can no longer be conceived as formal assemblages of symbolic elements connected by means of formal relations. They are now conceived as natural, organic, qualitatively self-organized and dynamically regulated wholes, as forms, Gestalts or patterns. The perspective is now organizational, dynamical, and emergental: structures emerge from substrata, be they internal (neuronal) or external, while symbolic discrete and sequential structures formally described by the classical paradigm are now equated with qualitative, structurally stable and invariant structures emerging from an underlying dynamics.”*

Petitot

Morphogenesis of Meaning

### 1.1 A Brain that Makes Meaningful Use of Language in Context: a Dynamic Distributed System

#### 1.1.1 Psycholinguistics and the Visual World Paradigm: Incremental Constraint Satisfaction over Multiple Knowledge Sources

For a while, the equation of core language processing to syntactic operations embedded in a strongly modular and competence-oriented perspective dictated the type of empirical studies and modeling that prevailed in the analysis of the functioning human language system. Stemming from both a long tradition of linguistic (Bloomfield, 1962; Chomsky, 1995, 2002) and cognitive (Fodor, 1983) theories, this so-called generative approach at first provided a strong paradigm against which to interpret empirical results. However it has become clear in the past decades that this fundamentally analytic approach, narrowly focused on cutting through the complex human language apparatus to carve out the core language organ, left untouched the equally complex problem of empirically and theoretically understanding how syntactic operation - and more generally linguistic operations - could be performed by and integrated within a body with specific sensory-motor interfaces to a physical and social world. Even within the generativist tradition the question of interfaces between linguistic and other cognitive systems became central (Jackendoff, 2002).

From an empirical perspective, the desire to quantitatively tap into the interactions between language and sensory-motor systems led to the development of a new line of psycholinguistic work using the Visual World Paradigm (VWP) (Huettig et al., 2011) in which subjects produce or comprehend utterances while they are presented with a related visual scene. Attention-related eye movements are recorded throughout the task, thus offering a way to link the time series of attentional focuses to the time series defined by the produced or perceived utterances. The results from this approach replaced a view of the language system in

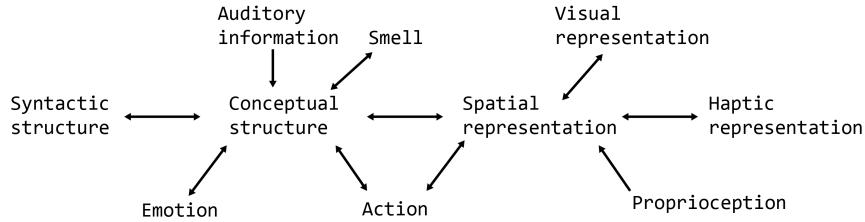


Figure 1.1: Informal model of the language processes anchored within a network of interfaces with sensorimotor systems (adapted from Jackendoff, 1997). Beyond the question of whether or not this informal model is correct or not (which is an empirical one), from a computational perspective, if one accepts that the data is at least strong enough to suggest that this type of complex network of interfaces and processes are the background against which linguistic operations are carried out, the challenge becomes immediately that of understanding how the behaviors of such a system come to be orchestrated to fit our communicative goals? Understanding these issues is the core of computational brain theory and therefore should be at the core of the computational cognitive neuroscience of language.

which syntax processing was a standalone module by a view of the language system functioning essentially as an incremental constraint-satisfaction system dynamically and opportunistically incorporating data provided by multiple knowledge sources that include linguistic, perceptual, and conceptual knowledge. Mayberry et al. (2006) summarize the key results of VWP as follows:

“First, on-line comprehension occurs incrementally and is closely time-locked with attention to the scene (Tanenhaus et al., 1995). Second, attention to objects in a scene before they are mentioned in an utterance shows that anticipation plays a vital role in comprehension (Altmann and Kamide, 1999). Third, all available information sources - linguistic and world knowledge, as well as scene information are rapidly and seamlessly integrated during online comprehension (Kamide et al., 2003; Knoeferle et al., 2005; Sedivy et al., 1999; Tanenhaus et al., 1995). Fourth, sentence comprehension is highly adaptive to the dynamic availability of information from these multiple sources. Fifth, these sources of information are coordinated: the interaction between language and visual scene processing is a two-way street.”

These five points are integrated in the best developed conceptual model of situated language comprehension, the Coordinated Interplay Account (CIA)(Knoeferle and Crocker, 2006). It is worth insisting and illustrating the online and opportunistic multi-source satisficing process at play during situated language processing. Altmann and Kamide (1999) showed how subjects incrementally combine linguistic, visual, and world knowledge information during a comprehension task. Faced with a visual display showing a cake, two toys, and a boy, the subjects were faster to fixate the cake upon hearing the verb “eat” in the sentence “the boy will eat ...” than upon hearing the more general verb “move” in the sentence “the boy will move ...”. The verb “move” imposes less constraints on the post-verbal patient (i.e., theme) that could be the cake or any of the two other toys, than the verb “eat” for which “cake” is the only plausible patient. The fact that such constraints can influence eye movements as soon as the verb is perceived shows the incremental nature of the combined impact of world knowledge and visual information on language processing.

Addressing the criticism that such an effect could simply be driven by the verb and not by true integration of linguistic, visual, and world knowledge information, Kamide et al. (2003) used a similar set up but, as in the reading time study described above, used a combination of agent and verb to constrain the possible post-verbal patient and its visual referent. Upon hearing the sentence “the man will ride ...”, the subjects made more anticipatory saccades to the visually presented motorcycle than to the merry-go-round, with the opposite pattern being measured when the sentence was “the girl will ride ...”. Tapping into production is empirically much harder but some important results have nevertheless been derived by analyzing descriptions of a visual scene produced by speakers and linking those to the way they attentionally parse a visual scene.

In addition, by manipulating the saliency landscape of the scene or by cueing attentional focus to specific regions of the visual scene, the VWP offers a unique opportunity to study the relations between the ways a scene is cognitively structured by attentional parsing and the types of constructions that are used to describe

it. Gleitman et al. (2007) showed how perceptual priming impacted syntactic structural choices made during the production of visual scene description. “Referential priming”, in which a scene referent is presented for observation before the rest of the scene is made visible, has also been investigated (Myachykov et al., 2011). Both cases outlined the links that exist between the temporal unfolding of attention, information structures and ultimately construction choices during production of visual scene descriptions. These results have led to incremental views of the system in which scene comprehension and sentence formulation occur concurrently mutually influencing each-other. However, a structural view has also been put forward that suggests that scene “apprehension precedes formulation” (Griffin and Bock, 2000) with the sentential structure determined by the conceptual structure built from a scene rather than direct perceptual prominence of individual perceptual items (Bock et al., 2004; Griffin and Bock, 2000) The two views seemingly describe mutually exclusive principles. However, language production may generally involve both an incremental and a preplanning mechanism (e.g. Levelt, 1993)s, and the production system may shift between these mechanisms based on the perceived information (Brown-Schmidt and Tanenhaus, 2006).

### **1.1.2 Neurolinguistics: Studying an Adaptive and Organized Distributed System and its Degradations**

Psycholinguistic behavioral results offer a unique window into the temporal dynamics of the complex interactions that weave together linguistic, perceptuo-motor, and cognitive systems in normal subjects. However, they make no attempt at linking the processes they describe to the underlying neural hardware. Their focus is on defining functional operations, the type of representations they rest upon, and the type functional structure that organizes them. It is neurolinguistics that attempts to address the question of how functional processes relate to neural architectures.

When faced with complex biological systems supporting cognitive functions, the question of the patterns of degradation the systems exhibit following lesions or perturbation remains both a unique source of information on the system’s organization and a healthy reminder than brain models need to be able to not only account for the healthy subjects performances but also display the graceful degradation and resilience patterns that are displayed by neurological patients.

#### **Neuropsychology**

Broca’s aphasics (expressive, or non-fluent aphasia) suffer from brain lesions that result in disfluent asyntactic speech usually lacking grammatical words, verbal inflections etc. (Gleason et al., 1975; Goodglass and Berko, 1960; Goodglass, 1968, 1976; Kean, 1977). Importantly, Caramazza and Zurif (1976) found that Broca’s aphasics were no different than normal subjects when asked to match a picture with canonical active sentences such as “the lion is chasing the fat tiger”, but were no better than chance for center-embedded object relatives such as “the tiger that the lion is chasing is fat”. However, performances of Broca’s aphasics was restored to the level of normal subjects for object relatives when world knowledge cues were available to constrain the sentence interpretation as in “The apple that the boy is eating is red”. This latter result led the authors to hypothesize a neuropsychological dissociation between two comprehension processes: a “heuristic” system based primarily on world knowledge information and an “algorithmic” system relying mainly on syntactic information. Sherman and Schweickert (1989) replicated the experiment while controlling for the possible combinations of syntactic cues, world knowledge plausibility and, importantly, picture plausibility. The conclusion of Caramazza and Zurif (1976) regarding the role world knowledge plays alongside syntax is largely admitted as a non-controversial empirical fact confirmed by subsequent studies (Ansoll and Flowers, 1982; Kudo, 1984; Saffran et al., 1998; Sherman and Schweickert, 1989).

Since this seminal work was published, it has been shown that the comprehension performances of agrammatic aphasics appear quite heterogeneous. The very notion that agrammatism reflects the impairment of an identifiable function of a syntactic system (as in the case of the Trace Deletion Hypothesis of Grodzinsky (2000)) is strongly challenged by the diversity of comprehension performances. In their meta-analysis of 15 studies published between 1980 and 1993 that reported agrammatic aphasics’ comprehension performances on sentence-picture matching tasks and included contrasts between active and passive constructions, Berndt et al. (1996) found that the 64 unique data sets (for 42 patients) could be clustered into three groups of approximately equal size, each reflecting a distinct comprehension pattern: (1) only active constructions

are comprehended better than chance, (2) both active and passive constructions are comprehended better than chance, (3) both structures are comprehended no better than chance. So far none of the theories linking agrammatism to a specific deficit in syntax processing has been able to account for this variety in performances. Rather than conclude that agrammatism does not constitute a useful neuropsychological syndrome for the understanding of the neural and cognitive structure of the language system (Caramazza et al., 2005) I suggest that this diverse set of data provides a good target for a new neurocomputational approach. The main counterpoint of Broca's aphasia and agrammatism is the case of Wernicke's aphasics who display mostly grammatical albeit nonsensical speech with very poor comprehension. As pointed out by Grodzinsky et al. (1999) and by Zurif and Piñango (1999), it is only when agrammatism for comprehension is associated to a Broca's aphasia diagnostic that it can be associated to specific lesion patterns, and that those lesions patterns should clearly differ from those predicted by a diagnosis of Wernicke's aphasia. This points to a necessity to try and account not only for agrammatism in comprehension but for the relations that entertain production and reception deficits.

## **Neuroimaging**

From the perspective of neuroimaging, far less is known about the language production system than about that of the language comprehension system. Neuroimaging studies, in great part due to the limitations of the tools they employ, have focused on comprehension. Far less is known about the language production system from this empirical perspective. Only a few key results about comprehension are presented here to set the stage. The past ten years have seen many studies highlighting the fact that the language system coarsely defined as the left perisylvian area was in fact composed of multiple anatomical and functional pathways that run between pSTS and Broca's areas. At least two different DTI studies demonstrated the existence of two ventral pathways (Anwander et al., 2007; Saur et al., 2008). Those two DTI studies suggest the existence of a single dorsal pathway but Catani et al. (2005) reported the existence of two different dorsal pathways with one connecting pSTS and PTr directly while the other connects pSTS and PO through IPL. Glasser and Rilling (2008) also reported two dorsal pathways connecting STG to PO and MTG to PTr respectively. The specific functions of these ventral and dorsal pathways are unknown, but there is a tendency to assign semantic processing to the ventral pathways and syntactic processing to the dorsal pathways (and possibly verbal working memory to one of the dorsal pathways) (Friederici, 2011).

The role of at least one dorsal path in direct sensory-motor coding involved in word repetition is well assessed and supported by both our and others' theoretical account of language processing (Arbib, 2010; Hickok and Poeppel, 2004; Saur et al., 2008). Recent work on primary progressive aphasia supported the role of the ventral path in semantic processing but also pointed to a role of the dorsal path in syntactic processing (Wilson et al., 2010a,b). A way to reconcile these two views on the role of dorsal path could be to assign the role of sensory-motor mapping to the arcuate fasciculus (AF) connecting pSTG to BA6 while the syntactic processing would rely on the superior lateral fasciculus (SLF) connection of pSTG to BA44 possibly through IPL.

This is consistent with the view considering that AF could be evolutionary older supporting the initial capacity to perform sensory-motor mappings necessary to bootstrap the parity required in symbolic communication while the SLF through IPL and connecting Wernicke area and BA44 through IPL, which are all brain regions that underwent considerable evolution between human and the non-human primates, supports the uniquely human grammatical capacities. Empirical advances in understanding the general connectivity pattern within the language system support the idea of that language rests on distributed computation in a multi-stream architecture similar to that of the visual system (how/what dissociation). This division between syntactic/heuristic and semantic streams or routes is also widely reflected in EEG results and models that have focused on the analysis of "semantic P600" and related work on semantic illusions (Bornkessel and Schlesewsky, 2006; Brouwer et al., 2012; Kim and Osterhout, 2005; Kos et al., 2010; Kuperberg et al., 2007; Nieuwland, M.S. and Berkum, J.J.A., 2005; van Herten et al., 2005).

## 1.2 NeuroCognitive Modeling: A (very) Brief Overview

### 1.2.1 Computational Modeling of Neuro-Cognitive Systems: A Look from the Language Processing Perspective

At the symbolic level, the **U-Space model** (U for Unification) of Vosse and Kempen (2000) uses an implemented version of a lexicalist grammar based on a formalism similar to that of tree-adjoining grammars (Joshi and Schabes, 1997). Based on tree-unification processes piloted by a process of dynamic competition and cooperation, the model is able to simulate both core psycholinguistic empirical results in normal subjects (effects of syntactic complexity, local and global syntactic ambiguity, lexical ambiguity) and agrammatics' comprehension performance results (Caplan et al., 1985) that focus purely on their capacity to process syntactic cues, but focusing on inter-patient variability rather than global averages. Moreover, the U-space model has been used as a core computational piece of one of the landmark neurolinguistic conceptual model of language comprehension, the **MUC** (Memory, Unification, Control) model (Hagoort, 2005, 2013). The U-space model represents a touchstone for modeling efforts to simulate incremental comprehension and associated deficits. However, this model suffers from important limitations: it does not incorporate semantics in any form, it only addresses comprehension, and it does not attempt to incorporate data regarding the neural or functional structure of the language system.

The **Lichtheim 2** model (Ueno et al., 2011) uses a structured network of neural layers, each representing a brain region, that replicates in its architecture the dorsal/ventral route distinction assumed to exist in the language system. By simulating lesions at various points of its architecture the model simulates aphasics performances on tasks involving word production (naming), recognition, and repetition but cannot be scaled up to account for more complex linguistic task involving sentence comprehension, production, or sentence-picture matching and therefore cannot simulate the patterns of comprehension of agrammatic aphasics.

Turning to the visual world paradigm, the **CIA net** model (Mayberry et al., 2006; Svantner et al., 2012) offers an abstract neural network level implementation of some of the key features postulated by the CIA conceptual model of situated language comprehension. The model rests on an Elman-type simple recurrent network and offers an interesting illustration of the close temporal coordination that can be established between attention and language processes. However, because of its use of localist abstract neural nets the model remains extremely limited in scope. Kukona and Tabor (2011) present a model that combines a self-organizing model of sentence processing **SOPARSE** (Tabor and Hutchins, 2004) and a dynamical system based on attractor landscapes to represent the dynamics of visual attention. Their effort focuses on modeling the impact of online grammatical processing on eye-movements.

Recently Brouwer et al. (2016) proposed a multi-level artificial recurrent network simulating the incremental interpretation of utterances. It showed how by, in part, allowing feed-back between discourse representation building layers onto lexical information retrieval layers, the system could model a series of ERP results that include both classic cases of N400 and P600 components, as well as P600 responses that have usually been classified as "semantic P600" (see above). A critical aspect of the model is that it does not employ, in order to deal the latter, a "multi-route" approach, a solution that has been suggested by most the key informal models tackling this effect (for a critical review of those multi-route approach, see Brouwer et al., 2012).

Going in the direction of understanding how dynamical system properties can yield language like behavior, Treves initiated a whole line of work studying how latching networks (an in particular Potts networks) can shed a light on sequence learning and generation in the prefrontal cortex (Treves, 2005). In doing, so, it continues the computational tradition of analyzing how to go beyond Hopfield pattern learning network (Hopfield, 1982), to build systems that can (learn and control) dynamical transition between patterns. Language generation taken as a sequence generating issue can be studies by such systems. Starting from an artifical grammar designed for learnability studies (BLISS (Pirmoradian and Treves, 2011)), a whole line of work as analyzed the capacity of such networks to learn to generate language. It is behind the scope of this brief overview to go in the details drawn by those approaches that, sacrificing the desire to build large-scale models, dive deeper into the theoretical questions regarding the relations between the biology of the language system, the dynamics it can support, and language. The reader interested is encouraged to review those studies (Russo et al., 2011; Russo and Treves, 2012, 2011; Pirmoradian et al., 2013; Pirmoradian and Treves, 2012; Kulkarni et al., 2016).

Finally beim Graben has offered some deep analyses of the relations between dynamical systems and

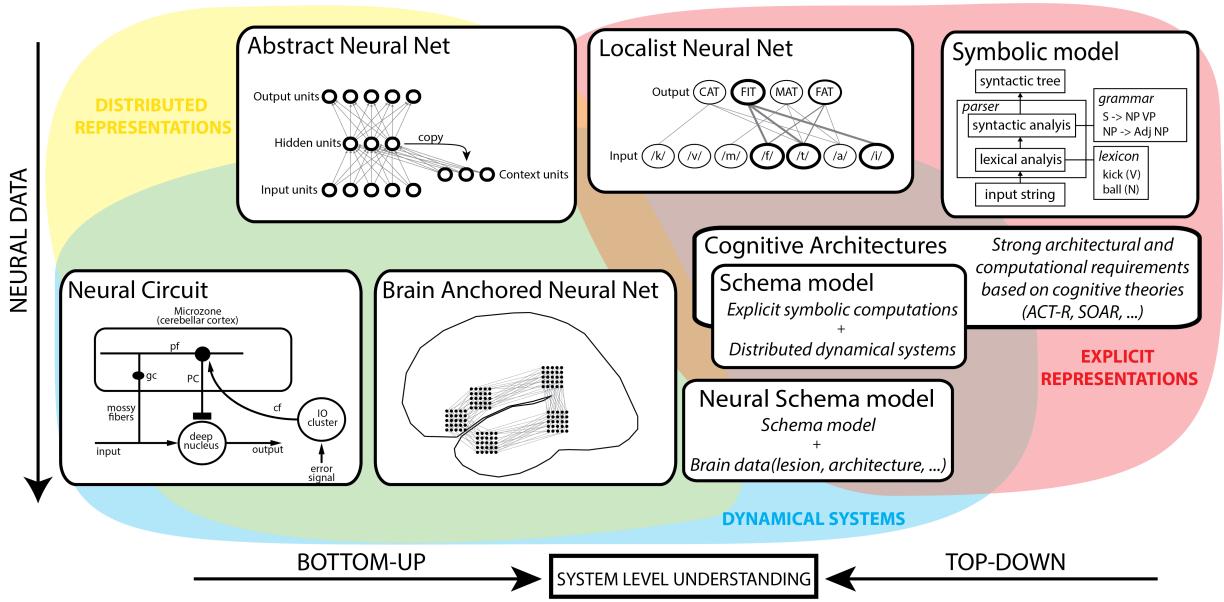


Figure 1.2: Computational modeling frameworks: Finding the right tools for the problem at hand. The complexity of neural systems imposes, if not necessarily at least given the current state of knowledge, that multiple computational methods be available to the researcher, each defining a particular quantitative epistemological angle. The present figure provides an overview of those computational framework, organized according to two perpendicular dimensions: Link to neural data, scale of the systems modeled. In addition, the background Venn diagram contrasts the frameworks based on the type of representation they use: explicit or distributed as well as on the type of mathematical foundations they chose as the basis of their operations: dynamical systems or symbolic systems. The present work will be based on and extend Schema Theory. Schema Theory (schema models) offer a hybrid modeling framework that combines symbolic and dynamic processes. As shown in the figure, it lies both at the boundary between dynamical and symbolic system, but also at the boundary between fully neurally anchored and algorithm oriented modeling approaches. The diagram distinguishes between Schema Theory and Neural Schema Theory in order to make clear the distinction between models that use the Schema Theoretic approach in its processing and architectural requirements and those that incorporate neural data in their design and can be qualified as part of Neural Schema Theory. The ideas behind Schema Theory will be discussed at length in sec. 1.4 and throughout the thesis.

language processing while offering methods to bridge between his theoretical insights and innovations to ERP data (beim Graben et al., 2008). In particular his line of work offer deeper analyses of the theoretical relationships that can be derived between symbolic and continuous dynamical systems. This vast topic, will not be detailed here. The work presented in the following chapter will consider the hybrid approach to linking symbolic operations and continuous dynamics. No attempt will be made to translate the entire model into a continuous dynamical system using distributed representations. For the reader interested in diving into the question of the general translation of symbolic onto distributed representation, see for example (Smolensky, 1990), or the line of work of beim Graben (Beim Graben et al., 2004, 2008; beim Graben et al., 2007).

Fig. 1.2 proposes a way to generally organize the types of computational modeling approaches. It insists on the questions of (1) scale of system modeled (horizontal axis); (2) link to neural data (vertical axis); (3) Style of computational approaches (symbolic v.s. dynamical systems); (4) Type of representation used (explicit vs. distributed). Schema Theory will be the focus of this work.

The vertical axis indicates how closely modeling framework attempt to fit neural data. At the top are framework that focus more on the study of algorithmic properties with often no attempts to link those to neural data. The horizontal axis focuses on the scale of the system that the framework can adequately model ranging from (from left to right) small scale neural circuits in which different neural types can be modeled as well as the data regarding their connectivity (or even at an even smaller scale, single neuron models), to Brain Anchored Neural Nets (BANN) that are often coarser in their individual model of small circuits but attempt to develop neural network models whose architecture (but not only) is constraint by empirical knowledge of brain connectivity, with finally Cognitive Architectures that usually focus on modeling complex cognitive tasks, often only indirectly linking their results to brain data.

The horizontal axis is however not expressed as a progression in a single direction to reflect the methodological stance that complete system level understanding of a nervous system can only be achieved through coordinate and mutually beneficial efforts to develop both computational models that aim to go from the cognitive level toward the neural level, and of small scale computational neural models that aim at integrating their findings in order to move towards offering an understanding of what is for now considered to be the realm of cognitive systems. This is methodological point that applies to the understanding of most complex systems, it remains silent as to the ontological question and should not be taken to advocate for some form of reductionist agenda.

Recent deep-learning models are a good example of Abstract Neural Net models that develop many new learning paradigms but do not qualify as brain models. Symbolic models have been heavily used in the development, for example, of computational linguistics, leading to key insights about the type of processing that can or cannot support certain cognitive operations (e.g. work on learnability for language models).

Localist Network are somewhat in between as they associate both the type of architecture and dynamic processes used in Abstract Neural Networks(ANN) while using explicit instead of distributed representations. Such hybrid frameworks that straddle across the symbolic and dynamic approaches, if they are often more difficult to use at scale and often resist deep formal analyses (due to their heterogeneity and idiosyncrasies), they play a crucial epistemological role as they help to bridge the gap between symbolic and dynamic approaches, offering possibility to use the explicit human designed knowledge that form the backbone of symbolic systems to understand the often difficult to interpret operations of ANNs. Conversely they can help refine the explicit knowledge of symbolic system on the basis of the results achieved by ANNs.

Localist Network are however too limited. They are a specific approach rather than a general framework. Schema Theory provides such general hybrid framework capable of tackling complex computational processes and system. It offers an powerful computational and epistemological tool for building bridges going from the top-down, helping moving towards system level models that display the capacity to be iteratively refined, following the changes in the state of scientific knowledge, offering platforms to attempt the integration of the computational advances put forward by smaller scale models.

### **1.2.2 Vision-for-Action: Integration of Bottom-Up and Top-Down Visual Attention Signal**

Although the computational modeling of the visual system will not be the focus of this work, most of the research presented in the following chapter are the results of a continued program of research that find parts of its root in the cognitive-level computational neuroscience of visual attention as an active process of a

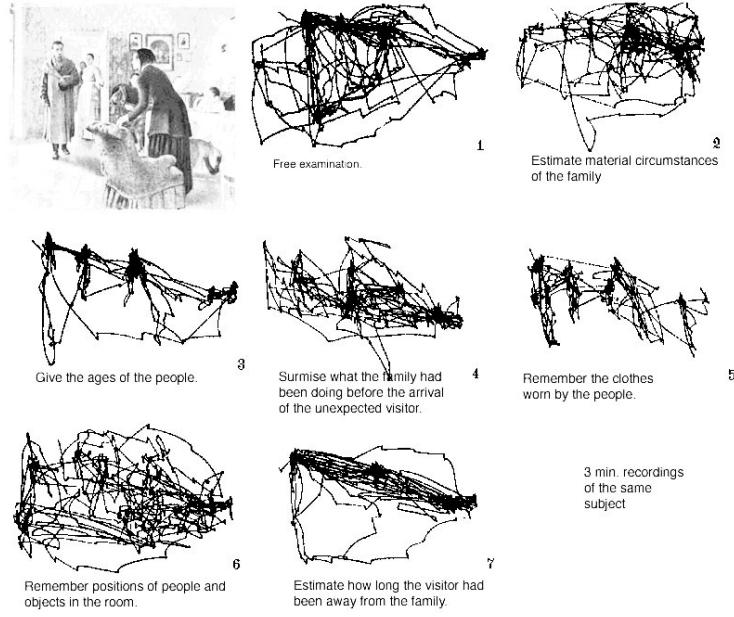


Figure 1.3: Influence of task on scan-path as observed in Yarbus' seminal work (Yarbus et al., 1967). In his seminal experiment, Yarbus recorded the scan paths of subjects as they were visually parsing a scene (here shown at the top-left) in order to fulfill different goals: (1) Free viewing case. No goal. (2) Estimate the material circumstance of the family. (3) Give the age of the people. (4) Surmise what the family had been doing before the arrival of the unexpected guest. (5) Remember the clothes worn by the people. (6) Remember the positions of people and objects in the room. (7) Estimate how long the visitor had been away from the family. Although here no quantitative analyses of the scan path is provided, the difference in scan paths are clear enough (at least between certain goal conditions) for this initial experiment to initiate a vast amount of research on the topic.

“vision-for-action” perceptual system. A model in particular, put forward by Navalpakkam and Itti (2005), played a key role in bootstrapping this research program: It is not an overstatement to suggest that one way to interpret the present work is as a continuation of this model. One of the narrative thread can be described as:

Assuming that we have access to models such as the one developed by (Navalpakkam and Itti, 2005), how can we use this as an advantage to move on and start building system-level cognitive models of situated language production and comprehension and of their interactions with visuo-attentional process.

### Visuo-Attentional System: Vision as an Active Sensori-Motor Process

Attention is key to allocate efficiently the energy of an organism towards the relevant information. In particular, visual attention, by directly orienting the eye of the animal towards locations of the visual scene that contain interesting information (overt attention) or simply by covertly allocating processing resources to a sub-part of the input that impact the retina (covert attention), constantly guides how and what is visually perceived. But what is interesting information? What justifies that a part of the environment should be paid attention to? Broadly construed, the organism wants to direct its visual attention to regions of the environment from which it can extract sensory information that is highly valuable for its survival and increase its capacity to control its relation with the world. It does not come as a surprise therefore that our human attentional system can be quite successfully functionally divided into two components. A bottom-up component that processes the retinal input in parallel and automatically orients the attention towards parts of the visual scene that can be defined by low-level characteristics such as quantity of movements, brightness, flicker, color, etc.

Psychophysics has long been interested defining such characteristics with debates persisting until today.

Simply stated those are reflex processes that allow the organism to survive by ensuring that it remains alert and can quickly react to environmental changes. The other component reflects the use of attention in the process of achieving some more or less conscious goal. This top-down component is not automatic or reflex but can be seen as more volitional (Baluch and Itti, 2011). Its role is to flexibly adapt the attentional processes so that they best serve the purposes set by the task the animal is currently performing. The first clear demonstration of the effect of those top-down processes on visual attentional guidance is often attributed to Yarbus who compared the sequence of saccades (scan path) produced by individual when inspecting a similar picture but while attempting to solve different tasks (determine the social class of the people, guessing what is happening to them, etc.). As the task varies, the scan paths display qualitatively changes (Yarbus et al., 1967) (see fig. 1.3). More recently Triesch and colleagues have used a simulated environment to show how subject can miss important change in the environment (such the change in dimension of a cube they are manipulating) if such changes, when they occur, are not relevant to proper unfolding of the task (Triesch et al., 2003). As they put it, what you see is what you need. Visual attention has been the focus of many computational endeavors (Borji and Itti, 2013; Filipe and Alexandre, 2013; Frintrop et al., 2010). However, most of those focus on the modeling of bottom-up attention processes and do not tackle the issue of its interaction with top-down control. Among the models that do tackle top-down attention, many are black box systems that do not provide an explicit account of the way top-down attention uses cognitive capacities to flexibly adapt the visual attention system to the task at hand (Ehinger et al., 2009).

### **Modeling the Impact of Perceptual Task on Attention: A Starting Point**

More than 30 years ago, Treisman and Gelade (1980) proposed their Feature Integration Theory of attention that stipulates what are the visual features that are used to orient bottom-up attention and, by analyzing pop-out effects in visual search tasks, how are they are combined.

Less than a decade later Koch and Ullman (1987) proposed a conceptual feedforward model of how such processes could be implemented using series of multi-scale spatio-chromatic filters processing the input image in parallel and that are finally combined into a saliency map. This saliency map is a topographic map that assigns a saliency value to each point of the visual scene, reflecting its conspicuity. Supplemented with a winner-take-all architecture this saliency map can be used to select the most conspicuous point of a scene, which becomes the focus of attention. Inhibition of return finalizes the picture by ensuring that the system explores sequentially the most conspicuous points, simulating bottom-up triggered overt shift in attention.

Another ten years were necessary before the first complete implementation of this model was proposed by Itti et al. (1998) that allowed the computation of saliency maps for natural scenes in a biologically plausible model. This implementation or related ones are at the heart of most of the attentions models (Borji and Itti, 2013), furnishing the core process supporting automatic bottom-up attention. Navalpakkam and Itti (2005) were the firsts to try and understand how this saliency map model could be integrated with an explicit reasoning module that could generate signals shaping saliency top-down to account for task-dependent attentional requirements.

The model proposes an architecture that can flexibly account for task-specific attentional guidance in real-world scenes. Importantly, flexibility here means that one of main goal consisted in designing a system that would not be limited to a specific type of task (e.g. single target search), but rather should accommodate the different types of requests. Keyword lists are therefore used to pass on the task specification to the system (e.g. “what is the man catching?” can be passed on by specifying man as a subject and catch as the action). The model incorporates some key elements that allow it to handle such tasks. It incorporates an explicit symbolic world-knowledge long-term memory. It is endowed with a symbolic working memory that can use this knowledge to reason about the world. Parallel to these symbolic processes, the model incorporates a visual long-term memory that stores object representations. Those can be learned and used in a visual working memory that also transiently stores the visual features of the relevant elements of the scene that have been explored (in the spirit of (Kahneman et al., 1992)). Both symbolic and visual working memories play a central role in creating a non-volatile higher-level visual processing layer whose states can control the volatile processes in the lower-level visual layer (Rensink, 2000). Indeed, the core ideas of the model is to use the visual and symbolic working memory to generate top-down signal orienting at each time the attentional system towards relevant objects. First it selectively tunes the bottom-up attentional processes by modifying the gains applied to the low-level feature-maps before they are linearly combined to generate

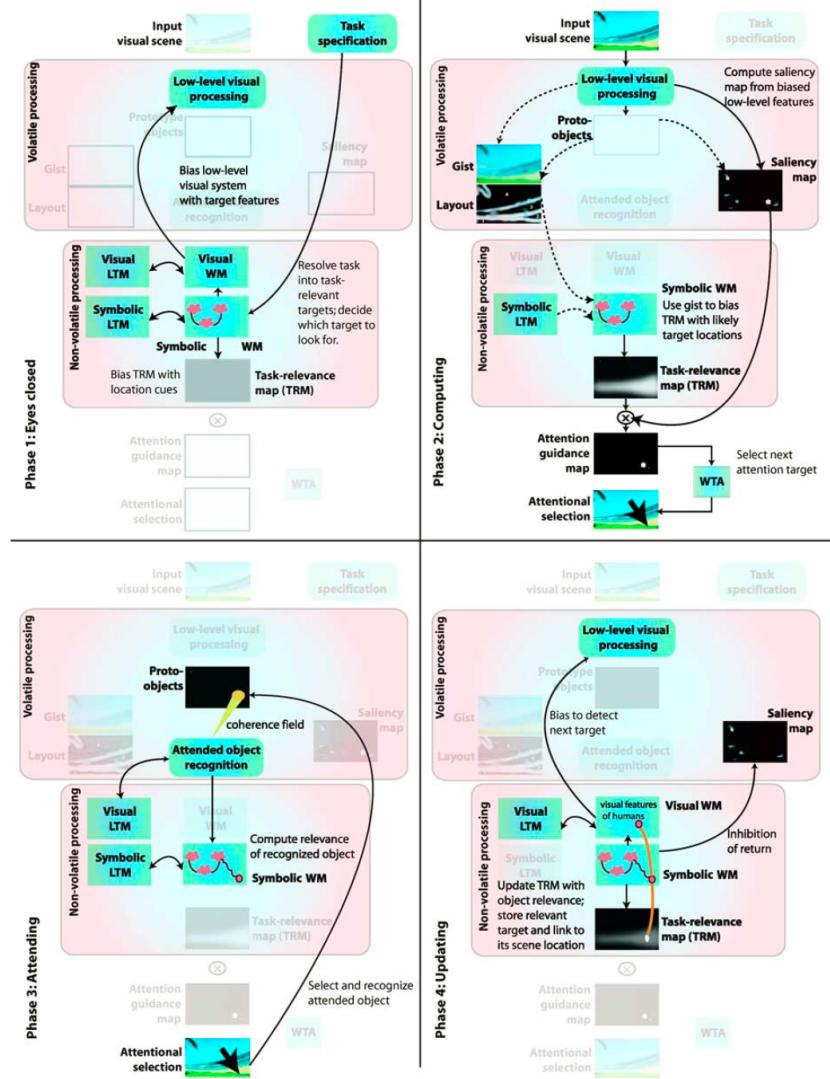


Figure 1.4: Overview of the model proposed by Navalpakkam and Itti (2005). Each panel represent the same model at different phases of processing. In each panel, the active components are highlighted. Phase 1 (top left): Eyes closed, Phase 2 (top right): Computing, Phase 3 (bottom left): Attending, Phase 4 (bottom right): Updating. As the authors wrote, “Following Rensink (2000) terminology, volatile processing stages refer to those which are under constant flux and regenerate as the input changes.” Phase 1: The system receives the task-related goals it will need to achieve (here find a given object). Those set up the states of both the Visual WM that holds active perceptual knowledge extracted from the Visual LTM, and similarly for the associated Symbolic WM that retrieves and organizes knowledge extracted from the Symbolic WM. Both the Visual and Symbolic WM are in a state in which knowledge relevant to the task at hand has already been retrieved. Prior to the visual input being received, high-level perceptual and symbolic knowledge is already engaged in defining task-based top-down guidance for the visual system by biasing the low-level visual processes. Phase 2: the scene input is received. Low-level massively parallel visual processes computes bottom-up saliency information based on low level features. In parallel gist information regarding the nature of the scene is quickly computed. The bottom-up computed saliency map is merged with a task-relevance map generated by Symbolic driven top-down expectations, to generate an attention guidance map that integrate both bottom-up and top-down information. Phase 3: The state of the attention guidance map is used to orient the attentional focus. The attended part of the scene is processed to extract its perceptual content (proto-objects and recognized object). This perceptual information is used to update the state of the Symbolic WM. Phase 4: All the states are updated again, which will lead to a new task-relevance map and to a new attentional shift. The process is repeated until the goal is achieved

the saliency map (in the spirit of the Guided Search model (Wolfe, 1994)). However one of the most unique features of this model is the use of a task-relevance map that can be seen as the top-down extension of the bottom-up saliency map. Based on the state of the working memories, the task-relevance map contains the topographic information regarding where relevant objects are predicted to be located. This task-relevance map is combined with the biased saliency map to generate an attention guidance map from which the focus of attention is selected to orient eye movements (see fig. 1.4 for an overview of the model).

The first key characteristics of this model is its use of the symbolic working memory to simulate reasoning based on world-knowledge and orient the attention towards entities that are relevant for the task. The world knowledge in the model is represented using hand-crafted ontology containing information about objects, actions and their relations (is-a, part-of, similar, etc.).

The symbolic working memory maintains a task graph that contains task-relevant object symbolic representation and their relations. This graph is updated using the ontology and each entity in the graph is associated with a relevance value. The best way to illustrate the functioning of this system is to look at how it processes the task information to generate the initial working memory state. If the task specifies MAN as a SUBJECT and CATCH as an ACTION, this is tantamount to the system to answer the question “what is the man catching?” The initial graph in working memory will contain a node for CATCH, a node for MAN. It will then use the ontology to find paths that connects the two (e.g. a path stipulating that the hand is part of a man and that it is used to catch objects) and such paths will expand the graph. By computing the relevance value of the nodes in the graph, the system might decide that HAND is the most relevant object. A top-down signal will therefore bias the saliency maps to boost hand saliency. Similarly, when an object is recognized by the perceptual system, the symbolic working memory will decide whether or not it is relevant by either checking whether it is already part of the task graph or by attempting to add it. This structure offers a way to model the dynamic and incremental recruitment of world knowledge guiding visual attention during a visual task.

A second important characteristic of the model lies in its capacity to learn object representations that are defined as feature-vector based of the same format than those used in the bottom-up Itti-Koch saliency system. Relying low-level features allows the system to use the same features during object recognition and during top-down biasing of feature-maps. As already mentioned, the biasing is done by changing the weights applied to the different feature-maps before their linear combination into a saliency map. For a given object, each feature-map is assigned a weight that reflects the variance and mean values of this feature in the object representation (the higher mean and the lower variance, the higher the weight). Such top-down biasing scheme is proved to outperform and be easier to generalize than other ways to implement target detection (and in particular the author compare their model with (Rao et al., 2002) showing the shortcoming of this approach).

Finally the authors show how the task-relevance map can be learned. Trained on visual input taken from the perspective of a driver, the system learns to detect the road as the task-relevant area to find cars.

This system was the first one to propose a way to integrate the successful and biologically plausible Itti-Koch model of bottom-up attention with an explicit reasoning system that can generate top-down biases flexibly adapted to various task demands. It introduces many important concepts that outline the type of computational elements necessary to perform such successful integration. Among the key ones are: non-volatile processes combining symbolic and perceptual working memories, gain modulation of low-level feature maps, and a learnable task-relevance map. Some elements are missing and are clearly indicated as requiring further investigation. The ontology is hand-crafted and cannot be learned. Gist is not used to quickly set-up task-relevance expectations. Objects visual representations do include the possibility to encode parts that could offer spatial guidance towards relevant area of the scene (if you found the hand of the man, you can quickly find his arm).

To this date, it remains a rather unique model since very few others can claim to have tried to explicitly implemented such flexible system of top-down attention control.

## 1.3 Grammatical Processes Beyond Syntax: Computational Construction Grammar

In the past ten years, a series of modeling works have used cognitive linguistics as a theoretical basis to generate computational implementation of language processing. While these by no means constitute a unified effort and span from embodied robotics to neuroanatomically specified neural models, they all share the common general assumptions that are put forward by cognitive linguistics, namely that language needs to be understood in terms of its use by a society of agents with situated bodies, and that language processing should not be studied in isolation from what is known of sensory-motor systems and of the type of representations they support. In addition, these share the view that syntactic, semantic, and pragmatic aspects of language cannot be properly analyzed as separate components but rather should be thought transversally as tied together into constructions. Constructions generalize the notion of form-meaning binding lexical item to encompass idioms (e.g. “kick the bucket”), partially filled expression (e.g. “the Xer the Yer” as in “the more the merrier”), or even formally abstract argument structures (e.g. the transitive construction), all carrying their own idiosyncratic mapping between form and meaning. Construction grammars therefore provide a fundamentally new way to look at language knowledge very much in the form a generalized lexical knowledge. I will here review four of the main computationally implemented models of construction grammar insisting on the novel perspective on the human language faculty they offer and on their relation to embodied theories of meaning. In particular I will insist on how the various framework incorporates roles for perceptuo-motor schemas in the language comprehension and production processes.

### 1.3.1 Cognitive Linguistics

None of the the models that were described above is able to tackle the core issue of semantics and how multiple levels of meaningful representations are articulated linking perceptuo-motor motor representations with linguistic semantics and grammatical representations. Cognitive linguistics has emerged as a solid alternative to the generativist linguistic perspective (Croft and Cruse, 2004). Cognitive linguistics focuses on language as a communicative tool that allows for meaningful actions to be performed between members of a linguistic community. The primacy of meaning goes hand in hand with a constant effort to link semantic/pragmatic operations to non-linguistic human cognitive capacities, weaving the language within a more general web of perceptuo-motor cognitive functions considered indispensable to a proper understanding of how language is used, has evolved, and is learned.

The various grammatical formalisms emanating from cognitive linguistic theories tend to fall under the rubric of Construction Grammars (CxGs) (Boas and Sag, 2012; Croft, 2001; Goldberg, 1995; Kay and Fillmore, 1999). Briefly stated CxG cuts across the generativist segregation of morphology, syntax, semantics, pragmatics and proposes to describe all linguistic knowledge as trans-componential, taking the form of constructions defined as form-meaning mappings.

### 1.3.2 Embodied Construction Grammar

The Neural Theory of Language (Feldman and Narayanan, 2004), and its core computational elements that are the X-Schemas and the Embodied Construction Grammar (ECG), offer the computational counterpart to the strong embodiment claim of Gallese † and Lakoff (2005). This framework seeks to explain language comprehension in terms of sensory-motor simulations on which linguistic meaning can be directly anchored in the case of concrete action sentences. In the case of abstract sentences, the idea is that the pervasive use of metaphorical mappings from an abstract target domain (e.g. International economics) onto an embodied source domain (as in “The liberalization plan stumbled”) makes such simulations possible. The concept of X-schemas (Narayanan, 1999) was developed as a way to computationally represent the hierarchy of pre-motor structures that package motor control programs into a limited set of parameters and can be used either to direct action in the world or to carry out offline simulations. Narayanan was successful in showing how a system that contains (1) abstract world knowledge about a target domain (knowledge of international economics coded as a Belief Network), (2) sensory-motor knowledge represented as a network of X-schemas, and (3) metaphorical mappings between the two, linking belief values to X-Schemas parameters, could generate correct inferences, when presented with a newspaper headline such as “Liberalization plan stumbling”: that

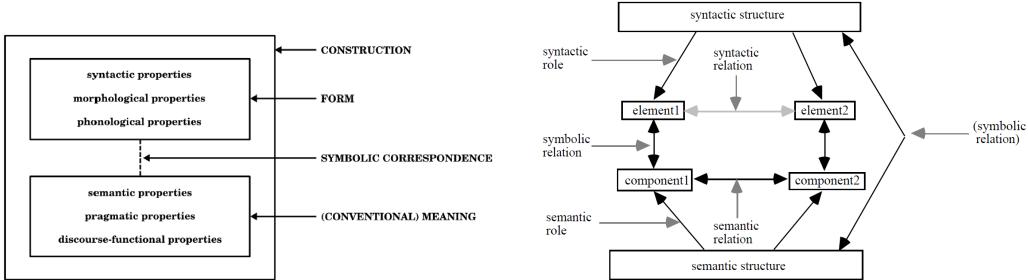


Table 1.1: (Left) General view of a construction (Croft, 2001); (Right) Exploded view of a construction into its components (Croft, 2005). (Left) A construction is a symbol(in the sense of Saussure) in that it defines a symbolic correspondence between a form and a conventional meaning. Construction grammars blur the distinction between grammatical rules and lexicon (the latter being canonically containing the elements that define form-meaning symbolic mapping). It also blurs the componential distinction between phonological, morphological, and syntactic properties that call participate in the definition of a construction's form, as well as the distinction between semantic, pragmatic, and discourse properties, all of which can jointly participate in the definition of a construction's meaning. (Right) Form and meaning poles of a constructions should not be thought of as monolithic entities but as complex structure themselves. The form and meaning pole are structured by internal relations (syntactic and semantic relations respectively). The symbolic mappings between those two structures is then itself composed of multiple symbolic relations. There is no agreement as to the exact definition of a construction and the reader should consider Croft's version only insofar as it illustrates quite well the main tenets of construction grammar that are shared among theories.

there is an ongoing economic plan, that it is facing difficulties, and that it is likely to fail. Such inferences are possible because the system can use X-Schemas to simulate the effect of stumbling on a WALK schema and map the resulting state (falling state unless a lot of force is applied) to the concept of difficulty and failure in the target domain of economic policy. Expanding on this core semantic processing computational model, later work introduced ECG as a way to use grammatical constructions to map more generally linguistic content onto conceptual schemas that can set the parameters of the X-Schemas which finally carry the sensory-motor simulations necessary to generate inferences (Bergen and Chang, 2005; Feldman, 2010).

Constructions in ECG span a large range of form-meaning pairs going from lexical to argument structure constructions, and ECG only covers comprehension. No work so far has addressed production. A precise presentation of the detailed formalism used to represent the constructions would be too tedious so I will insist on some of the key properties that are unique to ECG. Both the meaning and form pole constructions contain sets of features (respectively semantic and syntactic/morphological features) that can be used in a rather classic unification procedure similar to that used in unification-based grammars based attribute values matrices (e.g Head-Driven Phrase Structure Grammar (HPSG) (Pollard and Sag, 1994)). Unifying constructions in order to build a parse therefore largely inherits from constraint programming. This reliance on unification has the interesting processing characteristic that there is no difference in ECG between carrying an attempt to test the match between constructions and their actual successful unification (and we will see that in this ECG differs from Fluid Construction Grammar (FCG), sec 1.3.3).

Turning to the definition of the semantic pole of constructions, the type of semantic features postulated is strongly influenced by frame-based semantics (Fillmore, 1976) and by the formalism of image schemas introduced by Langacker (Langacker, 1986). The unification of construction results in a semantic interpretation of the output called the SemSpec (and not in a classic parse tree) (Shieber, 1986).

A unique feature of ECG consists in that constructions are all defined on an implemented inheritance lattice. Grammatical knowledge is structured in an ontology that adapts the type of representations that have been introduced in cognitive linguistics. This allows ECG to use the lattice during the parsing process to control which constructions are active, avoiding combinatorial explosion. Since construction grammars do not constrain the number of grammatical categories, grammatical features, etc., they did away with the formal parsimony that characterized generative grammars. As a consequence, they are much more flexible and adapt easily to the many “irregularities” of language use. But the price to pay for this flexibility is in

processing and control costs. Given a linguistic input to parse it becomes crucial to find a way to control how the search-space is built. The use of a structured grammatical knowledge is one of the ways to do so in ECG (in addition to ensuring the consistency of the system). Both because finding the optimal parse through an exhaustive search would still be too computationally expensive and because ECG aims at modeling the human use of language which is known to consist more in a satisficing problem (Simon, 1972) than in an optimal solution search, the best-fit analysis heuristic has been developed as a basis to build and explore the search space (Feldman, 2010). This heuristic uses a Bayesian-type of scheme to assign probability to various parses that reflect their plausibility, and it does so in a way that is incremental and robust, two other core aspects of human parsing. The use of the inheritance lattice makes ECG’s semantic more symbolic than procedural. However, the meaning pole of constructions is designed to parameterize the frame-based semantic content so that it could serve as an input to X-schemas. Therefore the SemSpec can be seen as a procedure to translate the semantic content of the input utterance into a set of parameters that will then control the simulations run on X-schemas. The result of this simulation in turn supplies embodied inferences that enrich the SemSpec.

The efficient integration of information sources, and in particular of contextual and linguistic cues, with a focus on co-reference resolution, has been formalized and analyzed (Bryant, 2008) who put forward the notion of “Best-fit analysis”, here again anchoring ECG strongly within the language as language-use scientific paradigm<sup>1</sup>. In parallel, Chang (2008) tackled the question of learnability and showed how ECG can model construction acquisition through language-use.

As a computational implementation of the strong embodiment claims of Gallese and Lakoff this work clearly illustrates how motor schemas can be harnessed to support the inferences that have to necessarily accompany language comprehension including in abstract domains. However, it also shows that sensory-motor simulations are useful only in that they can interact with the abstract knowledge of the source domain. Unfortunately in this work, the question of the nature of brain structures that support the metaphorical mappings are left unspecified.

### 1.3.3 Fluid Construction Grammar

Fluid Construction Grammar (FCG) emerged as a computational construction grammar framework tailored towards a use in embodied robotics. However it is also able to function as a standalone language model that could be used in any computational linguistic task (Chang et al., 2012; Steels and De Beule, 2006; Steels, 2011; Van Trijp et al., 2012). FCG aims at offering a toolbox for computational implementation of (any) construction grammar. Its achievement is therefore a computationally efficient and operational linguistic formalism to define and use constructions for both production and comprehension, formalism that nevertheless remains versatile enough that it does not over-constrain the characterization of the constructions allowing the community of researchers to explore many different options. Any discussion of FCG needs to clearly differentiate between the general formalism that leaves open many key choices and some of its specific uses.

Being a construction grammar framework, grammar is necessarily specified as a set of constructions in FCG. So any user defining her own grammar in FCG will have to specify for each construction the content of the syntactic and of the semantic pole. But for example, FCG does not dictate what general model of semantics should be used (first order logic, embodied robots sensory-motor parameters, lambda calculus, etc). This cannot be overlooked when discussing construction grammar in relation to neurolinguistics since it emphasizes right away that FCG will not in itself commit to any specific relation to sensory or motor schemas. One could use FCG within a predicate logic semantic system completely isolated from sensory-motor concerns or could use it in the framework of embodied robotics with semantics being directly defined in terms of sensory-motor parameters (or any position in between). For this reason I will first present briefly some core properties of FCG before turning to the question of its relation to sensory-motor schemas. Linguistic information in FCG is represented by two main structures: the transient structure and constructions. The transient structure defines all the information that is needed to either parse or produce a sentence. Initially this information is contained in an information buffer. In the case of the production the buffer

---

<sup>1</sup>Although the Best-fit analysis is bayesian in its implementation, I would contend that the general principles it outlines go beyond this implementational choice and can guide modeling work using other theoretical approaches to tackle heuristic based information integration

contain a semantic representation, while in the comprehension case it consists of all the surface structure of the sentence. I will consider only comprehension from now on but the production process is designed as symmetrical in FCG. Constructions in FCG, as in Embodied Construction Grammar (ECG) (see sec 1.3.2), define semantic and syntactic features that function in a way similar to features in unification-based grammars. In addition a construction can be seen as functioning as a kind of daemon. When the comprehension process starts, if part of the surface structure in the buffer matches the syntactic pole of a construction, this construction gets applied. Its syntactic pole merges with the current syntactic pole of the transient structure while its semantic pole merges with the current semantic pole of the transient structure (both defined as trees and both initially empty). As the syntactic pole of the transient structures grows, it becomes possible for constructions to get applied not onto the buffered sentence surface structure but directly onto the syntactic pole of the transient structure. So usually lexical constructions are first applied to single word-forms before more and more complex constructions start applying on top of the lexical constructions. This process shows some interesting computational features. First, lexicon is seen as primary in FCG, therefore even in absence of any complex constructions, a meaning can still be expressed as a set of words or some semantic content can be recovered during comprehension from the interpretation of content words. Second, FCG distinguishes between the matching and the merging process during the application of construction. There is a specific mechanism (match) that checks whether a construction could be applied onto the current transient structure (answering the question, does the form of this construction match the syntactic pole of the transient structure?) while another mechanism (merge) is in charge of modifying the structure if a match is found (i.e. if there is a match, merge both the syntactic and semantic pole of the construction with the transient structure). This implies that only constructions whose form match can modify the transient structure, and that if merge fails it can only be because of a semantic mismatch. Third, FCG uses a type of cooperative computation paradigm as a heuristic to explore the search space. A best-first search method is implemented with scores being attributed to nodes in the search space (corresponding to possible chains of construction applications). The score of a node is based on the scores of the constructions used so far in the parsing chain. The score of constructions is itself defined on the basis of their success in previous interactions and on the quality of their matching with the transient structure. Such heuristic is very interesting since it allows the process to be modified by usage. Finally, recent work by Wellens has started to explore the possibility to generate construction networks from use, organizing the grammatical knowledge, and allowing a construction that has been successfully applied to prime others that are known from past processing to be potentially relevant (Wellens and Steels, 2011).

After this very condensed presentation of FCG, I will turn to the question of its relation to models of semantics by summarizing the embodied robotics framework that predated the birth of FCG, triggered its development, and continued to be one of the main aspect of the research carried using this language formalism. Embodied robotics has been used to address the question of how (artificial) agents that perceive and act in the physical world can establish and use a language to communicate. In particular, the work initiated by Luc Steels has focused on using evolutionary language games repeatedly played within a community of embodied robotic agents to study the possible emergence of a shared lexicon and later the possible emergence of a shared grammar. The initial Talking Heads experiment consisted in a naming game (Steels, 1999). At each turn, two robots, selected from a population of agents, are placed in front of a visual scene (a scene composed of colored geometrical figures). One (the speaker) picks a figure in the scene (a topic) and orients its sensor (camera) roughly towards it. Then it tries to communicate to the other robot (the hearer) what it selected by producing words from its lexicon. The other robot is endowed with the capacity to use the sensor orientation of the speaker to orient its attention towards the generally relevant area of the visual scene. Upon hearing the words produced by the speaker, the hearer has to guess what the figure is and “points” towards it (by moving orienting its camera). If the hearer is wrong, the speaker then points to the correct figure. Given the proper learning rules, it was shown that, starting from random idiosyncratic lexicons for each agent, a shared lexicon will self-organize and stabilize in the population. Parity of meaning is therefore achieved as an emergent property of embodied language use that results in the alignment of the cognitive content (structural coupling). The embodied nature of the agent plays a central role in these results: parity emerges because the players share similar bodies (similar sensors and way to use those sensors), share a common physical environment, have already available sensory-motor schemas to orient their attention based on another agent gaze (camera orientation), share similar ways to encode the meaning of words as sensory parameters. In a next step vertical transmission across generations of agents was added and the linguistic

representations were expanded from lexical items to constructions grounded in sensory-motor representation using FCG. Beuls and Steels (2013), using similar types of language games, were able to show the emergence, evolution, and cultural transmission of grammatical agreement in a population of embodied agents, and this only through the repeated linguistic interactions between agents. This provides a preliminary computational insight into how the process of grammaticalization can result from a historical process of self-organized cultural invention improving the efficiency of language related cognitive operations.

### 1.3.4 Dynamic Construction Grammar

While ECG (see 1.3.2) and FCG (see 1.3.3) make little or no direct contact with neural data focusing on mirror neuron data in the non-human primate (ECG), or on strictly robotic/simulated situated language games (FCG), Dominey initiated a line of work that seeks to understand how neurally anchored neural net models can simulate the learning and use of constructions' form-meaning mappings (Dominey et al., 2006, 2009).

I will refer to this work as Dynamic Construction Grammar (DCG). One of the core element of this modeling work lies in the ties it builds between temporal sequence learning and language processing. The fundamental assumption is that sequential cognition, especially the capacity to learn certain abstract temporal sequences (e.g recognize "ABCBAC" motives) as well as to learn abstract sequence transformation (e.g map an "ABC" motive onto "CBA"), serves as a key scaffolding element of the human language faculty (Dominey et al., 2003). This perspective offers the interesting possibility to anchor modeling of the language system (focusing mostly on classic left hemisphere perisylvian regions) into previous modeling work done on sequential cognition in the macaque that was focusing on the oculo-motor saccade system. This work not only developed some key theoretical results on how to support sequential behavior using recurrent nets but also outlined clearly the central role that subcortical structure and in particular cortico-striatal connections play in sequence processing (Dominey et al., 1995; Dominey and Arbib, 1992). In order to link construction processing to sequential cognition, Dominey et al. (2006) introduce the idea that constructions can be recognized using the sequence of function words they contain. For example, the passive construction used in "The ball is kicked by the boy" could be characterized by the sequence "X Y by Z", while the object-cleft construction used in "It is the ball that the boy kicked" could be associated with the sequence "It is X that Y Z". Dominey shows how a neural net could learn to detect sequences of function words associated with different constructions and used this to map content words onto their proper thematic role (X onto Patient, Y onto the Action, and Z onto Agent in the passive construction). Hinaut and Dominey (2013) used reservoir computing as a way to boost the model's performances. The model has since also been turned into a model of language production (Hinaut et al., 2015). But training a production system is a much more difficult problem than training a comprehension system. Despite the apparent theoretical symmetry that the reservoir type of processing appear to display between production and comprehension, the large difference in algorithmic complexity between the two suggests that different processes could be at work during comprehension and production. Even though the linguistic representations used for production and comprehension can, perhaps, be represented formally by a single construction, the constructions used for production and comprehension at the neural level are distinct.

Although this work does not address the issue of how constructions can be used to generate more complex construction assemblage or how to go beyond the thematic roles to make contact with sensory-motor systems or conceptual knowledge (but see (Dominey and Boucher, 2005) for a embodied robotics perspective of situated learning of construction meaning based on visual sensory information), it provides a unique perspective on how construction processing can be computationally implemented using a neural architecture that is anchored on that of the language system (and including subcortical systems often forgotten in the neurolinguistic literature). In addition, the model builds links between language processing and active vision processes supported by the saccadic oculo-motor system, and in doing so, it makes an important move towards understanding better understanding how brain systems supporting various aspects of temporal sequential cognition could have been lifted to support linguistic functions.

The successes and drawbacks outlined by computational works reviewed above highlight the fact the main computational challenge will consist in building a model that both handles the attention-language interactions highlighted by the results derived from the VWP and accounts for the distributed and nature of the neural architecture that composes the language system while retaining the capacity to simulate symbolically complex

operations. If sub-symbolic neural network models offer an efficient way to integrate neural architecture within a model, they are very limited in the type of operations they can handle. On the other hand, symbolic models and in particular CxG based models display a great capacity to simulate complex linguistic operations and their relations to perceptuo-motor systems. Schema theory (Arbib et al., 1987; Arbib, 1989) offers a computational methodology fitted to build neuro-computational models that incorporate symbolic and dynamic operating principles. Template Construction Grammar (Lee, 2012) has been developed as a first step towards a schema-level model of language production of visual-scene description.

## 1.4 Schema Theory: System Level Cognitive Modeling Framework and Dynamic Cooperative Computation.

Schema Theory was introduced as a top-down counterpart to the bottom-up neural network modeling approach. Neural network models tend to adopt a bottom-up approach in which only small parts of a system is simulated in details leaving aside the question of their integration. Schema Theory offers a framework to build system-level models integrating the processes require to simulate the organization of goal-oriented behaviors. It focuses on the adaptive dynamic interactions between sub-systems, respecting the computational style of the brain (Arbib, 1989).

Schema Theory has already been successfully applied to the modeling of visual scene recognition, motor control, path finding, prey-catching behavior, etc. As more territories are explored, the Schema Theory formalism is refined and revised. This has led to a series of formalizations, each focusing on a certain aspect of the theory (mathematical analysis (Steenstrup et al., 1983), robot design (Lyons and Arbib, 1989), (Oztop and Arbib, 2002), locomotion (Corbacho and Arbib, 1995; Corbacho et al., 2005a,b), and system level cognitive architectures (Draper et al., 1988)). The purpose of the present research consists in building on these previous works and extend Schema Theory to language.

### 1.4.1 Schema Theory: A Brain Theory Framework to Model Organisms' behaviors within Action-Perception Cycles.

Tackling the vision-language interaction at a system level that encompasses multiple sub-systems as well as multiple sensori-motor loops integrated within a cognitive architecture requires to choose a modeling approach that fits the challenge by providing the right level of abstraction. In this work we propose to follow the Schema Theory approach to design brain models focusing on a top-down modeling methodology which consists in focusing on building a high-level architecture that encompasses the whole system, analyzing the various functions it needs to performs, the challenges of their integration in a brain like distributed concurrent computational architecture, with the aim to complement the bottom-up approaches which offer more detailed analyses, possibly at the neuronal level, but limited to very specific sub-systems.

Schema Theory, following in the footstep of early work on schema based cognitive modeling of memory processes (Bartlett, 1932), and on cognitive development (Piaget, 1965), was put forward by Arbib (Arbib, 1989; Arbib et al., 1987; Arbib, 1981) as a brain theory method to offer a principled approach to build symbolic processing system that respect the style of computation of the neural systems.

Schema Theory (ST) was designed to offer a way to start offering models of cognitive system for which we do not yet possess enough neural data to develop full-fledged neural models, while leaving open the possibility to re-factor and refine the schema theoretic models by replacing sub-systems by neural implemented models as knowledge regarding the neural underpinnings of certain cognitive functions become better known.

Schema Theory, as a computational framework rests on a few basic tenets. First of all, schemas are used to define basic cognitive computational units. Schemas can be either learned or innate and encapsulate basic processing units. The cognitive system can be seen as a structure set of schemas that dynamically interact to flexibly organize the course of an organism's behavior given a task and an environment. Schemas can be of different types: perceptual schemas, motor schemas, cognitive schemas. At each moment, the course of action chosen by an organism depends on how the sensory inputs are processed on the basis of the existing perceptual schemas, the state of the cognitive schemas, as well as the way motor schemas can shape the interactions with the environment. A strong emphasis is placed on always studying a cognitive, sensory, or motor, problem through its integration within one or multiple action-perception cycles coupling

the organism with its environment. A first core principle of ST, derived from the known architecture of the nervous systems, is that the ST architectures always aim at simulating *system-of-systems* (*SoS*) in which each sub-system can itself be composed of multiple schemas. In addition, the schema systems are implemented as distributed computational systems, with each schema performing its function asynchronously.

ST models share therefore one of the main assumptions behind neural network computational model: computation occurs in continuous time and in a distributed fashion. Another main characteristics of ST based models is their hybrid nature: although the functional schemas can be implemented symbolically, the interactions between schemas and therefore the overall computation supported by the schema architecture is governed by cooperative-computation (C2) dynamical system principles. Just as it is the case for neural systems, ST systems allow multiple schema functions to enter either in cooperation or in competition, forming cooperation or competition links. The result of those competitions and cooperations is the constant self-organization and emergence of flexible control structures organizing the system's behavior. The competition and cooperation between schemas impact the activation value given to each schema, activation that at each time reflects the relevance of the process it carries to the task at hand.

The C2 process leads to the creation of cooperating schema assemblage that serve as flexible control structure. This allows the system to consist to use schemas that are not necessarily tailored to solve a specific task but that can restructure their interactions on the basis of a task and resources at hand to control the behavior (Arbib and Caplan, 1979). It favors models that can engage in more direct interactions and integration with other modeling work that can use some shared set of schemas but for a different purpose. This sort of approach is already quite popular in robotics work where the community cares about developing autonomous agents that can flexibly reuse functions to best interact with their environment to solve multiple disjoint tasks. Robot Schema ( $\mathcal{RS}$ ) was put forward as a robotic oriented formalization of Schema Theory (Lyons and Arbib, 1989) (R.O.S., although not brain oriented, offer another example of distributed functional processing with reusable functional modules across tasks (Quigley et al., 2009)).

Schema theory provides a level of computational modeling that aims at facilitating later transfer to neural level implementation. It is both symbolic (whether or not a schema has been instantiated) and sub-symbolic (the activity level and parameter values of current schema instances). It allows for the distinction between a feature implicit in the operation of a schema and the activation of a schema that makes that feature explicit. For example, the color of an apple may enter into recognizing an object's shape en route to identifying it as an apple whether or not it enters explicit awareness that the apple is red or green. An initial schema-based model becomes part of neural schema theory if it addresses data from lesion studies, brain imaging, or single-cell recording to help us understand how this behavior is mediated by the inner workings of the brain.

ST also insists on the hierarchical aspect of schema processing in the various distributed sub-systems. To take the case of perception, bottom-up and top-down processes are integrated with low-level schemas offering bottom-up hypotheses regarding the nature of the sensory inputs directly anchored in the sensory modal features extracted by the sensory organs, those bottom-up processes are fed into and influenced higher-level perceptual schemas that offer top-down hypotheses and can in turn bias the lower level interpretation of the inputs.

Schema Theory has already been successfully applied in early models of speech perception. The HEARSAY-II model offered one of the first ST like architecture based in which the idea of blackboards for incremental and concurrent hypothesis sharing was introduced (Erman et al., 1980).

The HEARSAY-II speech understanding system also adopted the perspective of cooperative computation even though implemented on a serial computer. HEARSAY uses a dynamic global data structure called the blackboard, partitioned into surface-phonemic, lexical and phrasal levels. Processes called knowledge sources act upon hypotheses at one level to generate hypotheses at another. Arbib and Caplan (Arbib and Caplan, 1979) discussed how the knowledge sources of HEARSAY, which were scheduled serially, might be replaced by schemas distributed across the brain to capture the spirit of distributed localization of Luria (Luria, 1974). Today, advances in the understanding of distributed computation and the flood of neurolinguistic neuroimaging and behavioral data call for a new push at neurolinguistic modeling informed by the understanding of cooperative computation.

ST architecture modeling grasping motor control have been influential in organizing the knowledge around how grasping abilities are tied to action-oriented perception and to dynamic self-organization of various levels of motor-control .

Directly related to the current work are two ST models that have tackled respectively language processing

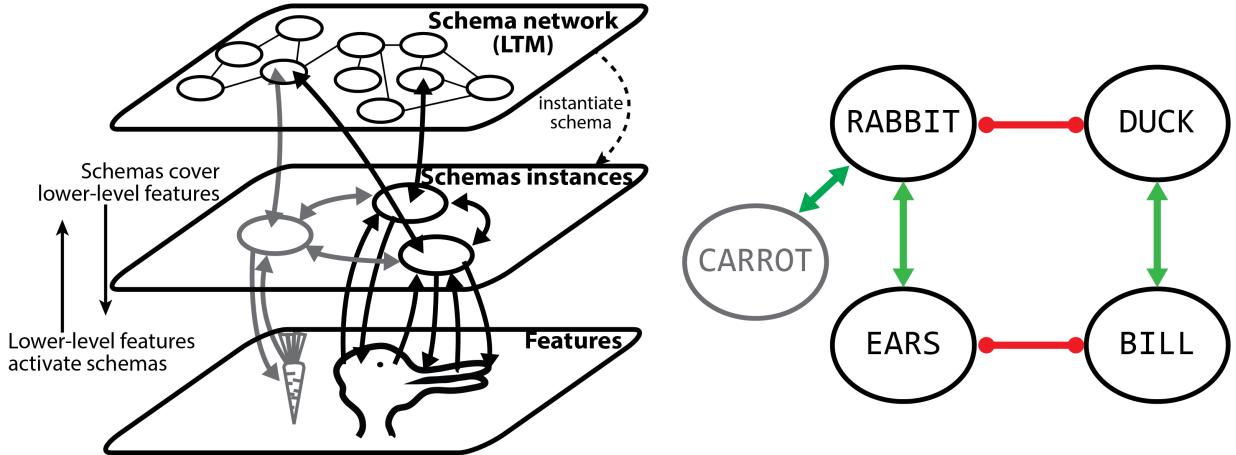


Figure 1.5: Schema Theory and Cooperative Computation (C2) (LEFT) Schema Theory integrates bottom-up perceptually driven with top-down knowledge driven processes within the hybrid symbolic/dynamic framework of Cooperative Computation between schema instances. At a perceptual-feature level (bottom) features drive the recognition of shapes that can be (and indeed usually are!) inherently ambiguous when interpreted at a perceptual semantic level (DUCK/RABBIT) (middle). Perceptual knowledge is stored as schemas in a Schema Network that form a Long Term Memory (LTM). Each schema represent a process that can be instantiated when it becomes relevant to solving the task at hand (in the present example, recognizing the identity of a shape). As brain theory teaches us that the nervous system is best described (in most instances) as carrying concurrent operations, schemas corresponding to competing processing hypotheses can be instantiated simultaneously, resulting in both cooperation and competition relations between schema instances. Here the perceptual instances that support the DUCK interpretations cover the same lower-level perceptual features as those supporting the RABBIT interpretation (although bundling them differently). Both set of cooperating perceptual instances, each forming an assemblage, are in turn in competition. (RIGHT) Informal summary of the cooperative computation situation: perceptual schema instances supporting EARS vs. BILL perceptual interpretation of lower-level features are in competition (red links, competition). They in turn cooperate/support respectively the perceptual schema instances that stand for the RABBIT vs. DUCK perceptual interpretation (green links, cooperation), instances that are in competition. The resulting dynamic system of cooperation and competition is, in abstract, symmetric. A final interpretation could then be reached due to noise leading eventually to the breaking of symmetry, or be bistable. But if most of the sense data are inherently ambiguous, expectations, previous general knowledge and context usually lead to an almost always unambiguous interpretation. Adding perceptual features corresponding to a CARROT interpretation, in a position that can be interpreted (based on our complex knowledge of the world), as being near the mouth of the ambiguous figure if it is itself interpreted as RABBIT (LEFT), breaks the symmetry (RIGHT). As a general principle, each schema instance can serve as a disambiguating context for the instances it cooperates with. This disambiguation by context is by no means guaranteed in absolute but is almost always the case in practice.



Figure 1.6: VISIONS: From visual input to scene interpretation. (Left) Image input. Segmentation: Low-level vision uses competition and cooperation at the level of local image features (edges, colors, etc) to grow edges and regions. This bottom-up process results in a first pass subdivision of the image that can ground semantic analysis. Recognition: High-level vision uses perceptual knowledge to link regions to perceptual schema instances. Instantiation can be purely bottom-up data driven (e.g. SKY), or results from higher level knowledge linking data to context (e.g. roof instance can be associated by a region both on the basis of the bottom-up data and due to the spatial relation of the region to the SKY instance). Finally, instantiation can be purely top-down driven (e.g. a WALL instance can be invoked as a hypothesis for what is below the roof). In this case, the system now interacts with lower-level vision to see if bottom-up data can be found to match this hypothesis). Schema instances compete and cooperate to interpret the different regions. (Right) The output of the process showing the final hypotheses (labels) on which the cooperative computation has converged.

(comprehension and learning) and visual processing.

#### 1.4.2 VISIONS: A Schema Theoretic Model of Scene Interpretation.

The core tenets of Schema Theory are best illustrated by discussing VISIONS Schema System (Draper et al., 1988), a schema theoretic model of scene interpretation. This model will also serve as our starting point for our scene description model.

VISIONS implements a knowledge-based (expert) system of scene interpretation. Given a scene, the model assigns an object-level perceptual interpretation to each region (e.g. Tree, House, Road, Human, etc) (see fig. 1.6). It does so by adopting the Schema Theory design philosophy . In order to insist on the schema theoretic aspect of the VISIONS process we will discuss the model from this perspective, following the interpretation of the model presented in (Arbib, 1989) (see Figure 1.7).

**Schema and Schema instance** In VISIONS a schema is a symbolic unit of high-level perceptual knowledge. Schemas organized into a schema network form the content of a Long Term Memory (LTM): the knowledge the system is endowed with over a certain domain. Figure 1.7 shows the LTM component on the right-hand side. It is composed of a network of object-level schemas, as well as scene-level schemas (see below). Each object schema carries defines (1) sources of positive and negative evidence supporting the presence of the object, (2) strategies for applying the object knowledge, (3) A function  $\mathcal{F}$  that defines how to map the evidence onto a confidence value.

When a schema is supported by enough evidence, it is invoked in a Working Memory (WM) as an executable copy: a Schema Instance (SI). Each SI represents an active hypothesis that an object is present in a region of the scene. Each runs as a separate process in a WM implementing a coarse parallelism well suited for high-level vision. Multiple instances of the same schema can be invoked in WM if the same object is hypothesized to appear at multiple location. The role of a SI is to gather support for its hypothesis. This support can itself be a SI for another object or object part (a WINDOW hypothesis can support a HOUSE hypothesis). SIs can therefore form cooperation networks. Each SI has a confidence value (derived through the  $\mathcal{F}$  function). This confidence value is dynamic since it varies based on the state of the system (new evidence can appear or be evidence can be removed).

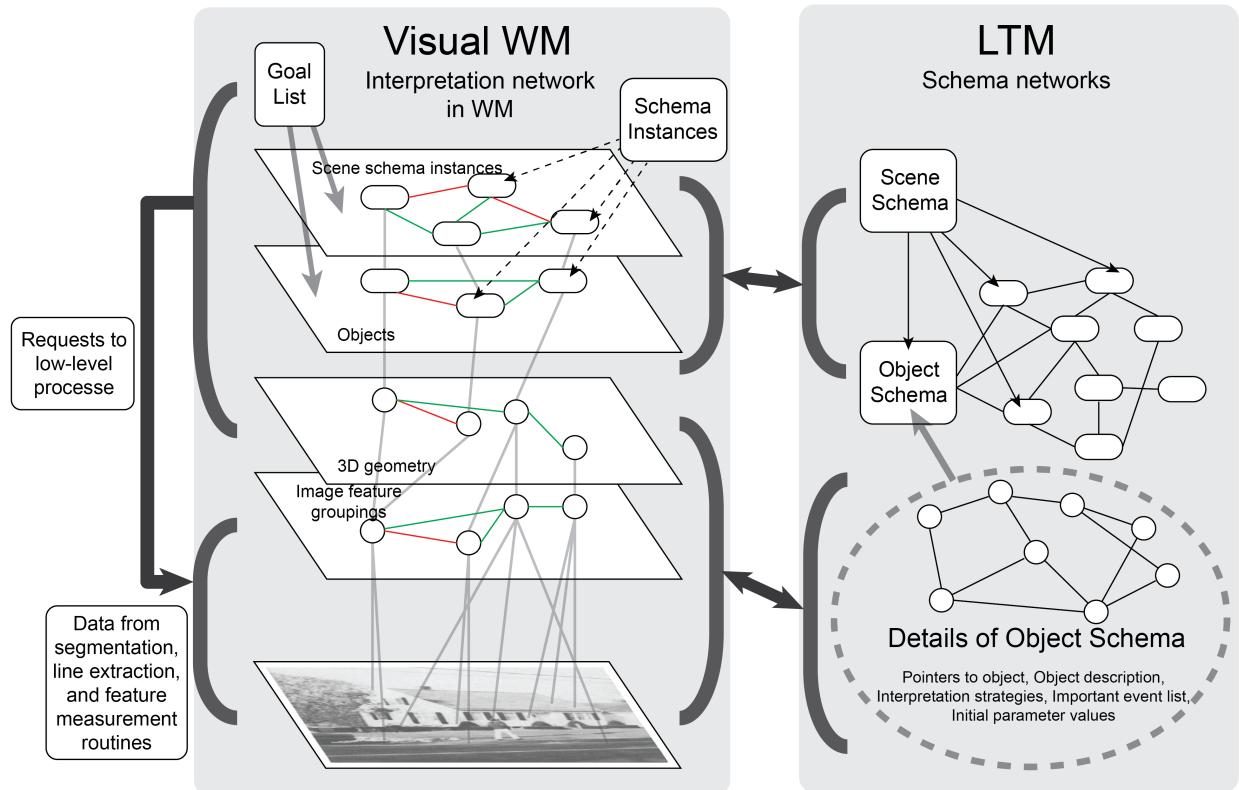


Figure 1.7: VISIONS. Informal description highlighting the schema theory design philosophy (adapted from (Arbib, 1989)). The Visual Working Memory (WM) of VISIONS interprets the current scene by a network of parameterized instances of schemas from Long Term Memory (LTM). These schema instances are linked to the visual world via an intermediate database (here represented by the image feature groupings) that offers an updatable analysis of the division of the world into regions that are candidates for interpretation as agents and objects, possibly in relation with each other.

The VISIONS system combines bottom-up and top-down processing. Low-level features can be extracted from the image, leading to the instantiation of object instances that provide hypotheses as to what those features represent (Bottom-up BU). But at the same time, a SI can predict the existence of an object, spawning a new SI which in turn will predict the existence of some low-level features. This prediction can then be verified by sending requests to low-level processes (Top-Down TD).

The system is seeded with some schema instances that represent the initial hypotheses. The simultaneous processes of BU and TD define the dynamics of the system.

Schemas in VISIONS are extended beyond objects. Schema can also define knowledge about perceptual contexts or configuration. A schema can be associated to a “sub-scene”, whose schemas are related to their parent sub-scene in predictable ways, in the same way an object part is related to an object. This introduces the critical idea that a visual is not simply a collection of objects but itself a complex construct consisting of multiple sub-scenes. The relation between object and scene/context is blurred in this view. As the authors note: “At a sufficient distance, a house is an object to be recognized as a whole. At closer range, the same house also functions as a context for its parts (roof, wall, etc.)” Beyond the features or object levels, the sub-scene defines a complex perceptual cognitive representation, that captures both the highly structured aspect of scene representation and the more ecological level of cognitive unit on which humans base their high-level perceptual operations (Itti and Arbib, 2006).

The network of visual schemas which forms the Visual Long Term Memory encompasses straddles perceptual and general world-knowledge. The knowledge that a HOUSE is generally composed of DOOR and WINDOW is both perceptual in the spatial relations it defines but also general world-knowledge.

Each set of active SIs and their relations correspond to a (partial interpretation) of the visual scene, spanning multiple levels of granularity (object parts, objects, sub-scenes, scene).

**Confidence value and representing uncertainty** At each time step, all schema instances are assigned a confidence value as a way to represent the degree of uncertainty associated with the perceptual hypotheses they carry. Confidence values play a central role in the dynamics of schema theoretic system since they are used to define the outcome of cooperations and competitions between schema instances (see below). VISIONS uses a simple five point scale for uncertainty, whose value is derived on based on purely heuristic combination of evidence. . SI invocation process rests however on a distinction between key evidence that are required to trigger the invocation of a given SI and secondary evidence that only impacts the confidence value.

**Cooperative computation (C2)** Implementing computation in the style of the brain, Schema Theory relies on cooperative computation (C2) as the driving force behind the system’s dynamic. Schema instances active in a working memory form cooperative interaction networks; cooperation occurs when a SI recognizes another SI as supporting its own hypothesis. They also form competition networks; competition takes place between SI representing conflicting hypotheses. Cooperating SIs reinforce each other increasing their confidence values while on the contrary, competing SIs inhibit each other. SIs are pruned out the working memory if their confidence value falls below a given threshold (usually as a result of competition). Cooperation and competition are always local to the SIs involved and conflict resolution is therefore carried out in a distributed fashion. Cooperating SIs can coalesce, forming schema instances assemblages. Each assemblage corresponds to a set of object hypotheses that are compatible both from a spatial and world knowledge perspective. Each assemblage is a (partial) interpretation of the scene.

VISIONS implements a crude version of C2 in which dynamics is greatly simplified by resolving competitions as soon as they appear (by choosing the SI with the highest confidence score), by relying on a discretized 5 points confidence scale, and finally by simulating the fully distributed aspect of computation within a working memory by using a blackboard architecture that keeps track of the state of each processes. If C2 and its dynamic is ideally suited to develop brain models, it comes with its own set of issues, in particular the problem of convergence. VISIONS does not guarantee convergence; limit cycles are not excluded even though they have not been observed in practice.

**Distributed computation** Distributed computation in VISIONS is implemented at two levels: at the system level and at the schema level. The description of cooperative computation between schema instances presents the first level of distributed computation: a scene interpretation emerges through the

self-organization of instances into coherent assemblage. Each perceptual hypotheses carried by a schema instances defines a local context for other hypotheses, and it is through those local interactions (cooperation or competition) that the system goes from an initial state to an interpretation.

Distributed computation also occurs at the system level between system components. The two main components are Long Term Memories (LTMs) and Working Memories (WMs) (see above). As seen in Figure 1.7, the system's knowledge can be partitioned: schemas are divided into classes/types, each defining its own LTM. Each schema type can be invoked in an associated WM. Each WM corresponds to a level of scene interpretation and interacts with other WMs (by linking to hypotheses that are active at a lower level, by supporting higher-level hypotheses, or by hosting SIs that results from top-down hypotheses).

VISIONS, as seen in the left-most panel of Figure 1.7, can be interpreted as a hierarchy of working memories, each keeping track of relevant perceptual hypotheses at different level of abstraction. Low-level routines that extract basic features such as color and edges serves as a basis for intermediate level routines deriving hypotheses regarding shapes, contours, texture etc. Those become the support for object and sub-scene hypotheses. Top-down interactions are simultaneously at work.

As mentioned above, distributed computation (C2 within WM + component interactions) is not implemented through direct acentric information exchange between SIs or between components. The system uses a global blackboard architecture to make the result of each process available to all the other; segmentation of the blackboard allows for some level of modularity. This points to the difference that exists between Schema Theory as a design philosophy and its implementation: a model can be schema theoretic even if it only approximates some of the computational tenets.

### 1.4.3 Dynamic Cooperative Computation: A Core Principle of Cognitive Modeling

The use of Cooperative Computation is a core step in building computational cognitive models. It has been shown to adequately capture the known properties of cognitive operations (McClelland, 1993). In particular, it captures:

1. The prominent role of context (state) in disambiguating upcoming inputs (cooperation, e.g. in character recognition (McClelland and Rumelhart, 1981)).
2. The graded/analogous nature of the confidence associated with cognitive representations as seen through their influences on many cognitive outcomes (e.g. category membership).
3. The temporally gradual nature of the propagation of information between interactive cognitive processes (e.g. influence of temporal distance between predictive contextual cue and target to recognize on reaction time), which also points to the gradual accumulation of information supporting a particular cognitive outcome.
4. The interactive nature of the cognitive processes in which multiple levels of computation can simultaneously support decision processes, with cooperation between levels ensuring the (e.g. scene recognition (McClelland and Rumelhart, 1981; Arbib et al., 1987)).

The C2 approach is also suited as a support for computational cognitive theories since:

5. Competition between mutually exclusive cognitive processes/representations has been shown to be an effective way to simulate cognitive level dynamics (Feldman and Ballard, 1982) (at the perceptual level (Grossberg, 1978) as well as in the motor domain (Cooper and Shallice, 2000)). Competition plays in particular a general role of figure-to-ground contrast enhancement Grossberg (1982, 1977) (a concept that can be lifted to the more general emergence of a cognitive pattern).
6. The parameter space of the C2 dynamics extends the symbolic operations and embeds them within a system that can account for idiosyncratic variability (based on the speaker, the situation, the mental state, the task), a good example being (Cooper and Shallice, 2006; Cooper et al., 2005)) where this dynamics has been shown to model the self-organization of motor programs and the deterioration of

the process in apraxic patients (and also (Vosse and Kempen, 2000) although the C2 properties of the system differ slightly from those mentioned here<sup>2</sup>.)

To avoid any confusion, it is worth noting clearly that in all these types of network models as well as in Schema Theory, units/schema networks do not have vocation to represent actual brain neural networks:

“The units do not correspond to individual neurons, nor do the connections correspond to individual synapses. Rather activations of units represent representational states of a processing system and connections capture constraints that hold among these representational states” (McClelland, 1993).

#### 1.4.4 Schema Theory: From VISIONS to COAST

Cognitive Architecture Schema Theory (COAST) was designed to provide a framework to implement schema theoretic models, framework that is then, in the course of this work, used to develop a novel model of language processing (Template Construction Grammar) and of vision-language interactions (Schema Architecture for Language-Vision InterActions (SALVIA)). This is the neuro-cognitive computational research program that will be detailed in the following chapters.

VISIONS highlights the role that Schema Theory plays in allowing for modeling complex high-level cognitive processes that cannot yet be fully reduced to models directly implemented by neural systems. Early components can be refined to closely approximate what is known of early visual processes, while the high-level ones uses symbolic representations and have vocations to be eventually replaced by schema systems as our understanding of high-level vision improves (which has proven to be quite a slow progress).

Schema theory based model require to choose a way to decompose a system into distributed components and stipulate how those should interact. COAST directly adopts this philosophy all models are defined as systems-of-systems (SoS) on which distributed computation takes place. This decomposition is key to building systems can evolve and be refined, integrating novel insights as the knowledge of a brain system changes but also highlighting the processes that require better understanding. The level of coarseness of a component/sub-system can vary and eventually be replaced by a schema system incorporating a more detailed model of the function it performs, without having to necessarily change the rest of the system. Fig. 1.8 provide a general description of the type of systems of systems-of-systems designed by COAST.

This system-level top-down approach to modeling offers a necessary counterpart to the bottom-up approaches which focus on detailed analysis of a single process. In distributed systems and in particular in those simulating brain functions the challenges of defining the hierarchy of functional decompositions and of understanding the principles that allows for the integrations of dynamically interacting parts into a coherent whole requires its own modeling framework as it raises questions that cannot be solved at the scale of single processes modeling.

VISIONS offered a direct view into the role played by the concept of schema and schema instances (SIs) in designing different types of knowledges and their application. COAST follows these features of knowledge design. COAST relies at its core on schema, schema instantiation as schema instances, as well as on the two main sub-systems Long Term Memory (LTM) and Working Memory (WM), which both retain a definition compatible to that given in the discussion of VISIONS.

COAST departs from some of the processes defined in VISIONS in order to move Schema Theory in a direction that brings it closer to the requirements imposed by brain theory. From a brain theory perspective, cooperative computation plays a key role in the design of schema theoretic models: cooperative computation (C2) belongs to a core set of fundamental brain operating principles .

Contrary to the VISIONS Schema System, COAST defines C2 as a dynamic process. COAST schema systems are hybrid systems. Schema instances have continuous activation values (confidence value), functions of time. The cooperation and competition is a-centric and distributed (no use of blackboard). Competition is therefore not instantaneous but rather the result of the system’s dynamics through the network of competitions links. Some simplifications are made in COAST. Instantiation strategies and confidence functions are shared among a schema type and are defined by the WM in which they interact. In the current model, most schemas do not define their own processes but rather store static data. But this limitation is not inherent

---

<sup>2</sup>Competition takes place between cooperation links rather than between units.

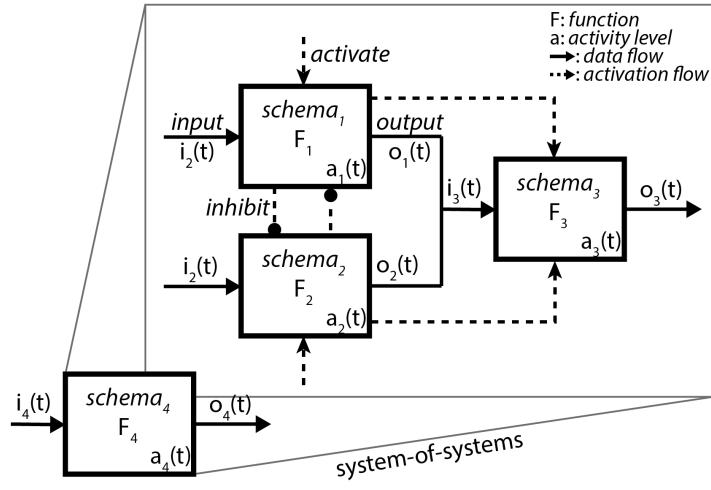


Figure 1.8: System-of-System approach to building neuro-cognitive models. Each box represents a schema carrying a function  $F_i$  and and a time-dependent activation value  $a_i(t)$ . Each schema receives inputs both in terms of information flow (solid lines) and activation signals (dashed lines) from other schemas. The main addition to the classic idea of algorithmic/functional composition of concurrent asynchronous systems, lies in the use of the activation values that can dynamically reshape the way the computational processes are used: it can flexibly re-organize the composition of processes of the system while the actual sub-processes are kept stable in their function. This is a key characteristic of brain systems: a single sub-system can perform a function that will then be shared amongs multiple functional systems. It is the role of brain theory and brain science to understand what those functions are, how they are organized, and to do so in a way that is not task specific.

to the COAST framework and results from the current state of implementation. The details of COAST are given in Appendix A

Schema Theory embraces the idea that the goal of computational cognitive neuroscience consists in building models both from the bottom-up: starting with well defined small networks that are accessible to detailed neurobiological experimentation, and from the top-down: modeling complex cognitive systems for which we have little access to detailed biological data but, for what data has been gathered and existing phenomenological computational models, require computational understanding of how a function can be both decomposed in a concurrent system. Only the dialectic between those two approaches can help navigate between Charybdis of the explosion of fine-grain myopic perspectives and the Scylla of hyperopic systems whose desire to synthesize combined with the difficulty to revise their theory result either in the Ptolemaic approach of constantly adding epicycles and amendments or in the idealist tendency to disregard as false or unimportant the bottom-up models that do not match their view. Schema Theory, among other approaches, offers a methodology to build top-down model whose very objective is to be revised, specified, re-organized, as soon as more data and model emerge from bottom-up approaches.

## 1.5 From Gaze to Speech: the Template Construction Grammar Research Program

Template Construction Grammar (TCG) has been at the center of a research program focused on building a schema theoretic account of language processing with an emphasis on the role of visual schemas and on the way humans are able to talk to each other about what caught their attention, producing or comprehending description of visual scenes. Initial work was able to show how models of goal oriented attention guidance models ((Navalpakkam and Itti, 2005) encompassing both top-down and bottom-up attention mechanisms, can be used to analyze how linguistic and visual information interact (Itti and Arbib, 2006). In particular, this work put forward the question of how the temporally unfolding sequence of subsenes, that are built by an active visual system parsing a visual scene, can be linked to unfolding sequence of grammatical constructions used to express the semantic content of the scene, or vice-versa, how the grammatical constructions received during comprehension interact with the way the visual system will parse a visual scene. (Arbib and Lee, 2007, 2008) lifted the classic VISION schema model of visual scene recognition (Draper et al., 1988) to incorporate the semantic representations necessary to package the content of visual schemas into a format that can be made accessible to the language system (SemRep). They developed in parallel the TCG construction grammar framework that supplemented schema theory with “linguistic schemas” capable of mapping the SemRep onto utterances through a dynamic and incremental process of cooperative computation. Making contact with behavioral data, later work showed how the model was able to account for the qualitative variations of type of utterances subjects generate when asked to describe a scene under various time pressure conditions (Lee, 2012).

## 1.6 Outline of the Thesis

The next three chapters focus on language production. Chapter 2 and 3 introduce describe how the novel Schema Architecture Vision-Language InterAction (SALVIA) cognitive model simulates the dynamic and incremental interactions between the visuo-attentional system and the language production system. By simulating the production of scene descriptions, SALVIA provides an explicit framework to study the coordinated distributed processes that support visual scene apprehension, conceptualization, and grammatical processing leading to utterance formulation. Those chapters follow Kuchinsky (Kuchinsky, 2009) and reframe the psycholinguistic debate regarding the relations between gaze patterns and utterance forms: moving away from a dichotomy between serial modular (Griffin and Bock, 2000) and interactive views (Gleitman et al., 2007), they show that those can both be explained simultaneously as two ends of a spectrum by simulating the impact on the system’s dynamics of variations in the relative temporal characteristics of processes and in their modulation by task requirements and scene types. Chapter 2 introduces the SALVIA model while Chapter 3 presents the simulation results. SALVIA simulates the effect of attention capture on the order of gaze patterns and show how the model can account for preferential emergence of meaning to form mappings

that package in their information structure the effect of perceptual saliency, with a particular focus on the active vs. passive construction. However, as a necessary first step, it is necessary to start by showing how SALVIA can model empirical results regarding the impact of time pressure on the quality of utterances produced (structural compactness and grammatical complexity). This first step is fundamental: it is through (temporal) constraints imposed on the system that the complex interactions between language and visual attention are systematically put to use and therefore reveal themselves at a behavioral level. SALVIA not only provides computational foundations for the empirically-based informal account of Kuchinsky, it also formalizes her theory and suggests new empirical questions, offering a computational framework on which a dialogue between experimentalists and modelers can be grounded.

SALVIA makes use at its core of a novel model of grammatical processing: Template Construction Grammar (TCG). Chapter 4 offers a formal overview of TCG as a computational construction grammar (CompCxG) framework for language production. TCG was developed as part of a brain theory modeling effort to build a system-level neuro-computational model simulating the dynamics at play during language-vision interactions in the context of online visual scene description production or comprehension. Rather than developing a CompCxG handling a wide scope of grammatical constructions, the TCG framework focuses on modeling dynamic adaptive interactions between incrementally built semantic and grammatical structures during online language processing. Schema Theory (ST) provides guidelines to implement cognitive-level hybrid computational computational models that operate in style of the brain. TCG extends schema theory to language.

Chapter 5 operates a double shift in perspectives. First, in order to fulfill the overall goal to model language use, production cannot be the end of the story. Language as a social tool for communication cannot exist but as supported by both a production and comprehension system. The next two chapter will therefore tackle the challenge of turning SALVIA into a model language comprehension. Second, so far SALVIA has only made contact with behavioral results. If it does qualify as a schema theoretic cognitive model, it does not yet qualify as a neuro-cognitive model. As the focus is turned to language comprehension, it will also be turned toward neurolinguistic data and in particular neuropsychology results regarding agrammatic aphasia. The second goal will therefore be to incorporate and simulate neurolinguistic data into the design of the SALVIA model of language comprehension. This lead to the extension of the conceptual modeling discussed in (Barrès and Lee, 2013) that proposed to build the comprehension model as a two-route model comprising both a Grammatical and a World Knowledge route. This chapter provides a conceptual overview of the challenges and of the SALVIA comprehension model.

Chapter 6 is the computational counterpart of the Chapter 5. It details the computational underpinning of TCG and SALVIA as computational model of language comprehension. It details successively the processes supporting the Grammatical and World Knowledge routes before tackling the question of their integration within the Semantic WM. It provides simulations results that describe the model at work. Finally, it goes back to the question of agrammatic comprehension and shows how SALVIA provides a novel interpretation of the tripartite divisions in their comprehension performances patterns on active vs. passive voice sentences described by (Berndt et al., 1996).

Chapter 7 is its own self-contained work. It once again changes perspective, this time to tackle the question of how computational model of language can make contact with brain data and in particular with EEG recordings. It extends the technique of Synthetic Brain Imaging (SBI) introduced by Arbib et al. (1994) (see also (Arbib et al., 2000)) to include a tool to generate Synthetic Event Related Potentials (Synthetic ERP). This chapter only makes a first step towards building the Synthetic ERP toolbox, but presents simulation results that highlight the importance of developing quantitative methods to bridge the gap between conceptual and computational neurolinguistic models.

SALVIA offers a cognitive model, and, in its link to aphasia data, it made a first step toward becoming a neurocognitive model. However, much remains to be done in order to go from cognitive to neurocognitive computational modeling. The last chapter thererofre opens up the discussion of the neural substrate of the language production and comprehension system. Since proposing a direct mapping between SALVIA and brain systems would be doomed to failure, this chapter rather first proposes a summary of the neurolinguistic results regarding the neural architecture of the language system supporting comprehension. It insists on the consensus that has emerged regarding the multi-route architecture of the comprehension system, and in particular on the distinction between ventral and dorsal routes that could provide a coarse anatomical correlate of the distinctions between semantic and syntactic processing respectively. It places these results in

relation with the multi-route architecture of the visual system and puts forward direction for future research on the neural level interactions and interfaces between language and visions. Narrowing the focus to a single brain region, the hypothesized roles of Broca's area in language comprehension and production are then analyzed. It seemed important to do so given both its historical significance in neurolinguistics and its relation to brain functions that are more and more commonly regarded as having played a key evolutionary role in the emergence of our linguistic capacity (be it manual dexterity (Arbib, 2012), or cognitive control). Finally, the difficult question of how the brain constructs meaning from language is discussed. The goal of this final section is both to highlight the many confusions that contaminate this field of research, to offer a map of the terrain, and to put forward the position that computational neurolinguistic theory requires to do away with syntax centered approaches for a focus of the meaning creating processes, contrasting language against the many other ways in which humans make sense of the world in order to act upon it.

# Chapter 2

## SALVIA: An Implemented Schema-Theoretic Framework for Investigating the Linkage of Vision and Language Production

*“Figuratively speaking, every animal subject attacks its objects in a pincer movement - with one perceptive and one effective arm.”*

Von Uexkull

A Foray into the Worlds of Animals and Humans: With a Theory of Meaning

*“Contra vim mortis non crescit salvia in hortis.”*

Medieval adage

### 2.1 Introduction

Psycholinguistics has made great strides those past twenty years in investigating language production in the here-and-now.

The Visual World Paradigm (VWP) provided an empirical procedure to study the dynamic interplay, at the behavioral level, between language production and visual attention (Tanenhaus et al., 1995; Henderson and Ferreira, 2013; Knoeferle et al., 2016).

The basic procedure consists in asking a subject to produce a description of a scene while her eye-movements are recorded using eye-tracking. The VWP yields two time series corresponding to two overt behaviors: the sequence of fixations and saccades, and the utterance produced.

We propose the Schema Architecture Language-Vision InterAction (SALVIA) as a novel implemented cognitive-level model of vision-language coordinated and dynamic interactions in the context of the production of visual scene descriptions. SALVIA represents a new chapter in an effort to develop a modeling framework based on brain theory to analyze the vision-language interactions (Lee, 2012; Barrès and Lee,

2013; Barres, 2017; Arbib and Lee, 2008).

Classic models of language production (Fromkin, 1984; Garrett, 1980; Bock and Levelt, 2002) were developed on the basis of the offline analyses of speech error data. All converge on a multi-stage/module organization of the language production system (each stage being justified on the basis of a type of speech error). They broadly agree on a division between a message level, a grammatical level and an execution level: the message level stipulates the semantic content of the message to be verbally encoded, the grammatical level (that can be divided in sub-modules including functional and linear ordering) is hypothesized to generate the translation of the message onto an executable language production motor program (formulation), while the execution level covers the motor production of the verbal message.

Language processing however, as any neuro-cognitive process, is in essence dynamic and incremental and those conceptual models largely ignored the temporal aspects of the processes in their description. In addition, they cannot provide any clue as to how the message is itself generated. The psycholinguistic work on the VWP addresses those shortcomings. In this framework, the message is generated through the process of apprehension of the visual scene. And the data collected lead to the revival of two main hypotheses regarding the nature of the interactions between apprehension, formulation, and execution.

The first one can be traced back to Hermann Paul and proposes a direct connection between the fixation order and word order (Paul, 1970). According to this hypothesis apprehension and formulation go hand in hand and the sequentiality of language directly reflects the sequentiality of the concepts that are retrieved on the basis of the perceptual parsing of the scene. This behaviorist take on the problem finds is cognitivist counterpoint in the opposite claim that was first put forward by Wundt and then by Lashley according to whom apprehension and wholistic conceptualization of the perceptual content necessarily precedes formulation: The speaking subject encodes events, at least coarsely, before starting formulation. Formulation is therefore considered to be the sequentialization of a wholistic structure (Wundt, 1970; Lashley, 1951).

Through the application of the VWP, each position has found a modern counterpart and at least partial confirmation. According to Griffin and Bock 2000 apprehension of the event structure, at least at a coarse level, precedes the formulation process. However, after production starts during extemporaneous speech, eye-movements appear to be piloted by the formulation process as indicated by the orderly relation that exists between gaze sequence and word order.

The authors conclude that, in the spirit of the model proposed by Wundt, apprehension sets the stage for formulation to begin, formulation that is then an incremental process involving feedback effects on the visual-attentional processes, but always on the basis of the initial event apprehension.

Using Attention Capture Manipulation (ACM) to subliminally control the subject's initial fixation position, Gleitman et al. (2007) found effects of initial gaze position on formulation which points toward some degree of influence of visuo-attentional dynamics on formulation. A closer look at their results leads them to endorse an even stronger hypothesis in which apprehension and formulation are in full interaction. Gleitman et al. offer a modern version of Paul's view in which there is a reliable relationship between initial gaze patterns and word order. Against a behaviorist view however, the authors postulate an active role for intermediary cognitive structures, mediating the impact of visuo-attentional dynamics on linguistic form.

In parallel with the question of independence or interaction of apprehension and formulation, Kuchinsky proposed to organize the two conceptual views supported respectively by Griffin and Bock on the one hand and Gleitman et al. on the other as "**linguistic guidance**" vs "**perceptual guidance**" respectively (Kuchinsky, 2009).

The perceptual guidance view, in the wake of Paul, emphasizes the role of bottom-up driven attentional shifts on apprehension and formulation. In contrast the linguistic guidance view, in the wake of Wundt-Lashley, hinges on the idea of "**seeing-for-saying**" (similar to the notion of "thinking-for-speaking" put forward by Slobin (Slobin, 1996)), and emphasizes the role of top-down driven attentional shifts in the generation of scene description, with the top-down signals originating within the language system.

We propose that those two views taken together can be synthesized into a model that involves two types of incremental dynamics.

The Wundt-Lashley view posits a sequentialization of an already built wholistic semantic structure which suggests a serial relation between apprehension and formulation, while the linguistic guidance principle that emerges from Griffin and Bock suggests an important role of feedback from linguistic system to perceptual system (again "**seeing-for-saying**"). That is, the system starts with event level apprehension that gives a, possibly coarse, wholistic semantic structure on which is anchored the formulation process, formulation

	Griffin and Bock 00 (M1)	Gleitman et al. 07 (M2)
Original conceptual positions	Wundt-Lashley	Hermann Paul
Recorded ACM effects	-	+
Attentional processes $\leftrightarrow$ Linguistic processes	<b>Linguistic guidance (<math>\leftarrow</math>)</b>	<b>Perceptual guidance (<math>\rightarrow</math>)</b>
Apprehension $\leftrightarrow$ Formulation	Serial ( $\rightarrow$ )	Interactive ( $\rightleftharpoons$ )
Main level of linguistic incrementality	Structural incrementality	Lexical incrementality

Table 2.1: Summary of the differences between the two VWP-based models of visual scene description put forward by Griffin and Bock (2000) (M1) and by Gleitman et al. (2007) (M2).  $X \leftrightarrow Y$  indicates a question regarding the nature of the interactions between processes  $X$  and  $Y$ . M1 and M2 respectively correspond to modern extension of the original conceptual positions of Paul Wundt-Lashley and Hermann. M1 presents apprehension as temporally preceding formulation, while M2 posits an interactivist account of the relation between apprehension and formulation. M1 insists on the top-down impact that linguistic process can have on the attentional processes (linguistic guidance) while M2 focuses on evidence pointing to the causal impact of saccades sequence onto linguistic form (perceptual guidance). M1 and M2 both highlight the incremental aspect of the linguistic processes but M1 focuses on the role of structural incrementality (incremental building of the semantic and syntactic structures) and M2 on the role of lexical incrementality (word choice).

process that can then send top-down signals back to the visuo-attentional system. Here the incrementality emerges from an outer feedback loop from the formulation system to the visuo-attentional system. The work by Gleitman et al. establishing the validity of perceptual guidance suggests that incrementality can be observed at the level of the transfer of the dynamically apprehended scene onto the formulation stage. This incrementality is driven by attentional shifts triggered by bottom-up saliency signals. Our model therefore incorporates both types of interactions and show how perceptual guidance and linguistic guidance can be seen as the two end of the spectrum of the system's behavior, and are triggered by the impact on the system's dynamics of two key task dimension: scene types and time pressure.

The main question that is not addressed is the timing of the interactions in the system, including that of the combination of BU and TD cues. The use of both prepared and extemporaneous speech production introduces a key element in the analysis of the interaction dynamics between visuo-attentional and linguistic system: that of time pressure. Changing the temporal requirements of the task offer not only an important empirical test case, but it also points to a core test case for computational models that should be able to properly handle the variation in production dynamics introduced by time pressure.

## 2.2 The Schema Architecture Language-Vision InterAction (SALVIA) Cognitive Model

### 2.2.1 From VISIONS to SALVIA: Towards a Schema Theory Approach to Language Processing

The Schema Architecture Language-Vision InterAction model (SALVIA) applies Schema Theory and extends the architecture proposed by VISIONS to a model of visual scene description. VISIONS focused on interactions within the Visual WM between various types of perceptual schemas. SALVIA zooms out to tackle a cognitive system linking perception (vision) to action (language production). It focuses on modeling the interactions between the various working memories linking the dynamics of visual attention to that of utterance production.

Figure 2.1 presents an informal outline of the cognitive processes that needs to be articulated within SALVIA (the processes that are directly tied to sensory-motor systems - e.g. saccade system - are not represented here). Linguistic structures, at a cognitive level, are to be understood as interacting with a network of modal representations (Jackendoff, 1997). During language use, those directly or indirectly contribute to generating (production) or interpreting (comprehension) the conceptual structure or message (for a more in depth review of the question of semantic processes and in particular of the modal/amodal debate regarding the nature of semantic representations cf. Ch. 8, sec. 8.4).

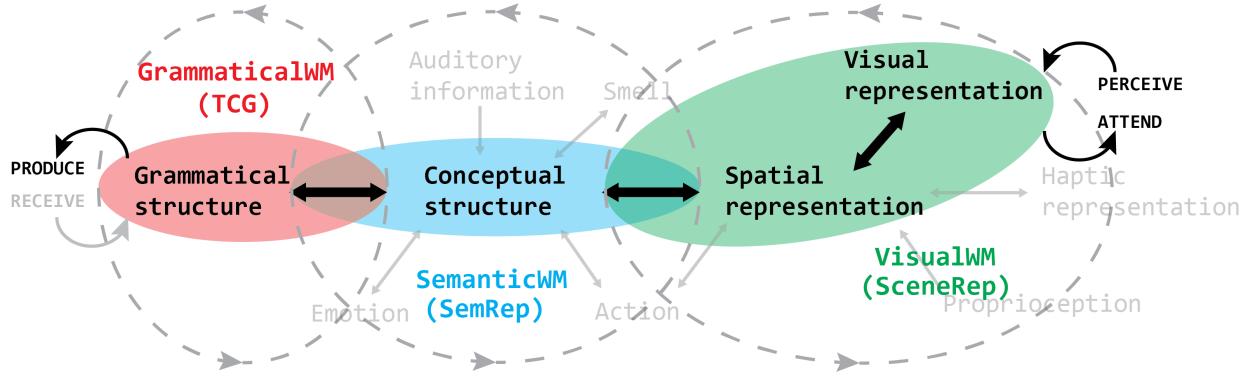


Figure 2.1: Informal view of how SALVIA extracts the visuo-linguistics interactions from the network of interfaces that was proposed by Jackendoff (cf. ch. 1, fig. 1.1) and reframes them in terms of a network of three working memories (WM) (VisualWM, SemanticWM, and GrammaticalWM). The Semantic WM plays the role of hinge articulating the visual and grammatical processes. In comparison to the static picture provided by Jackendoff, SALVIA concerns itself with the incremental and dynamic aspect of the coordination processes that take place. Dashed circles mark highlight the fact that the states of each working memory is time dependent and that the different states of the three WM are coupled and interpendant at the interfaces. Finally, SALVIA moves away from the description of the cognitive structures to highlight their interactions with sensori-motor systems Attention-Vision (right) and Utterance generation (left).

The SALVIA model goes beyond the static informal representation of relations between representations (Fig.2.1): representations and knowledge are interpreted as schemas that become part of dynamically interacting systems (here working memories). SALVIA focuses only on the vision-language interaction, and applies the schema theoretic method to highlight the interactions of what are now modeled as three working memories (WM): a Visual WM which encompasses both visual and spatial representations, a Semantic WM in which the conceptual structure is built, and a Grammatical WM in which the grammatical structure is built. We chose the term grammatical structure instead of syntactic structure to highlight that we approach grammatical processing as going beyond the purely syntactic level.

By focusing on the vision-language online interactions, SALVIA addresses a key computational question raised by theories of online situated language production: How does the human cognitive system orchestrate the “seamless” integration and coordination of multiple time-dependent processing states ultimately resulting in the coordination of two overt sequential behavior, utterance production and attentional scene parsing. (see the Coordinate Interplay Account model (Crocker et al., 2010)).

SALVIA expands the scope of schema theory by lifting and updating the architecture of VISIONS to a model of language production of visual scene descriptions. Figure 2.2 presents an informal view of the SALVIA architecture that highlight its extension of the VISIONS schema system.

The model offered in VISIONS, if it provides a core basis for the schema theoretic analysis of visual processing, remains silent with respect to the two key aspects our work focuses on: (1) attentional processes are absent from the visual system which focuses on object recognition. Incrementality of processing in the model only derives from the C2 dynamics that pilots the self-organization of visual schema into cooperation assemblage combining bottom-up and top-down perceptual knowledge (incrementality from internal dynamics), but does not include the incrementality stemming from the visuo-attentional parsing of the scene (incrementality from active-perception). It is worth noting that another type of incrementality is tackled by neither VISIONS nor by the model we propose: the incrementality due to the inherently dynamic nature of the external environment (incrementality from external dynamics). (2) In addition the model stops at the level of object recognition and does not attempt to make any connection to language.

VISIONS can nevertheless be seen as the historical starting point for our current model. Navalpakkam, Itti and Arbib (Navalpakkam and Itti, 2005; Itti and Arbib, 2006) went one step further. The Salience Vision and Symbolic Schema (SVSS) model remains one of the only existing model to incorporate bottom-up and top-down saliency signals in simulating saccadic system. Their model added a symbolic WM (which they

call Short Term Memory STM) that extends the visual WM of VISIONS by storing symbolic/conceptual schemas which, in conjunction with a symbolic LTM containing conceptual world knowledge, can be used to perform basic inferences that in turn shape the top-down saliency signal orienting the visual attention.

In order to focus our work on the Vision-Language interactions, we will abstract away from much of the detailed modeling of the detailed sensory-motor schemas that form the two-way interaction of the SALVIA model with the outside world. We will assume that we are working with the output of a visual processing system that incorporates the insights of both VISIONS and SVSS offering high level object recognition as well as the possibility to orient attention on the basis of both bottom-up and top-down saliency signals. The detailed neuronal implementation of the saccadic motor schemas will not be incorporate but assumed to support the high-level algorithmic gaze orientation schema (but see (Dominey et al., 1995; Dominey and Arbib, 1992)). Similarly, the model will not tackled the motor articulation schemas supporting speech production (but see (Tourville and Guenther, 2011; Bohland et al., 2010)).

SALVIA is broadly divided between two main system types: Long Term Memories (LTM) and Working Memories (WM). LTMs store the long term knowledge in the form of schema networks. The systems defines three main types of LTM: A Perceptual LTM which contains the perceptual knowledge and broadly corresponds to VISIONS's LTM, a Conceptual LTM which contains the conceptual world knowledge, and a Grammatical LTM which contains the grammatical knowledge in the form of construction schemas (see below for definition of constructions).

Each working memory is a simplification of systems that could be represented as having their own complex architecture. The various LTMs are not independent of one another: the grammatical knowledge needs to allow the system to express concepts which in turns have to be related to perceptual schemas (among others). The model makes three simplifying assumptions regarding the LTM architecture: no internal complexity within an LTM (LTM is defined as a single sub-system), interdependence of LTMs limited to what is built in at the representational level (i.e. no dynamic interactions), and no learning.

VISIONS defined its Visual WM as a WM network in which schema instances representing various types of perceptual knowledge engaged in cooperative computation (C2). SALVIA follows the same theoretical premises and expands them to define a visuo-linguistic WM that consists of a WM network involving a Visual WM (summarizing the complex Visual WM of VISIONS), a Semantic WM in which conceptual schema instances are invoked to build a semantic representation (SemRep), i.e. the semantic content to be included in the verbal description, a Grammatical WM in which grammatical schema instances are invoked to generate a mapping from the SemRep onto a verbal sequence, and finally a Phonological WM that holds the sequence of phonological information while it is relevant to the generation of an utterance (Phonological LTM not represented here). This network of WMs can be broadly divided into a Vision System and a Language System with the Semantic WM and the SemRep it contains serving as an interface allowing the interaction between the two.

Beyond the local dynamics, the scene description production processes perpetuate system level cycles (Fig.2.2, left): bottom-up attentional cues guide scene perception process impacting utterance production down the line, meanwhile top-down feedback signals emanate from the language system whose goal is to orient attention toward a location that contains the perceptual information required to pursue the utterance production (see below).

The following sections describe the various parts of the model in more details. The dynamics involving schema instances in a WM will only be discussed in detail for the Grammatical WM for which SALVIA implements the full C2. The dynamics of instances in Visual, Semantic and Phonological WM are simpler (no C2) and can be easily inferred from that of the Grammatical WM.

## 2.2.2 SALVIA: Schema Architecture

Beyond the informal framework described above, SALVIA offers a novel implemented cognitive-level model of vision-language interactions in the context of the production of visual scene descriptions. It is worth noting that this computational model also serves as a case study for Schema Theory as a general brain-theory based modeling framework (cf. ch. 1, sec. 1.4 and Ch A).

Schema Theory offers a top-down counterpart to the bottom-up neural network modeling approach. It focuses on the adaptive and dynamic nature of the interactions between distributed computational units, respecting the computational style of the brain. SALVIA is part of a new chapter in the application of

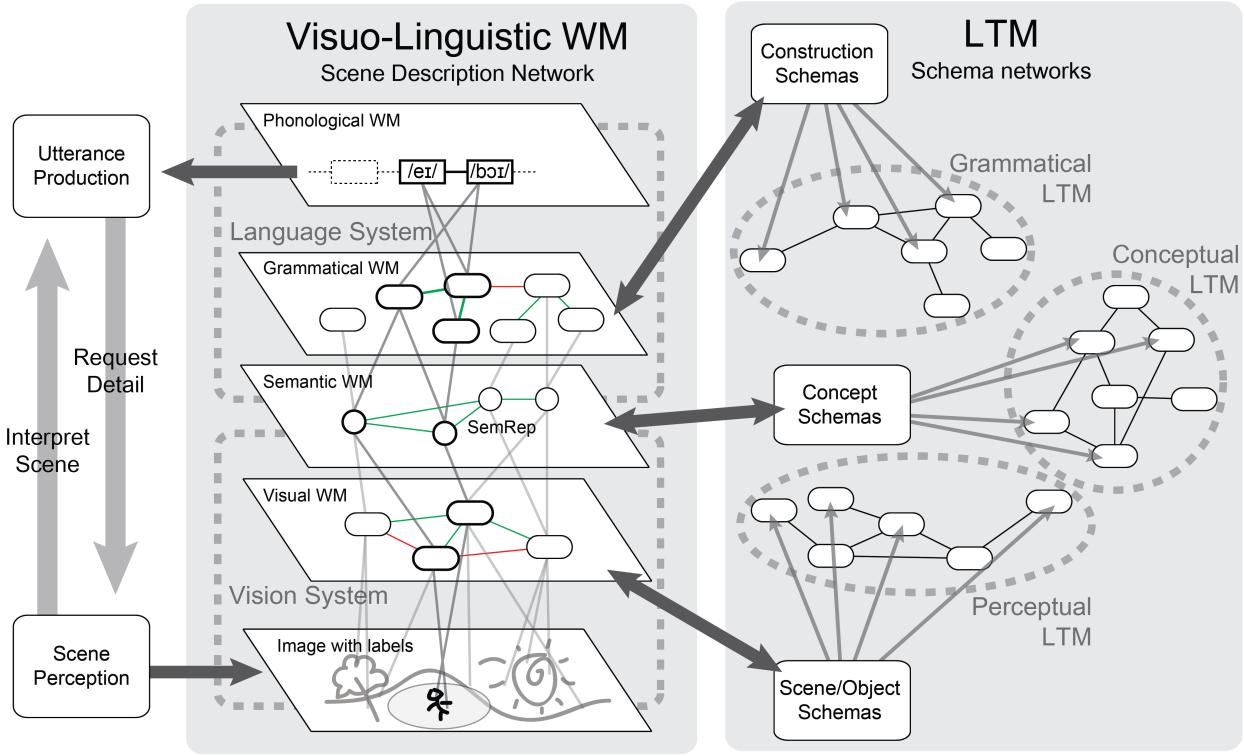


Figure 2.2: Schema Architecture Language-Vision InterAction (SALVIA) model (informal view). The model is presented here following the same conventions as that used to describe the VISIONS system in Fig.1.7 to facilitate the comparison between the two models. The right side describes the Long Term Memories (LTMs) that specify the three main sources of knowledge: Grammatical, Conceptual, and Perceptual knowledge. Those are defined as schema networks composed, respectively of Construction, Concept, and Perceptual schemas. The left side presents the Working Memories (WMs) in which the processing takes place. The Visual WM of vision has been extended by the addition of a Semantic WM, a Grammatical WM, and a Phonological WM, each associated with the LTM containing the schemas it can invoke (the Phonological knowledge is not represented as our focus will be on the grammatical processing). If the perceptual processes of VISIONS are greatly simplified, SALVIA adds visuo-attentional processes that are not present in VISIONS: The scene perception takes place through the incremental process of attentional parsing. The working memory network can be divided into a Vision and a Language system on the basis of the nature of the schemas it involves. The interactions between those two systems hinges on the state of the Semantic WM. As in VISIONS, schema instances active in WMs are not simply interacting through cooperating computation within a WM but also through linkages that span across WMs. The left most side of the figure indicates the constant interactions that exist between on the one hand bottom-up directed attentional parsing of the scene generating semantic and eventually linguistic descriptions and, on the other hand, top-down requests for perceptual information. In the Visuo-Linguistic WM, the bolded schema instances indicate those that are directly or indirectly linked to the vision of the scene currently under attentional focus (a boy). It highlights the interactions between schema processing and visual attention, without here stipulating whether the attentional focus was the result of a top-down signal propagated through the WMs, or if the bolded instances are the results of incrementally growing the states of the WMs following an attentional focus shift toward the boy (bottom-up).

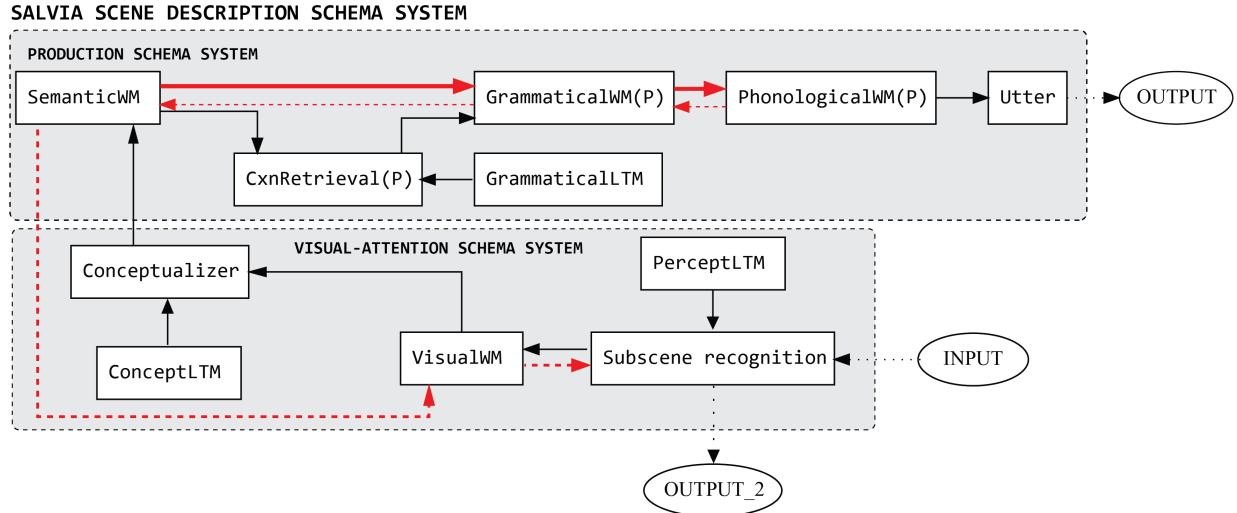


Figure 2.3: Overview of the computational SALVIA model. Each box corresponds to a system with arrows indicating message passing. As a Schema Theoretic model, SALVIA is designed as a system-of-systems. The PRODUCTION SCHEMA SYSTEM box outlines the language production system that was described in Ch. 4 (cf. fig. 4.1). The model receives a pre-processed visual scene as INPUT. OUTPUT1 corresponds to the utterance output (time stamped string sequence), OUTPUT2 to the location and size of the visual attention window at each time. The model makes use of the Schema Theoretic distinction between Working Memory (WM) and Long Term Memory (LTM) (see below). The core of the system lies in the articulation and temporal coordination of the three main WM systems: VisualWM, SemanticWM and GrammaticalWM. The key addition that SALVIA brings to the language production model is its integration with visual-attentional system, not simply as a system that receives incremental semantic content from the latter, but as a system that can in turn impact the visual processes, organizing the vision-language relations into an interactive loop. Red arrows indicate feedback signals. The suffix “(P)” is used as a reminder that those systems are part of language production processing and here no assumption is made as to their relation to the comprehension processes (cf. ch. 6 and ch. 5).

schema theory to language (Arbib et al., 1987)

At the heart of the model is the schema theory based model of grammar: Template Construction Grammar (TCG). TCG is a novel implemented computational construction grammar framework. It is part of a more general effort to develop a neurolinguistic model of vision-language interactions and follows the tenets of Schema Theory as a cognitive-level brain modeling philosophy Arbib (1989).

TCG has already been used in an implemented model of visual scene description production (Lee, 2012). But if the previous model was only approximating some of the schema theoretic computation using traditional computing paradigm, SALVIA offers a fully schema theoretic implementation of the TCG based scene description model. This chapter will only present (often informally) the aspects of TCG that are key to understand the SALVIA model. Chapter 4 is dedicated to a formal presentation of Template Construction Grammar.

The Schema Architecture Language Vision InterAction (SALVIA) model is defined as a schema system, following the general framework of Schema Theory (cf. ch. 1, sec. 1.4). In particular, the SALVIA model is a network of systems defining a system-of-systems with in particular the general Scene Description Schema System including a sub-system: the Production Schema System which encompasses the language related systems. All the processes defined by the various systems are concurrent. The arrows define the fixed connectivity between the different systems, the direction of the arrow defines the direction of the message passing.

Figure 2.3 presents the system level architecture of SALVIA, the formal, implemented, counterpart of the informal model presented in Fig 2.2. The model takes a pre-processed visual scene as INPUT and outputs words (OUTPUT1) as well as location and size of the visual attention window (OUTPUT2). Working

Memory (WM) and Long Term Memory (LTM) are defined following the principle of Schema Theory (see below). Incrementality of processing is at the core of the model with a particular focus on the temporal coordination of the three main WM systems: VisualWM, SemanticWM and GrammaticalWM (cf. fig. 2.1).

The state of VisualWM is incrementally built to contain an up-to-date representation of the perceptual content gathered through attentional parsing of the scene. The state of the SemanticWM abstract away the perceptual details that might be useful for perceptual purposes but irrelevant for verbal description purposes conceptualizing the perceptual representation. The SemanticWM therefore constantly updates the semantic representation that define the conceptual content of the description to be uttered. Finally the GrammaticalWM builds on top of the semantic representation by applying the appropriate language schema (constructions) to build a mapping from meaning (semantic representation) to form (phonological representation). Each of those three systems hosts time dependent states and the main challenge is to properly handle their temporal coordination. In red are indicated the feedback messages sent from the language system to the perceptual system which add another layer of complexity to the coordination problem. The PhonologicalWM simply hosts the current state of word sequence that have been already chosen as the basis for an utterance. Those are sequentially uttered by the model as OUTPUT1. The various LTM contains the knowledge representations that are required for the WM processes to take place (see below for a description of the role of LTM and WM in Schema Theory). Finally, the Control system impacts the dynamics of the language production systems based on the goal and task requirement.

In broad strokes and starting with a feed-forward path, an input, defining the visual scene to describe (see sec.3.2.1) is given to the Subscene recognition system. This system algorithmically simulate the process of retrieving the relevant percept schemas based on the attentional dynamics (including gaze positions and focus scope), the content of the input, and the state of the perceptual knowledge defined in the Percept LTM system (see sec. 2.3). At each time step, the percept schemas that correspond to the region of the input fixated are instantiated within the Visual WM, updating the state of the scene representation (SceneRep). At each time the fixation location and the focus scope correspond to OUTPUT2.

While the Visual WM keeps active the percept schema instances that have been so far retrieved based on the attentional parse trajectory, the Conceptualizer defines a simple algorithmic system whose role it is to conceptualize the state of the Visual WM (SceneRep) into the semantic representation (SemRep) held in Semantic WM.

The Conceptualizer in the present model is kept simple, mapping percept onto concepts with a many-to-one mapping (see sec. 2.3). This is of course much too simple as conceptualization involves many complex operations which can depend on factors such as the general task, the communicative goals etc. (see Spranger et al., 2012; Spranger and Steels, 2015) . However, we deliberately choose to include the Conceptualizer system, even in a simplified version, in order to both open avenue for future work and allow for comparisons with other models, as well as to make clear that this important step should not be overlooked within the general framework of modeling the vision-language interactions.

Starting with the Semantic WM, the systems are all part of the Production Schema System and are specifically language related systems. It is possible to run the Production Schema System by itself if the semantic inputs are directly provided to the Semantic WM. We will make use of this option at times by directly approximating the incremental gathering of perceptual information through the direct definition of an incremental semantic input (see ch. 3, fig. 3.1). A subsystem of the model can therefore be used to directly test the impact of the time dependent semantic WM state (message) on the language processing and ultimately on the type of utterances generated. Focusing now on the language-related systems, the construction retrieval system (CxnRetrieval(P)) systematically attempts to find constructions schema in Grammatical LTM whose semantic pole (SemFrame) match a sub-graph of the SemRep, i.e. are candidate to participate in mapping a sub-part of the message onto a linguistic form. Selected construction schemas are instantiated in Grammatical WM in which they can enter in the type of cooperative computation described in ch. 1, sec. 1.4.3) and that will be detailed below(see sec. 2.5).

Based on the dynamic competition and cooperation between construction instances, construction assemblages are incrementally built in Grammatical WM. When the system is ready to produce an utterance, the winning assemblage is read-out, and the resulting word sequence becomes part of the state of the Phonological WM from which the Utter system can algorithmically generate the utterance output (OUTPUT1).

The Control system is an algorithmic system whose state includes parameters defined by the task (time pressure) and some characteristics of the speaker (importance of placed on utterance continuity, utterance

compactness, etc.). The Control system, on the basis of these parameters, can impact the other language systems and affect the production dynamics (cf. sec. 2.7.2).

The model is not feed-forward but incorporates feed-back connections between modules. The state of the Phonological WM can impact the grammatical processing by forcing the system, in the case of the production of fragmented utterances, to try and generate utterance fragments that although disconnected conform to a certain level of syntactic continuity. The feedback connections between the Grammatical WM and the Semantic WM plays a key role in the case of the production of grammatical incomplete utterances. The system can indeed produce an utterance based on a construction assemblage that lacks information regarding some of the semantic content (e.g. “the boy kisses ...” where the patient is missing). In this case the Grammatical WM can send a feed-back request to the Semantic WM indicating that the part of the SemRep graph corresponding to the missing information should be expanded in order to be able to continue the utterance to which the system has committed itself. In turns, the Semantic WM can send a feed-back signal to the Visual WM which correspond to a top-down attentional signal, promoting the system to guide the visual attention towards the region of the scene that contains the relevant information.

The SALVIA architecture therefore reflects the core modeling assumptions that had been put forward by psycholinguistic models while both expanding their scope by incorporating an explicit treatment of the semantic level (message level) at its core (placing the focus on meaning rather than on syntax), by linking this level to a model of the active-vision system, and most importantly, by operationalizing those processes in a way that places time and incrementality at the heart of the modeling problem.

## 2.3 From Perceptual to Linguistic Meaning via Conceptualization

SALVIA links the language system-of-systems architecture with visuo-attentional subsystems that can pilot the incremental parsing of the visual scene, keep in perceptual working memory the relevant perceptual information retrieved, and conceptualize this information into a message that can be process by the language production system.

Fig. 2.4 highlights the visuo-attentional components of SALVIA. Some sub-systems are only given a simple implementation. It is however the strength of schema theory to allow the building of models that can have sub-systems specified at various degrees of granularity, opening the possibility for future refinement or integration with other models. In the present case, one way to understand the visual-attention schema system is as a place-holder for a model such as the one proposed by (Navalpakkam and Itti, 2005) and discussed in ch. 1 fig. 1.4.

The INPUT given to the model is a scene associated with a predefined sub-scene hierarchical structure (see fig. 2.10). The OUTPUT provides, at each time step the coordinate of the center of gaze in the scene as well as the size of its associated attentional window, standing for the fact that, for a given fixation point, the associated covert attention focus can vary in scale from narrow, detail oriented to large. The Visual WM holds the perceptual schema instances incrementally retrieved from PerceptualLTM.

The Conceptualizer only host simple mapping processes from perceptual to conceptual representations, but its presence as a sub-system is crucial if only to indicate that a large chunk of the work that has to be done to understand the language-vision interactions have to focus on the question of how such mappings are learned and applied on the basis of communicative goals.

The feedback signal emanating from the SemanticWM makes explicit the core hypothesis that SALVIA implements: during language production, the language production system can opportunistically interrogate the visual system and pilot the attentional systems from the top-down. The hypothesis here is that this interaction takes place through the Semantic WM. However, given some of the questions regarding the embodied nature of the linguistic meaning (cf. Ch. 8), other avenue of communication (possibly more direct) could also exist, but are beyond the scope of this work

### Visual WM

While VISIONS applied processes to the entire scene at once, in SALVIA, as the attentional focus lands on various regions of the scene, those regions receive an interpretation in the form of perceptual schema instances instantiated in Visual WM. SALVIA does not replicate the work of VISIONS but assumes that the perceptual system is able to deliver high-level scene structures and representations (e.g. not simply objects

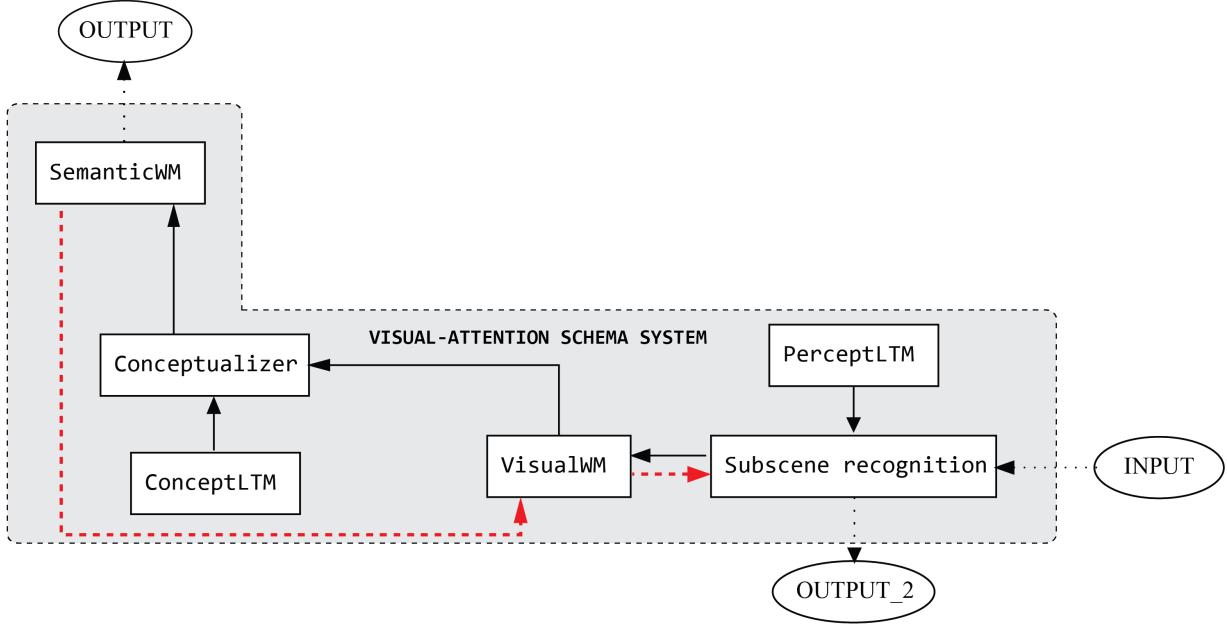


Figure 2.4: Focus on the Visual-Attention Schema System. SALVIA only implements a simplified visuo-attentional system. The goal of this system is to conveniently carry out the minimum set of operations necessary to highlight and simulate the interactions between feedforward (solid black arrows) and feedback signals (red dashed arrows) affecting the attention driven building of a high level cognitive representation of the scene at hand.

tags but also, actions as well as relations). The Perceptual LTM contains a perceptual schema network that defines the perceptual knowledge endowed to the system (see sec. 2.3)

SALVIA defines a minimal perceptual knowledge with perceptual schemas simply defining type and tokens hierarchy. Those schemas do not have vocation to be applied to the processing of the image input which will be assumed to be handled by a system akin to VISIONS not implemented here. The actual perceptual knowledge associated with each schemas that would allow it to actively participate in the scene interpretation process need not be stipulated.

SALVIA essentially assumes that the perceptual schema instances that can be invoked in VISUAL WM to form a scene interpretation is given as input. The incremental order in which those schemas instances are invoked, updating the state of the Visual WM, depends on the visual-attentional process (see below). Each perceptual schema instances active in Visual WM, as in VISIONS, carries a pointer back to the spatial region it covers and (partially) interpret.

### Perceptual Knowledge

Fig. 2.5 presents the content of the perceptual knowledge that seeds the PerceptLTM, defined as a schema network.

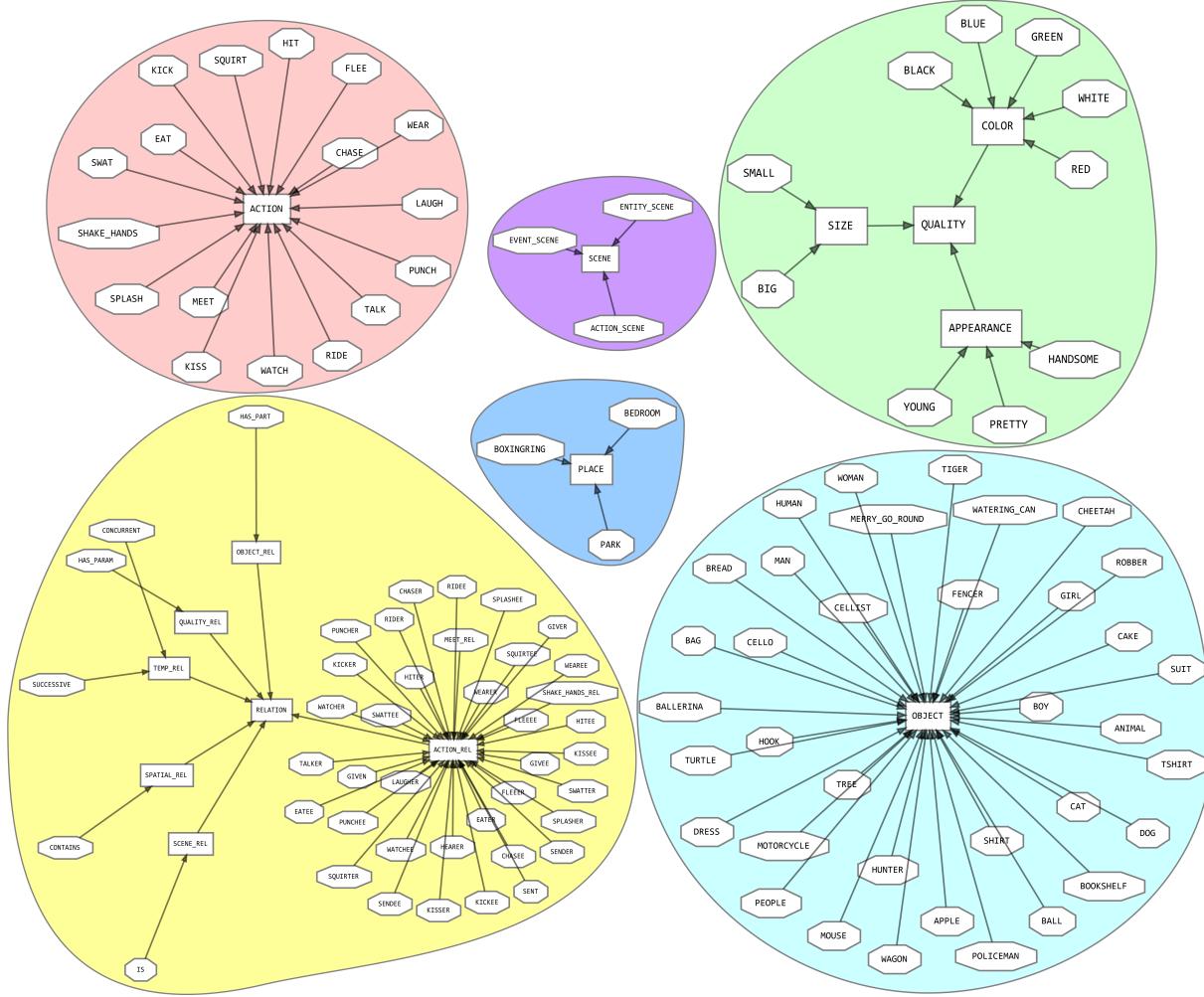


Figure 2.5: Example of perceptual knowledge format used to define the state of the SALVIA PerceptLTM. In this example, a simple subsumption ontology is used (arrows  $X \leftarrow Y$  signify that “ $X$  IS-A  $Y$ ”). Rectangular nodes correspond to type of perceptual knowledge whereas hexagonal nodes stand for tokens of perceptual knowledge. ([victorbarres.github.io/media/SALVIA\\_percepts.png](http://victorbarres.github.io/media/SALVIA_percepts.png))

Although the actual perceptual processes are not implemented in SALVIA, the distinction between types and tokens stand for the distinction between type level perceptual knowledge (for example COLOR can be defined over the hue spaces) and specific tokens of such types (for example GREEN correspond to a sub-space of the hue-space, or a specific person X can be linked to a specified HUMAN perceptual schema.) Of course this approach has an important drawback: It’s approach to the gradedness of the perceptual knowledge is simplistic and it does not capture the pervasive role of analogical similarities, the flexibility of the categorizations, etc. However, it insists on the fact that categorization does not only occur at a result of conceptualization. Colors can be conceptualized differently according to the linguistic background of an individual, but this does not mean that at a pure perceptual level, if behaviorally relevant, some other types of categorizations take place , irrespective of their mapping onto linguistic driven concepts, mapping that might not capture the distinctions made at the perceptual level. Conceptualizations and concepts are constrained by their social use through language acts, while perceptual categorization are essentially private. In a nutshell, the type-token distinction at the perceptual knowledge level can be seen as reflecting perceptual construals that are somewhat independent of the linguistic ones.

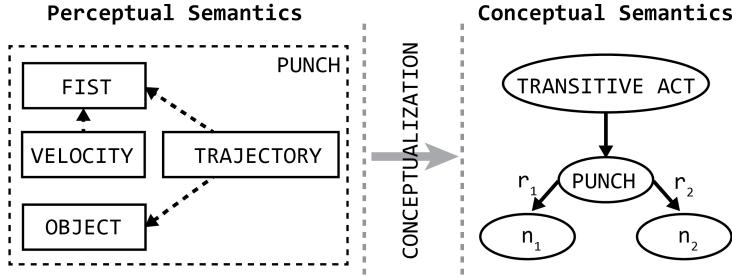


Figure 2.6: From cognitive level scene representation (SceneRep) to a semantic representation (SemRep). Although SALVIA does not handle the complex processes underlying the conceptualization of perceptual content, it holds that such processes cannot be overlooked and, insists that future work will need to focus on this issue. The figure illustrates informally the core difference that exists between perceptual and conceptual semantics. (Left) An informal example of a high level perceptual recognition of a PUNCH action. It involves recognizing, among other things, objects (FIST), its dynamic profile in space, its trajectory in relation to a object goal. All those are very complex perceptual recognition processes that can, at some level, be perceptually represented as parametrized form of PUNCH perceptual schema. (Right) This perceptual representation can then be conceptualized and abstracted into a semantic conceptual representation that only retains a silhouette of the action that, for example, profiles it as a transitive action linked to two thematic roles. Here most of the parametric information has been discarded.

### Conceptualization

Conceptualizations are defined as percept to concept mappings. The conceptualizer maps percepts schema instances invoked in visual WM to concept schema instances invoked from Concept LTM into Semantic WM. Although much work could be dedicated to the conceptualization process, the current model keeps this stage very simple. The underlying purpose is both to show the necessity of including this step in any model tackling vision-language interactions while leaving the tackling of this crucial problem for further work.

Having a conceptualization module, allows a more direct comparison of the SALVIA model with other models and in particular with the Fluid Construction Grammar based suites of architectures for embodied language use, acquisition, and cultural evolution<sup>1</sup>. In addition, in the context of building a schema theoretic model with the explicit goal to incorporate the possibility to test for the impact of various types deterioration on the system’s behavior, having a conceptualization module will offer the possibility to analyze, at least at a coarse level, the difference in overt linguistic behavior caused by functional deterioration of the VisualWM compared to a deterioration of the SemanticWM, as well as, potential deterioration in the conceptualization process itself.

The Conceptualizer system is initialized with the conceptualization knowledge (see sec. 2.3). The function of Conceptualizer is made very simple by constraining the conceptualization mapping (taken to be the set of all the mappings) to be a function. Therefore, given a state of the VisualWM, the conceptualizer can simply apply the conceptualization function to the percept schema instances forming the SceneRep to generate the unique corresponding set of concept schema instances that are to be instantiated in SemanticWM.

As concept schemas are instantiated in SemanticWM conceptualizing the (changing) state of the VisualWM, the initial activity of the concept schema instances is a function of the current activity of the perceptual schema instances. Conceptualization always results in cross WM links being built between the concept schema instances and the perceptual schema instances they conceptualize. This allows for activation signals to flow across WMs with changes in activation levels in perceptual schema instances impacting the state of the Semantic WM. Such cross WM links play a key role in coordinating the activity levels at the system level.

<sup>1</sup>We refer the reader to (Spranger and Steels, 2015) for a thorough investigation of the acquisition of spatial relations conceptualizations.

## Conceptualization Knowledge

As discussed above, the conceptualizer in SALVIA simply maps perceptual schema instances onto concept schema instances (see sec. 2.3). It serves to update the state of the SemanticWM based on the changing state of the VisualWM: this is not here an active process in the sense that conceptualization is deterministic and fails to capture the goal oriented decisions that have to take place at this level (but see work in both FCG and ECG for in depth discussions of this issue Spranger and Steels, 2015; Chang et al., 2006). The presence of this subsystem is nevertheless crucial as a placeholder in the system for further investigation of the conceptualization function.

Fig. 2.7 provides a snapshot of the conceptualization knowledge SALVIA is endowed with.

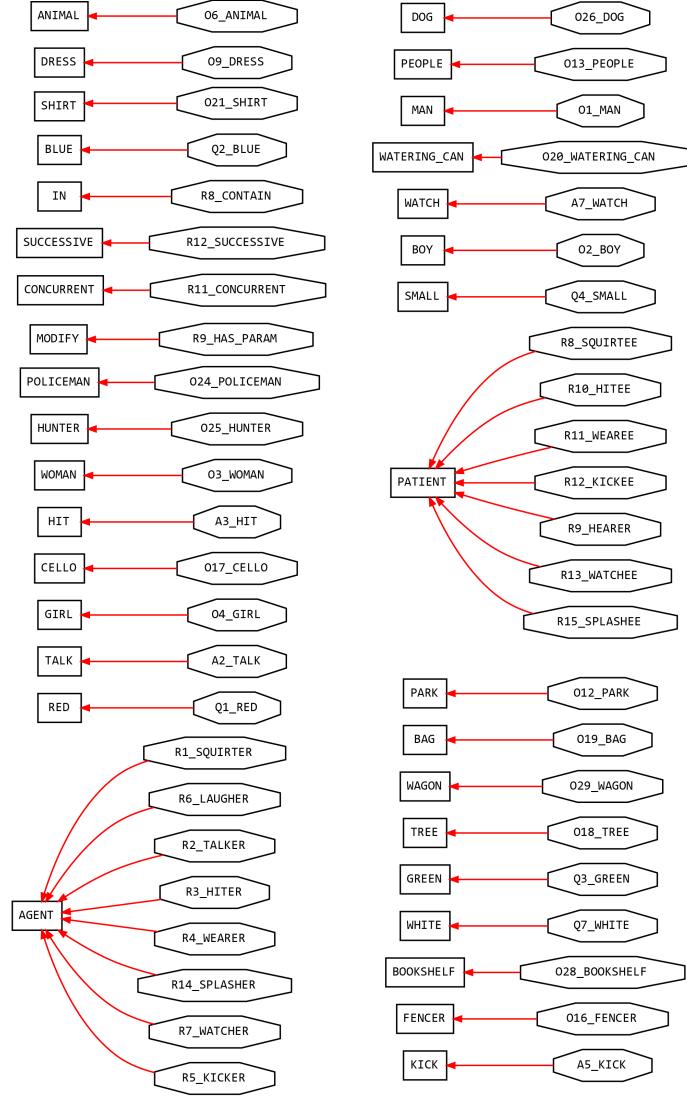


Figure 2.7: Conceptualizations as many-to-one mapping. Multiple perceptual schemas can be mapped onto the same concept schema. ([victorbarres.github.io/media/SALVIA\\_conceptualizer.png](https://victorbarres.github.io/media/SALVIA_conceptualizer.png))

## Semantic WM: Building the Semantic Representation

As shown in Figure 2.2, the Conceptual LTM defines a network of concept schemas. This forms a repository for a semantic network model of world knowledge in which concept schemas are connected through hypernym

(IS\_A) relations. Concept schemas are limited to four types: ENTITY, ACTION, PROPERTY, and semantic RELATION.

Conceptual schema instances are invoked in Semantic WM to form a Semantic Representation (SemRep). The instances are invoked if they contribute to the conceptualization of the perceptual scene representation (state of the Visual WM). At each time step, the SemRep specifies, the semantic content of the message that has to be expressed linguistically as shown in Figure 2.8.

The expressiveness of the semantic representation is limited in order to focus on its time dependent nature as an incremental and dynamic semantic structure and on its role as coordination structure between incremental visual processes and incremental language processes (Discussed below, see Fig. 2.14).

Since all the conceptual relations are binary, the SemRep is conveniently expressed as a labeled (not necessarily connected) directed graph: edges correspond to RELATION, while nodes correspond to ENTITY, ACTION, or PROPERTY concept schema instances. No cooperative computation is implemented within Semantic WM (i.e. the semantic message does not contain any conflict).

The SemRep only encompasses the information that is relevant for the language system. There might be a large amount of visual information that is used in building the Scene Representation but that is not transferred into the SemRep. For example, the red color of a roof, as perceptual information, can be a useful cue to help in the segmentation of the roof from neighboring sky and walls, and in turn, the recognition of the roof might be an important step in recognizing a house. However, at the SemRep level, all this perceptual information might lead to the instantiation only of a HOUSE concept, the semantic information to be linguistically conveyed, abstracting away from all the perceptual details that have supported the recognition of the object the concept refers to.

In SALVIA the conceptualization that pilots the invocation of conceptual schemas is kept simple and consists of deterministic many-to-one mappings between perceptual schemas and conceptual schemas. This does not capture the active processes by which a conceptualization is chosen amongst many possible candidates. However, it provides the minimal architecture that captures the dynamic nature of the interactions between visual and semantic representations.

In the context of vision-language interactions, as in VISIONS, and in line with the theories of situated cognition (Ballard et al., 1997; Pylyshyn, 2000, 2001; Kahneman et al., 1992), if a SemRep abstracts away much of the perceptual details used by the visual system in the process of attentionally parsing a scene, any part of the SemRep can serve as a ‘deictic pointer’ Ballard et al. (1997) capable of re-orienting the attentional focus back to the visual region the concept refers to. This enables the system to flexibly use the ‘external world as memory’ O’Regan (1992): The SemRep can be incrementally updated by requesting more details from the perceptual system through the reorientation of attention towards the region of the external world where this information is most likely to be found.

From the focus of the current work is on the online dynamics of the vision-language interactions (in the context of visual scene description production) follows that the model stresses the incremental and dynamic nature of the semantic representation. As more perceptual information is gathered through visuo-attentional exploration of the visual scene, the content of the SemRep graph incrementally grows to dynamically update the content of the message. The goal of grammatical processing is to generate a flexible grammatical structure articulating the incrementally built SemRep and the production of utterances.

### 2.3.1 Conceptual Knowledge

The conceptual knowledge of the system is defined as a knowledge network. It defines a simple subsumption ontology. The structure of the model and of the processes it supports however, does not restrict the model to using this type of representation. The main requirement is that measures of similarity can be performed on the representations. Pilot work using vector space representations has been carried out, but for the present purpose, the ontology is sufficient.

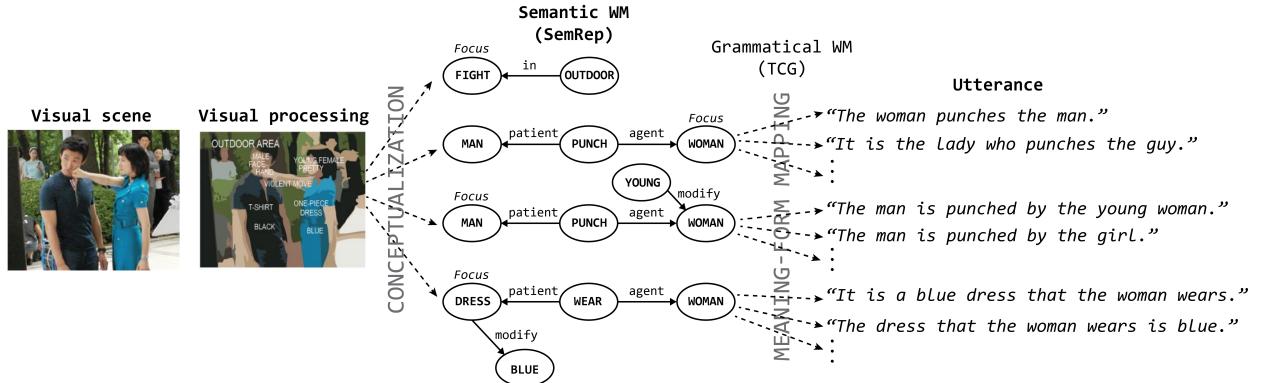


Figure 2.8: From visual processing to utterances: A many-to-many mapping. From left to right. The visual scene SALVIA has to describe. The visual processing assumes the existence of complex perceptual processes able to generate high-level perceptual interpretations of the scene input. VISIONS (see below) outlined what could be part of such a system. Recently, an great amount of models have been proposed to generate scene interpretations. However powerful their abilities are compared to the historical VISIONS model, they usually suffer from the same serious limitation in what they count as perceptual representations, usually limiting it to tagging, unable to reach structured interpretations involving both entities and their relations (in particular in the case of actions). So it will be assumed here that such models have been able to overcome this limitation. Based on the outcome of perceptual processing, the perceptual representation can be conceptualized in many ways. Conceptualizations can vary in the semantic content they encode (for example in its scope: an outdoor fight vs. a woman wearing a blue dress), but also in terms of what semantic information is highlighted (Focus). For the punching action, the focus can be on the agent (WOMAN) or on the patient (MAN). At the utterance level, resulting from grammatical processing generating meaning-to-form mappings, a given SemRep can yield different linguistic forms: YOUNG(WOMAN) can be expressed as “young woman” or as “girl”, a focus on the agent (WOMAN) can be expressed in the use of an active voice (mild focus) or in a cleft subject (strong focus), MAN can be lexicalized as “man” or “guy”. The paths from given perceptual representation to utterances form a one-to-many mapping. However, for a given scene, there are actually many ways a perceptual representation can be built. The whole system consists therefore of a many-to-many mapping.

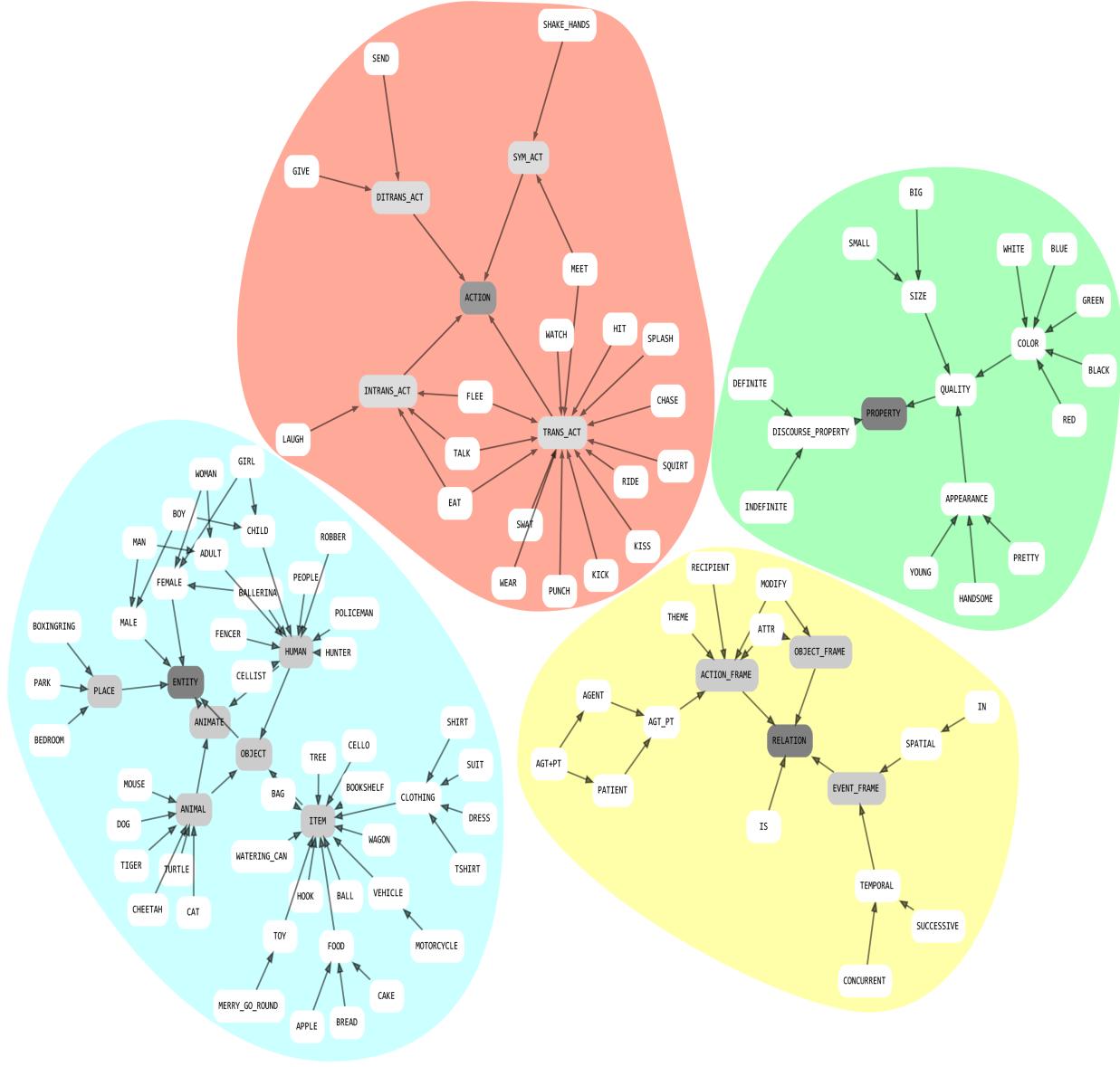


Figure 2.9: Example of conceptual knowledge format used to define the state of the SALVIA ConceptLTM. In this example, a simple subsumption ontology is used. Arrows  $X \leftarrow Y$  signify that “ $X$  IS-A  $Y$ ” ([victor-barres.github.io/media/SALVIA\\_concepts.png](http://victor-barres.github.io/media/SALVIA_concepts.png)).

### 2.3.2 Long Term Memories: Various Types of Knowledge

The model defines three types of long term memories (LTMs), each corresponding to a specific domain of knowledge: perceptual knowledge (Percept LTM), conceptual knowledge (Concept LTM), and grammatical knowledge (Grammatical LTM) (see fig. 2.2). The content of the knowledge schemas is defined declaratively and for now there is no learning taking place in the model. We do not claim that those types of knowledge are not at all related when it comes to the architecture of the brains semantic system, however, we consider that for our present purpose, they can be thought of as distinct as a first approximation. The question of their interrelations both synchronically, ontogenetically, and even diachronically (in the case of the cultural transmission of language and of world knowledge) corresponds to an entire line of investigation on its own, far beyond the scope of the present work.



Figure 2.10: An example of scene representation (SceneRep) as a hierarchy of subscenes (See text for details.)

## 2.4 Visual Attention

### 2.4.1 Subscenes: A Cognitive-Level Scene Structuring Principle

As was already mentioned in the previous sections, SALVIA assumes that, on the visual side, the system is able to build complex scene representations. At a cognitive level, a scene representation (SceneRep) cannot be taken to be a simple set of objects. Itti and Arbib (2006) proposed that visual scene representations are organized around the construct of (minimal) **subscenes**. As the visuo-attentional system parses a scene, in the context of a given task and goal, information about the scene is structured in a hierarchical network of subscenes that accumulate and package the perceptual content into meaningful cognitive entities. Crucially, such structure need not be “for language” but can be tied to different goals: for a similar scene, one will perceptually construct high-level scene representation differently depending on whether the goal is to find an object, perform an action, or describe the scene. From an evolutionary perspective, the notion of high level complex scene representation need also not be specific to humans but is likely share with non-human primates<sup>2</sup>.

The evolutionary question of how such complex perceptual cognitive structure could have been the basis on which similarly complex structures specifically “for language” (SemRep) appeared is far beyond the scope of this work but is part of the overall research program that it is inscribed into (Arbib and Bonaiuto, 2008; Arbib, 2016b; Arbib and Lee, 2007; Arbib, 2016a, 2010)

### 2.4.2 Subscene Recognition

In SALVIA, the sub-scene representation level plays a key role. The Subscene Recognition system functions as an algorithmic placeholder for a host of complex perceptuo-attentional processes. It focuses on implementing both attention orientation as well as focus selection

Ignoring for now the role of focus and of top-down attentional signals, the Subscene Recognition schema, will successively generate saccades to each of the subscenes defined in the scene input based solely on their saliency value (starting from the most salient subscene to the least salient subscene). Upon orienting the attention to a given subscene, the Subscene Recognition system triggers the instantiation of the perceptual schema instances associated with the subscene into VisualWM. An uncertainty value associated with a subscene impacts the duration of this process.

For now the smallest chunk of perceptual information instantiated at a time in VisualWM by the SALVIA model is that of a subscene (ie a perceptual schema structure). However, there are for now no limitation as to what perceptual schema content is required to form a subscene, so technically each perceptual schema could be given its own subscene, leading the system to proceed a perceptual schema at a time.

<sup>2</sup>All organisms build structures of perception for action, but here the assumption is that the precise nature of those representations might be shared with non-human primates.

The current version of the model implements a very basic mode of inhibition of return (IOR) (the content of each subscene can only be retrieved once)<sup>3</sup>.

### 2.4.3 Attentional Parsing: Building and Navigating Hierarchical Scene Representations

As mentioned above, the visuo-attentional parsing of a visual scene is not only incremental in the eye movements but also in the succession of areas that are in focus. This first attentional process corresponds to attention *shifting*. However, to take into account the multiscale, hierarchical nature of the cognitive representations of visual scenes, shifting attention is at work alongside two other attentional processes *zooming-in*, *zooming-out*.

Instead of simply necessarily working at the level of the entire visual scene, the state of the Subscene Recognition system includes, at each time step, the specification of a *focus area*. At each time step, the process of retrieval of subscene related perceptual information is bound by the current focus area: everything outside the focus area is ignored until there is either nothing left to perceive within this specific focus area (in which case the system revert to the whole scene as a default focus area), or until the focus area is modified based on upcoming signals from another module.

The inclusion of a varying focus area in addition to the saliency driven attentional process, endows the rather simple Subscene Recognition system with the possibility to model not only the incremental retrieval of perceptual information based on bottom-up saliency, but also the impact of top-down signals emanating from the language system (see next paragraph), as well as various **perceptual strategies** (e.g. breadth-first vs. depth-first) that emerge from, among other thing, interactions between goal, cognitive constraints, and input type.

Perceptual strategies will play an important role in the way SALVIA simulates psycholinguistic results (cf. ch. 3, sec. 3.6).

## 2.5 Grammatical Processing

The goal of grammatical processing is to generate a flexible grammatical structure articulating the incrementally built SemRep and the production of utterances. The discussion of the model’s grammatical processing will focus on the key points that are key to understand its functioning in relation to the problem at hand. We refer the reader to ch. 4 for a full exposé of the theory behind the grammatical model.

### 2.5.1 Template Construction Grammar (TCG)

We propose Template Construction Grammar (TCG) as the basis for a schema theory model of grammatical processing. TCG, as a computational construction grammar, builds on the insights of more complex symbolic models (Embodied Construction Grammar and Fluid Construction Grammar) Steels (2011); Feldman (2010); Bergen and Chang (2005). TCG however significantly reduces the complexity of the semantic and grammatical representations tackled in order to better focus on the use of the constructions as language schemas engaging in C2. A first version of TCG has already been presented both conceptually and computationally in (Arbib and Lee, 2008, 2007; Lee, 2012). The present work, builds on top of the previous work while adding some important changes to the model.

A core tenet of construction grammar is that syntax and semantics are not disjoint: each unit of grammatical knowledge forming a symbolic whole linking meaning and form knowledge. The example of the verb alternation contrast between the *Theme-object construction* “X V Y in/on Z” (Cxn1) and the *Goal-object construction* “X V Z with Y” (Cxn2) illustrates this point (Pinker, 2000). They are both abstract argument structure constructions but their semantic requirements for the verb slot V vary.

(1) “John pours water in the glass”

(2) \*“John pours the glass with water”

---

<sup>3</sup>The process of integrating SALVIA with the SVSS model (Itti and Arbib, 2006) will tackle in much more details the question of modeling bottom-up attentional processes, including IOR.

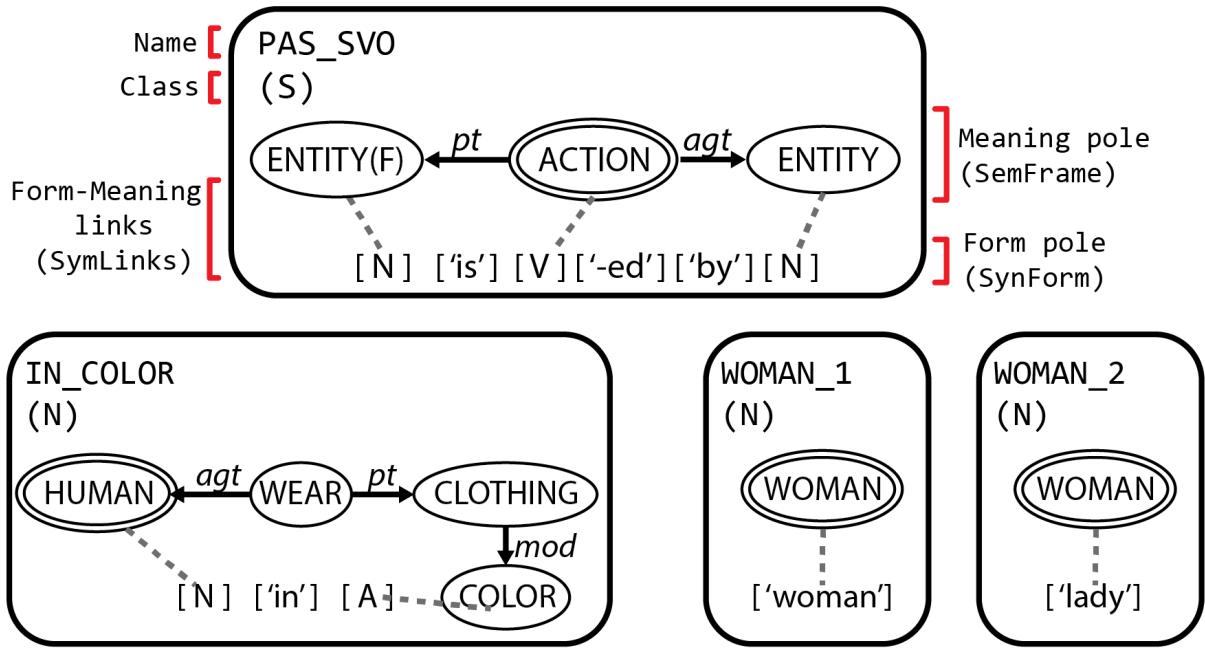


Figure 2.11: Constructions in Template Construction Grammar: A few examples. Constructions have a meaning pole (SemFrame) defined as a semantic graph following the SemRep format defined above. A form pole (SynForm) is defined as a sequence of slots (undetermined form) and phonological forms (fully determined form). Slots have vocation to be filled by the SynForm information from another construction and include class constraints limiting the types of construction that can play such a role (e.g. [N] defines a slot with the class constraint N). Symbolic links (SymLinks) define the relations between form and meaning: which part of the SynForm expresses which part of the SemFrame. Those are limited in the current framework to link between SemFrame nodes and slots. Each construction is assigned a class which in the present case is chosen to resemble canonical syntactic classes, but need not be. Constructions range from lexical constructions (WOMAN\_1, WOMAN\_2) that are fully lexicalized, to partially lexicalized constructions (IN\_COLOR), all the way to constructions with little or no phonological specification (PAS\_SVO). Double circle SemFrame nodes mark head nodes.

(1') \*“John filled water in the glass”

(2') “John filled the glass with water”

Cxn1 accepts “pour” but not “fill” in its verb slot, while the reverse is true for Cxn2. Both share the same syntactic class, and are very close in term of general conceptual/world knowledge. But they differ in the fact that “pour” only specifies the path to the object and not the change of state of the object, “fill” only specifies the change of state, not the manner by which it is achieved. This difference between the two concepts “pour” and “fill” reveals the difference in semantic constraints associated to the V slot in Cxn1 and Cxn2. Some verbs can satisfy equally the constraints of both constructions (cf. “sprinkle”).

A slot in a construction vary in its sensitivity to meaning, stipulating “grammatical semantics” constraints that differ from the full-fledge multi-facetted world knowledge associated with a lexical item: grammatical semantics usually represent a much “lighter” or “bleached” version of world-knowledge, keeping only the linguistically relevant features (which in some case can end up being so bleached that they can be simply summarized by a “syntactic class”). In the case of the ditransitive construction Subject Verb Object1 Object2 (e.g. “Bill kicked Bob the ball”), the construction is sensitive to the specific contrast between (1) verbs of instantaneous causation of ballistic motion which are acceptable (“I kicked/tossed/rolled/bounced him the ball”) and (2) the verbs of continuous causation of accompanied motion that are not (\*‘I carried/hailed/lifted/dragged him the ball’). The construction however, is not sensitive to the many world-knowledge differences that exist between the verbs within each group.

Figure 2.11 presents a few construction examples illustrating those features. Each construction is assigned a class. If for simplicity the classes used here are similar to the classic syntactic classes, there is no a priori constraint on the number or nature of those classes. Following the main tenets of cognitive linguistics focusing of language in use, linguistic knowledge is not divided into components (phonology, syntax, semantics, and pragmatics), rather any construction can potentially cut across all those strata. For this reason, constructions ran the gamut from lexical constructions (e.g. WOMAN\_1, WOMAN\_2) all the way to argument structure constructions (e.g. PAS\_SVO).

The meaning pole of each construction (SemFrame) is represented using the SemRep format with additional features. A head node indicates the semantic head of a construction. A focus feature F can be associated with a node to encode the information structure features carried by the construction’s meaning pole (cf. PAS\_SVO).

The form pole of constructions (SynForm) is limited to representing sequences of phonological forms and slots. Slots play a key role as variables that need to be filled by the form of another cooperating construction. Slots also express constraints on the constructions that can be used as filler (set of admissible construction classes).

The mapping between meaning and form is defined through symbolic links (SymLinks, dashed lines) linking semantic to form elements, denoting that a specific form element symbolizes a given part of the meaning. In the current format, symbolic links only appear between a node and a form element, any semantic element that is not associated with a symbolic link is assumed to be de facto symbolized by the construction although the nature of this symbolization is not stipulated. Semantic relations (SemFrame edges) are always symbolically represented in the form (e.g. as sequential relations). Similarly some semantic nodes can appear in the meaning pole that are not symbolically linked to any form element (c.f. IN\_COLOR construction).

*Preference* and *Group* features can be added to the constructions. *Preference* captures usage preferences (e.g. defined from usage frequency) and during processing modulates the initial activation value of construction schema instances. *Group* defines construction subsets (e.g. lexical and grammatical constructions) that can then be processed differently.

A grammar  $\mathcal{G}$  is a set of constructions  $\{Cxn_i\}$ . As a construction always includes a SemFrame which is defined in terms of concepts, a construction, and by extension the whole grammar, is necessarily defined in relation to a conceptual knowledge. The model does not impose a particular content for the grammar and offers the option to write and test new grammars using simple json format.

A language schema or *construction schema* defines a functional unit of grammatical knowledge. The construction schema is defined as a tuple

$$(Cxn, act^0)$$

where  $Cxn$  is a construction as defined above, and  $act^0 \in [0, 1]$  is a scalar value used to define the initial activation value when an instance of the schema is invoked.

Although schema theory hypothesizes that long term memory (LTM) should be represented as a schema network, TCG in its current version simply models Grammatical LTM as the set of all construction schemas defined based on the grammar<sup>4</sup>.

Based on the state of the Semantic WM in which the message to be conveyed is incrementally built (SemRep), construction schemas can be instantiated in Grammatical WM if the meaning-to-form mapping they carry represents a possible candidate to participate in building the general translation of meaning to form. In Grammatical WM, construction instances enter in cooperative computation (C2). Through the process of competition and cooperation, they generate construction assemblages, each representing a potential (possibly partial) self-organized program to translate the message (SemRep) into a phonological form.

Each construction schema instances can be thought of as a “constructor”, which, in cooperation with other, can flexibly build a program articulating meaning and form. The next sections details the process C2 processes.

## 2.5.2 Grammatical WM

At each time step, the state of the Grammatical WM is defined by the construction schema instances that are currently active as well as by the cooperation and competition links (C2 links) that they have established and, forming a competition-cooperation network (C2 network) that governs the cooperative computation.

Figure 2.12 illustrates key points of the cooperative computation process through a simplified example. Construction instances compete and cooperate to form a winning construction instance assemblage that will express the information contained in the SemRep shown at the center (since we show together the construction instances and the SemRep, this combines the states of the Semantic and Grammatical WM which we consider to form the Linguistic WM).

Lexical level construction instances attempt to map individual SemRep nodes onto linguistic forms with competition taking where lexical synonymy appears. WOMAN\_1 and WOMAN\_2 compete as they proposes two different hypotheses for the mapping of semantic content of the WOMAN node onto a lexeme. One can note that WOMAN\_1 is already winning the competition, possibly reflecting idiosyncratic preferences or other usage-based differences between the two constructions.

Argument structure constructions instances express the whole SemRep frame of a transitive action performed by an agent onto a patient. Here again, competition takes place between the SVO representing the active voice transitive construction, and PAS\_SVO representing the passive voice transitive construction. The two compete as they overlap on the SemRep edges and therefore represent different ways of linguistically expressing the semantic relations.

Through their open variable in their SynForm (SLOTS), SVO and PAS\_SVO are also built on top of the lexical constructions with which they form cooperative links. The figure shows that the PAS\_SVO instance's activation is higher than that of the SVO instance, resulting from the fact that MAN node in the SemRep has a higher activation than WOMAN node which favors its position as FOCUS, as stipulated by the information structure of the PAS\_SVO SemFrame (but not by that of SVO.)

### Construction Schema Instantiation

As shown in Figure 2.2 the Grammatical LTM consists of a network of construction schemas. At each time step, the state of the Semantic WM can be updated, following the incremental process of building the semantic representation (SemRep). When new SemRep nodes or edges (i.e. conceptual schema instances) are invoked in Semantic WM, constructions whose SemFrame semantically match (SemMatch) a SemRep subgraph that contains those new elements are invoked as instances in Grammatical WM (see ch. 4, sec. 4.4.4). A semantic match between a SemRep subgraph and a construction schema SemFrame indicates that the construction expresses in its form, at least in part, the semantic content of this subgraph and is therefore a candidate hypothesis for participating in the mapping of the SemRep onto a linguistic form in Grammatical WM (see fig. 2.12, SVO and PAS\_SVO have been invoked as their SemFrames are a semantic match with the SemRep.)

---

<sup>4</sup>Future work will need to explore the possibility of using a dynamic network allowing for cross-priming effects between construction schemas. But see (Wellens and Steels, 2011)

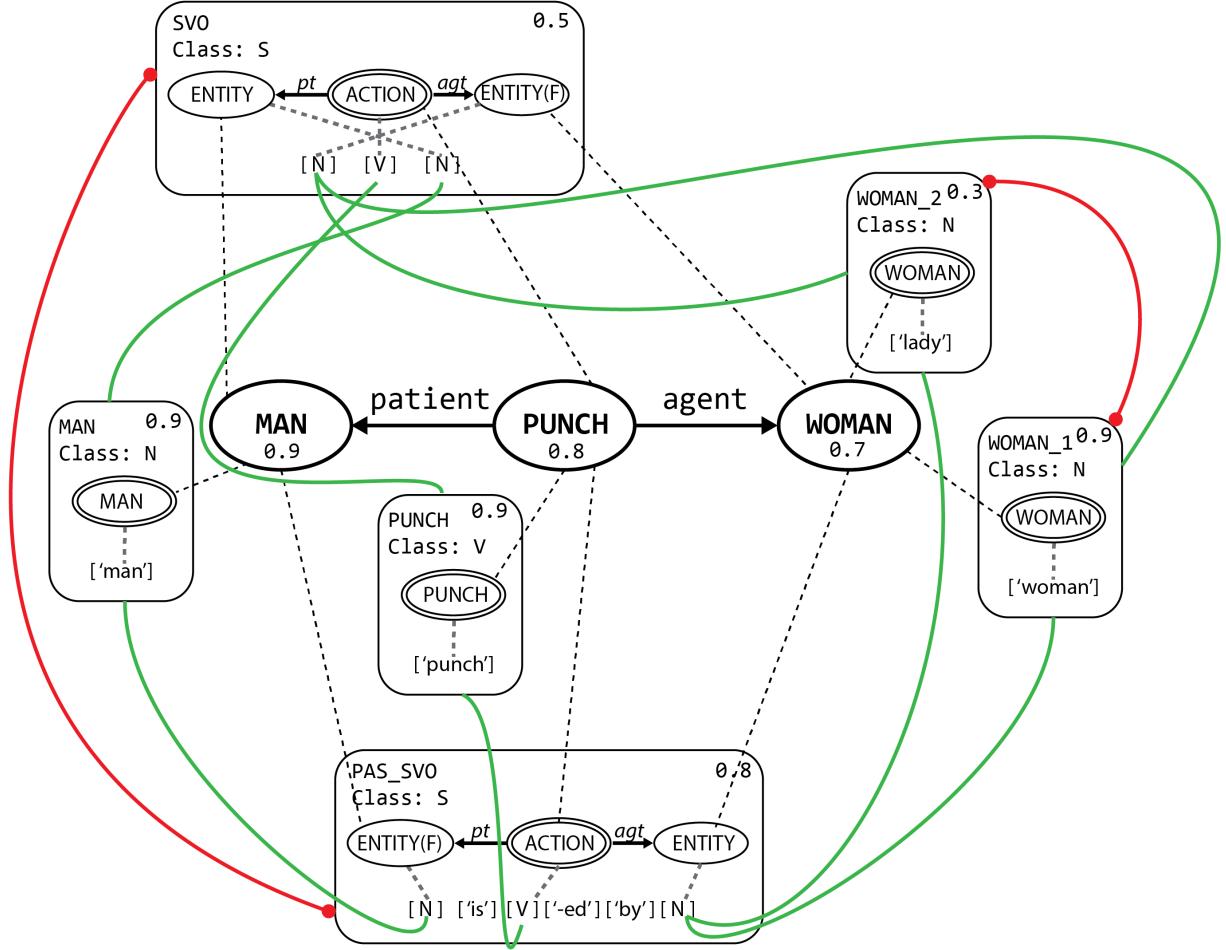


Figure 2.12: Example of cooperative computation (C2) in Linguistic WM (Semantic WM + Grammatical WM). The SemRep is shown at the center and corresponds to the state of the Semantic WM. The constructions are shown forming a cooperation (green) and competition (red) network which corresponds to the state of the Grammatical WM (C2 network). Taken together, Semantic and Grammatical WM form a Linguistic WM. The dashed lines linking constructions' SemFrames to SemRep nodes represent the portion of the SemRep that each construction covers and expresses (partially) in its SynForm. They correspond to cross WM cooperation links. Each construction is shown with an activation value on the top-right corner. Competition takes place between lexical constructions: WOMAN\_1 and WOMAN\_2 compete as synonymous lexical constructions, with WOMAN\_1 winning. Competition also takes place at the more abstract level of argument structure/voice: PAS\_SVO and SVO compete as they both build on top of the same portion of the SemRep but express the agent-patient semantic roles in different ways in their SynForm (SVO places the agent as subject while PAS\_SVO places the patient as subject). PAS\_SVO is winning reflecting the fact that it stipulates a patient focus and that the patient MAN as a higher activation than the agent WOMAN. (See Ch. 4, fig. 4.5 for version of this figure as output of SALVIA computation.)

### **Cooperative Computation: Building the Competition-Cooperation Network**

The goal of the Grammatical WM consists in incrementally building mappings to express the semantic content of the SemRep (itself built incrementally) in a linguistic form. Construction schemas that correspond to relevant meaning-form mapping hypotheses are invoked in Grammatical WM (see above) where they enter in cooperative computation (C2).

Each construction instance carries an activation value, whose initial value is modulated by the preference value stored in the schema, representing the idiosyncratic usage preferences of the speaker (to which can be added a factor reflecting the quality of the semantic match). They organize into a C2 network, whose dynamics defines at each time step the values of the instances activation values. If a construction instances activation value falls below a given threshold, the instance is pruned out of the Grammatical WM. The C2 network is therefore intermittently reshaped following either the invocation of new constructions instances or the pruning of construction instances that “lost” the competitions in which they were involved.

C2 links are built based on the “Match” operation (for details see ch. 4, sec. 4.4.5). Two instances that do not overlap in their coverage of the SemRep do not form any C2 link. Informally, if two instances overlap in their SemRep coverage, one of the constructions (child) needs to provide a SynForm that can (partially) fill in the missing form information of the other construction (parent). The core constraints are that the feature of the child construction needs to match both the syntactic constraints and semantic constraints carried by or linked to the slot of the parent construction. Going back to fig. 2.12, the MAN construction instance’s meaning-form mapping overlaps with that of the SVO construction instance since they both cover the MAN patient node of the SemRep. MAN can enter in cooperation with SVO (green link) through the first slot of the latter. Indeed, MAN is of class (N) as required by the slot, and the head node of the MAN construction (its “semantic class”), fits with the semantics of node symbolically linked to the first slot in SVO for the simple inclusion requirement ( $\text{MAN} \subseteq \text{ENTITY}$ ) (the same analysis holds for the cooperation between MAN construction instance and PAS\_SVO construction instance).

Each construction instance active in Grammatical WM carries a mapping hypothesis of a portion of the current semantic representation onto a linguistic form. Cooperation emerges between two constructions whose mapping can be composed to generate a new mapping covering a larger portion of the semantic content, or refining the mapping. Competition, on the other hand, is triggered when two constructions represent incompatible mapping hypotheses. (cf. WOMAN.1 and WOMAN.2, or SVO and PAS\_SVO in fig. 2.12).

C2 links are created incrementally: each time a new construction instance is invoked it is matched against the ones that are already active in the Grammatical WM (cf. fig. 2.16).

### **Cooperative Computation: Dynamics**

The construction schema instances invoked in Grammatical as relevant meaning-form mapping hypotheses (see sec. 2.5.2) enter in cooperative competition (see sec. 2.5.2). At each time step, the Grammatical WM contains a network of interacting construction instances consisting of cooperation and competition links (C2 network). The activation levels of construction instances are updated following a leaky-integrator dynamics.

The principles that govern the dynamics of the instances follows directly in the footsteps of most of the other cognitive modeling efforts based on hybrid dynamic-symbolic systems (McClelland, 1993) mentioned in Introduction (See ch. 4, sec. C.1.3 for details).

### **2.5.3 From Construction Assemblages to Utterances: Phonological WM and Utterance Production**

Through the process of competition and cooperation, construction instances generate construction assemblages, each representing a potential (possibly partial) self-organized program to translate the message (SemRep) into form content.

At each time step, a constructions assemblage corresponds to a network of cooperating construction instances. The hypothetical meaning-to-form mapping it represents carries its own activation value derived from that of the assemblage component instances and that reflects its relevance as a meaning-form mapping solution.

Looking back at Fig. 2.12 it appears that lexical constructions WOMAN\_1 and WOMAN\_2 compete as synonymous lexical constructions, with WOMAN\_1 winning. At the more abstract level of argument structure/voice: PAS\_SVO and SVO compete as they both build on top of the same portion of the SemRep but express the agent-patient semantic roles in different ways in their SynForm. PAS\_SVO is winning due its patient focus that is a better semantic match for the high activation patient MAN. Assuming that WOMAN\_2 loses the competition and is pruned out, we are left with two construction instance assemblages, corresponding respectively to the use of active and passive voice. If forced to choose, the system employs a winner-take-all strategy and the passive would win since PAS\_SVO has a higher activation value than SVO, yielding an assemblage with a higher activation.

When the system is required to generate an utterance, the winner assemblage is selected, the constructions instances are unified, and the form of the resulting meaning-form mapping is sent to the Phonological WM as the basis for generating the utterance. Constructions instances continuously receive external activation from the concept schema instances of the SemRep they cover. Such external activation, across WMs, injects, in the temporal dynamics of the activation values of the construction instances information about the relevance of the semantic content they map onto form content. When an assemblage is selected and used to generate a meaning-form mapping, all the elements of the SemRep it has expressed onto form content are mark as expressed and stop sending activation to the construction instances, reflecting the fact that they do not need to contribute anymore in the generation of utterance content. This ensures that the state of the grammatical WM adapts to the state of the message (i.e. what parts have been expressed, what parts remained to be mapped onto utterances).

Informally, going from a set of cooperating constructions (construction assemblage) to a sequence of phonological forms involves unifying the SynForms of the cooperating constructions by replacing, where a cooperation link exist, the slot of a parent construction by the SynForm of the child construction it is linked to. For a detailed account of the process refer to ch. 4.

The Phonological WM plays an important role as the system might be required, in order to continue the incrementally production of utterances, to take into account the phonological content of previous utterances.

#### 2.5.4 Linguistic WM: Hierarchy but no Tree

Fig. 2.13 gives an informal view of the state of the system, subsuming both Semantic and Grammatical WM into a unique, multi-layered, Linguistic WM. It assumes that the competitions have been carried out and that the losing construction instances have been pruned out (v.i.z WOMAN\_2 and SVO). The bottom layer corresponds to the Semantic WM with the SemRep state. On top of this layer, construction instances are applied, forming a hierarchical structure that correspond to the final Grammatical WM state. Lexical construction instances form the first layer of construction instances. From there more and more abstract construction instances pile up. More abstract construction instances build upon and organize the content of less abstract construction instances that in turn specify information requested by the more abstract instances. The text associated with each instance corresponds to the utterance that would be produce, would the system only read-out the WM content at this level.

The lexical construction instances express the content of the SemRep nodes, while the argument structure construction maps onto the lexical constructions and bundles them into a single SynFrom. This hierarchical view of the processes aligns with the design of VISIONS in which the Visual WM was composed of various WM layers, each handling perceptual schemas of increasing scope and abstraction.

#### 2.5.5 Good Enough Production of Utterances: Speaker and Task Relevant Parameters

Focusing on language use, to fully understand the nature of the vision-language interactions requires embracing production performance as it is rather than sterilizing it, discounting the utterances that are not well-form and fluently generated, or limiting the production to predefined sentence templates (e.g. single active clause, conjoined subject NPs, etc.). Much work on language comprehension has by now outlined the necessity to understand the comprehension process as solving a satisficing problem (Simon, 1972) finding an interpretation for an utterance that is good-enough to solve the problem at hand while satisfying the

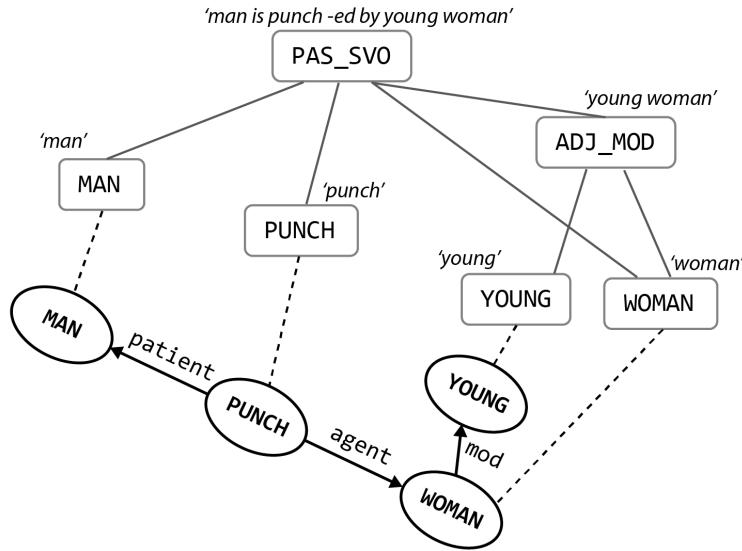


Figure 2.13: The weakly coupled Semantic and Grammatical WM taken together can be considered as forming Linguistic WM. This figure informally summarizes what could be an end-state of the situation described in fig. 2.16. The lexical construction instances express the content of the SemRep nodes, while the argument structure construction maps onto the lexical constructions and bundles them into a single SynFrom. This hierarchical view of the processes aligns with the design of VISIONS in which the Visual WM was composed of various WM layers, each handling perceptual schemas of increasing scope and abstraction. Dashed lines indicate cross-WM activation links through which the concept instances of the SemRep can impact the activation of the construction instances that map them onto meaning (for simplicity, links between concept relation instances and construction instances are not shown). Above each construction instance, the text corresponds to the utterance that would be generated by the construction assemblage that would select this instance as the top instance.

constraints defined by the task as well as by the system itself. To this “good-enough comprehension” principle to comprehension (Ferreira and Patson, 2007; Ferreira, 2003; Christianson et al., 2001; Patson et al., 2009) we propose that should be added a “*good-enough production*” principle: the output of the language production system corresponds to a good-enough solution to a given task. Whether or not fluency and well-formedness are the overarching constraints depends on the task at hand. SALVIA accounts for the fact that the processes can function at various regimes and can be impacted by task-related requirements.

Parameters defines the characteristics of the dynamics taking place both within and between WMs. In doing so, the core temporal behavior of the model with respect to incrementally received inputs is set.

The main parameters of the system are those that define the dynamics of each WM (in particular their relative characteristic times). They set the core temporal behavior of the model with respect to incrementally received inputs is set.

We propose that those be supplemented by another parameter reflecting constraints on the GrammaticalWM dynamics. To simulate the impact of (cognitive) time pressure on utterance production,  $t_{time\_pressure}$  constrains the model to attempt the production of an utterance at each  $\Delta_T = t_{time\_pressure}$  intervals. Crucially, the system has to do so whether or not all the required semantic information has been gathered, and also whether or not the state of the GrammaticalWM has converged to a unique solution (no more competition).

## 2.6 Language-Vision Interactions: Visual Guidance vs. Verbal Guidance

Attentional parsing of the visual scene is piloted bottom-up by the interaction of the bottom-up saliency values associated with the subscenes and the successive attentional windows (focus areas, which can restrict the set of subscenes that compete in the bottom-up attentional process). A key aspect of the Subscene Recognition system lies in the possibility to orient visual attention based on top-down requests received from VisualWM.

As mentioned above, the perceptual schemas instances active in VisualWM function also as pointers to the area they (partially) interpret. If more information is needed regarding a given perceptual instance, the VisualWM can send a request to the Subscene Recognition schema to change its current focus toward the area covered by the instance in question, triggering the retrieval of all the information that is spatially located within this area.

The interactions that take place between bottom-up and top-down signals orienting visual attention are informally illustrated by fig. 2.14. As shown on the left panel, eye-position and focus-area (spatial attention window) can be oriented by bottom-up by saliency characteristics of the scene (here towards the center of the scene where the action takes place) and top-down signals. This defines the area of the scene currently under attentional focus.

The perceptual schemas associated with this area, forming a subscene, are instantiated in Visual WM. In the case depicted, two entity perceptual schemas are instantiated, MAN and HUMAN, the latter reflecting the lack of information regarding the perceptual entity it denotes, one action schema (PUNCH) that links the two participants (represented as edges). Such high level perceptual schema instances are used as stand-ins for the highest level of perceptual representation achieved in a model such as VISIONS. It assumes that many more perceptual schemas have been put to work in order to be able to generate such perceptual representation, while also making the hypothesis that only the relevant perceptual schemas will be used to support the creation of the semantic content.

Those perceptual instances active in Visual WM are then conceptualized to generate the SemRep, state of the Semantic WM. SALVIA does not tackle the conceptualization process per se and therefore simply assumes many-to-one mapping between perceptual schemas and concept schemas (e.g. many types of perceptual relations denoting an actor here PUNCH will be mapped onto the concept schema PUNCH and the conceptual relation of agent (agt) and patient (pt)). It is important to note that, as in VISIONS, the invocation of schemas instances in semantic WM on the basis of the content of the Visual WM results in cross-WM links between the instances.

As shown in Fig.2.14, those links can serve as back pointers allowing the passing of top-down request to the visuo-attentional system from the Semantic WM. In this case, in order to request more information

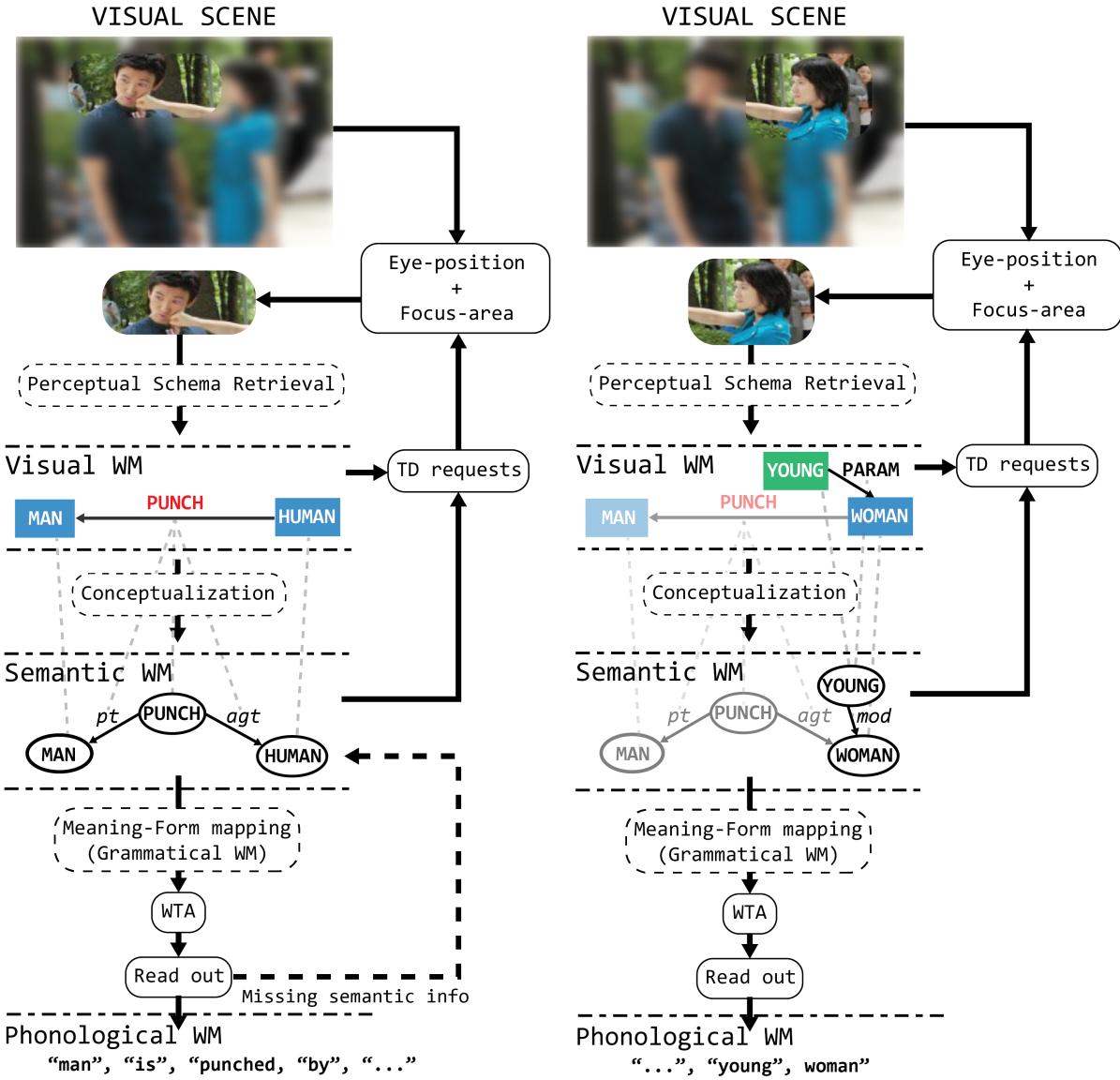


Figure 2.14: Interaction between visual-attentional processes and language production. (Left) Eye-position and Focus-area defining the current attention window result from bottom-up saliency signals. Through local saccades, the visual sub-scene associated with this scene region is extracted and the corresponding perceptual schemas are instantiated, updating the state of the Visual WM. Following their conceptualization, they update the state of the Semantic WM (SemRep), expanding the content of message to be verbally conveyed. In this case we assume that the WMs were initially empty and therefore the final semantic representation correspond to the sub-scene perceptual information that has been extracted in this bottom-up process (a man is punched by an unknown human). Top-down signals can emanate from the Semantic WM or the Visual WM to orient the attention window toward an area of the scene where information relevant to the current process can be found. Verbal guidance results from such top-down attentional requests. As the system has already committed to the utterance “man is punched by” but is lacking specific semantic information about the agent of the action (only specified as an HUMAN), a top-down signal can be sent that will orient the visual attention towards the area of the scene that is linked to the HUMAN concept instance (through the intermediary of the associated perceptual instance). (Right) The attention focus-area is now directed at the HUMAN (PUNCHER). This allows the SemRep to be updated on the basis of novel perceptual information (HUMAN is specified to be a WOMAN, who in addition is YOUNG). Filling in the missing semantic information allows the system to continue the utterance smoothly using the PASSIVE pattern it had chosen.

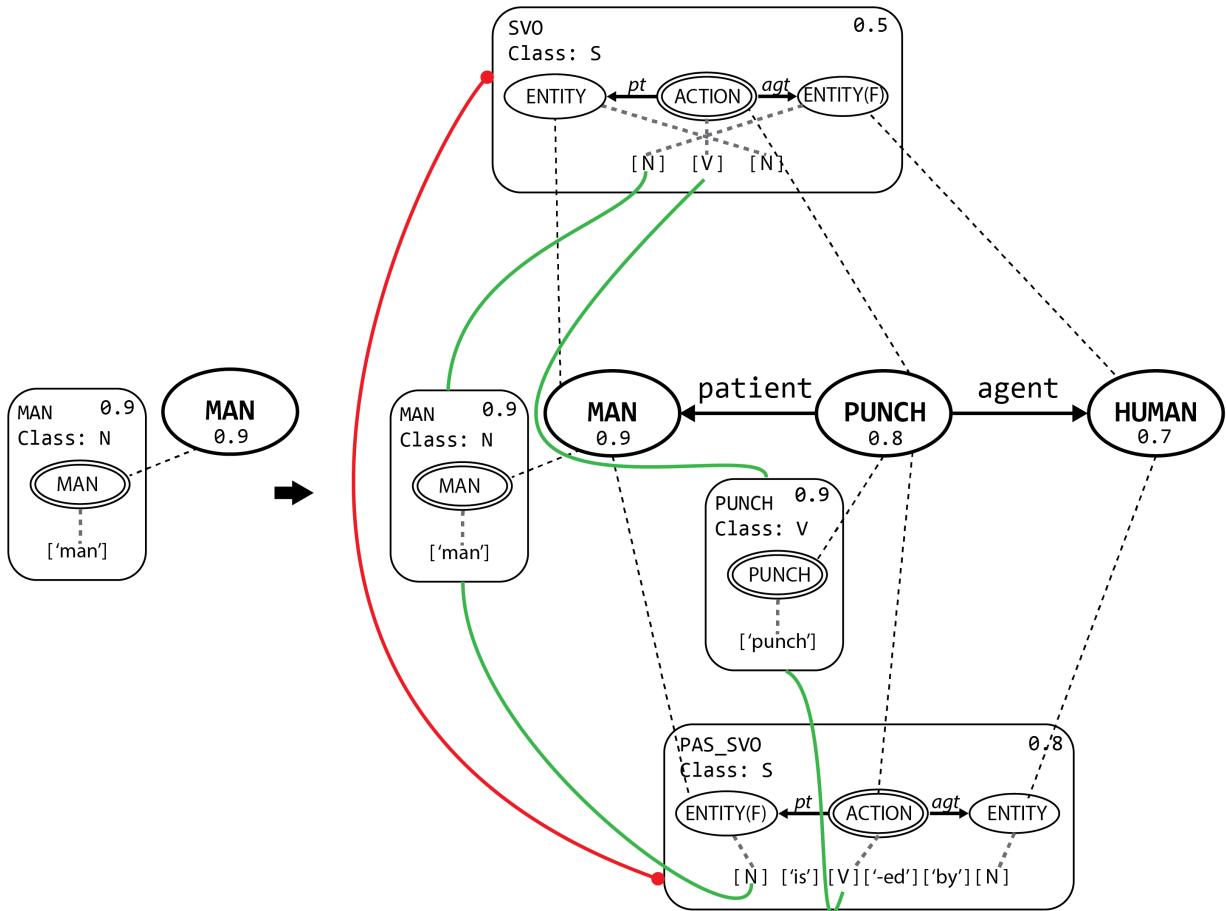


Figure 2.15: Generating a flexible grammatical structure mapping on top of an incrementally built semantic representation (SemRep). (Left) The SemRep only consists of a MAN node covered by a single MAN construction instance. (Right) Semantic content has been added to the previous SemRep. The MAN construction remains while other construction instances have been invoked, adapting the grammatical structure to the new SemRep state. fig. 2.12 presents yet a following state in which the agent HUMAN has been specified as WOMAN.

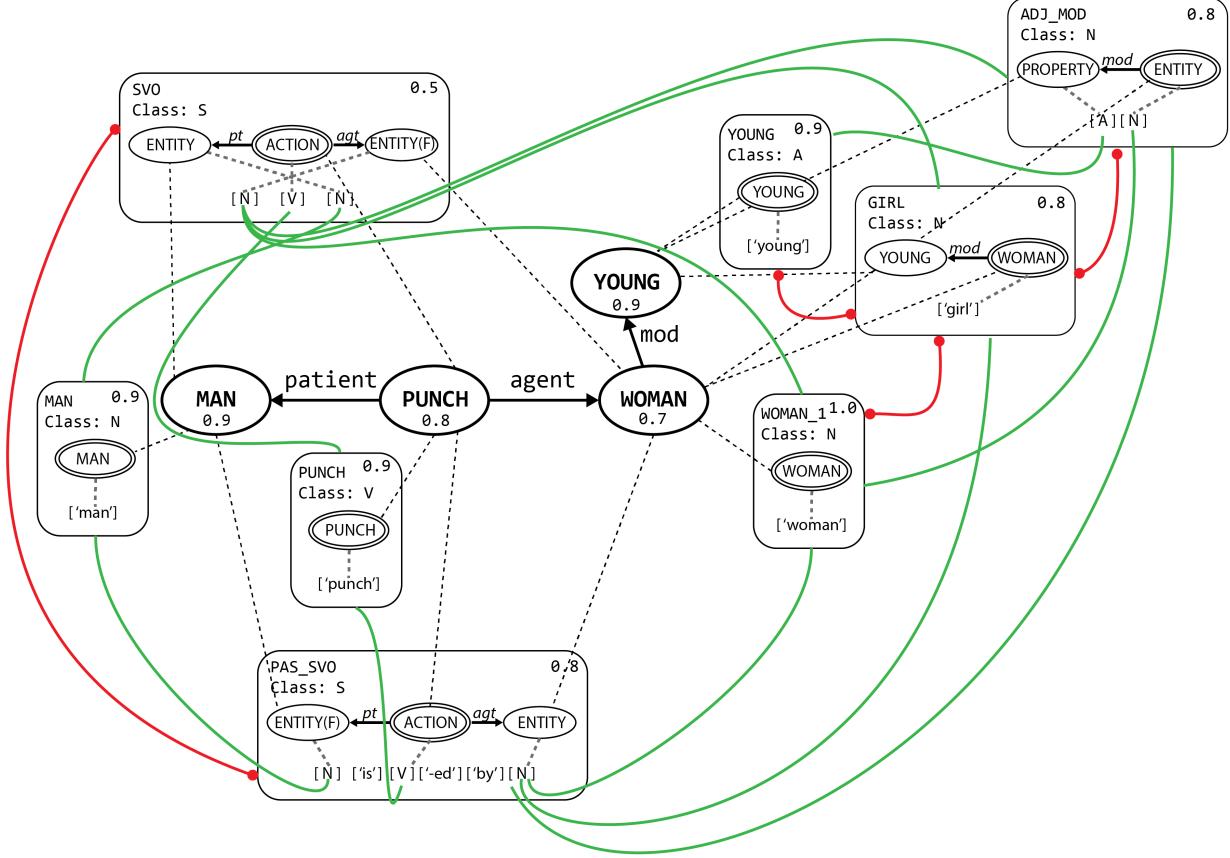


Figure 2.16: Example of state that can follow the one described in Fig. 2.12 following novel attentionally retrieved perceptual information. Additional semantic information is added to the WOMAN concept instance node: a property (YOUNG) that modifies WOMAN. On the other hand, the WOMAN\_2 construction instance has now lost the competition against WOMAN\_1 and has been pruned out of the Grammatical WM. The addition of the new semantic content resulted in incremental invocation through SemMatch of both the lexical YOUNG construction instance and of the ADJ\_MOD construction instance, following by the creation of the new C2 links between those new constructions instances and the existing ones through Match. The C2 network changes naturally followed the incremental changes that took place at the SemRep level.

regarding the nature of the agent HUMAN, the Semantic WM can orient the visual attention TD towards the relevant region of the scene through the intermediary of the perceptual schema instance it is linked to, instance that itself contains information regarding the spatial area of the scene that it covers as a perceptual hypothesis. This allows the combination of both bottom-up and top-down attentional cues in the process of visual parsing a visual scene.

Fig. 2.16 therefore represents the states of the Semantic and Grammatical WM that follows the one described in fig. 2.12. The interplay of bottom-up and top-down visuo-attentional signals resulted in the updated of those states, allowing for a smooth continuation of the utterance.

Given a SemRep, construction instances enter in cooperative computation in Grammatical WM until a winning construction assemblage is chosen as the basis to generate the meaning-form mapping that will result in a description utterance.

Following a Winner-Take-ALL (WTA) and a ReadOut process that combines the construction instances SynForms in order to generate a phonological sequence in Phonological WM, the system commits to uttering "man is kicked by". However, semantic information regarding the nature of the agent is missing. This triggers a Top-Down signal requesting information about the HUMAN SemRep node that for now represent

the patient. This type of utterance-driven TD attention orientation represents a form of “verbal guidance” which, in scene description generation, co-occurs with the BU driven “perceptual guidance”, as the driving force underlying the relations between saccade and word sequences.

Note that the assumption is that more information about a given perceptual hypothesis can be found in the vicinity of the area it interprets. This is obviously a strong simplification since in many cases, relevant extra-information related to a given perceptual schema might not be found in spatial proximity (e.g. getting the information about what a person looking at requires much more complex spatial reasoning).

## 2.7 Language Production Schema System: Modeling Language Use

### 2.7.1 Incrementality at its Core

Fig. 2.16 presents a conceptual view of a continuation of the states of both the Semantic and Grammatical WM, compared to the state previously described in fig. 2.12, once the visuo-attentional system, triggered by verbal guidance, led to the perceptual and from there semantic information regarding the agent to supplement the SemRep. In doing so, it triggered the instantiation of new construction schemas (YOUNG and ADJ\_MOD, and GIRL).

It is worth noting how the GIRL construction instance enters in competition with a whole noun phrase mapping the same conceptual content onto “young woman”

Fig. 2.17 presents a possible final assemblage  $A$  following the C2 process in Grammatical WM as described in figs 2.15, 2.12, and 2.16.

All the competitions have yielded their results, and only cooperation links remains. The system is not guaranteed to converge, or might not have converged when the system is required to make a decision and start an utterance in which case competition are terminated with the winning assemblage being selected. Each cooperation link can be seen as a unification link between two construction instances. Figure 2.18 illustrate some of the resulting stages of performing all the unifications, yielding ultimately a single construction instance  $eq\_inst_A$  that contains the sum of all the information carried by the construction instances in the assemblage<sup>5</sup>.

Generating a form from a construction instance assemblage rests on the possibility to unify the cooperating construction instances. Given the state reached in fig. 2.17, fig. 2.18 provides an informal example of the unification process at play.

Ch. 4 presents in detail the TCG schema theoretic model of language production.

### 2.7.2 Seeing-for-Saying

The previous section on “Good-Enough Production” already presented the way the SALVIA system of language production incorporates task and speaker related parameters. Pressure to produce can be applied to the system reflecting situations in which, for example, a person is producing language extemporaneously. Other speaker related parameters can be incorporated as factors defining the criteria used to define the scores of an assemblage (favoring utterance continuity, seeking brevity and semantically compact utterances, etc). Those can also be defined through constraints imposed on WM parameters as well as on constraints imposed on them (limiting for example the complexity of the structures they can hold) (cf ch. 4)

The production system does not seek to produce well formed sentences but utterances that fulfill the communicative goal of the system given the constraints it faces and inherent preferences.

These constraints imposed on the language processes are critical to understand and simulate the interactions between language and vision. Without them, the system would be essentially in an idealized situation of simulated competence instead of performance, situation artificially created that abstract away precisely the effect that SALVIA is interested in modeling: those emerging from situated language use.

---

<sup>5</sup>PAS-SVO serves as a much simplified shortcut for a complex process that should involve at least argument-structure (transitive) and voice (passive) constructions.

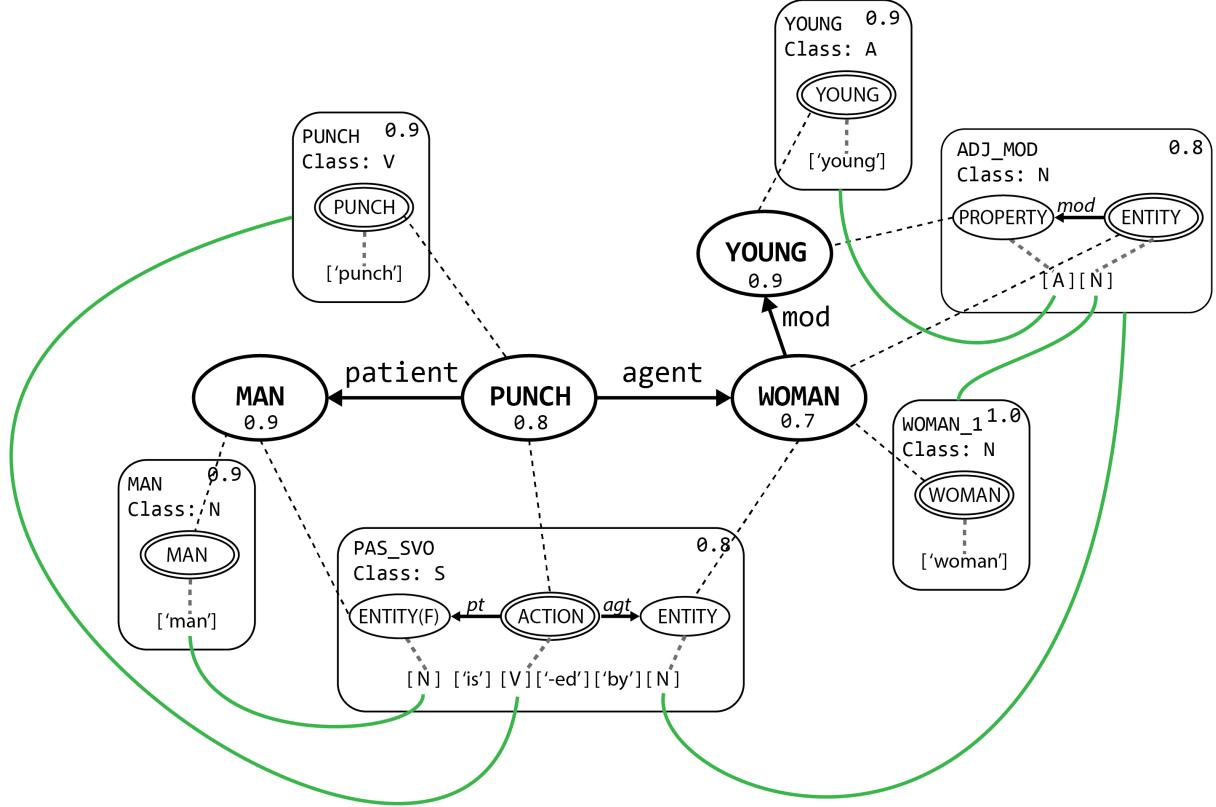


Figure 2.17: Possible winning assemblage on which the C2 process in Grammatical WM could converge (or competitions could be prematurely terminated using winter-take-all if the system is required to make a choice before convergence). In this example, the state of the Grammatical WM described in fig. 2.16 yielded a winning assemblage in which the PAS\_SVO construction instance won the competition against SVO resulting in the latter being pruned out of the WM. In this assemblage, each construction instance contribute to mapping a part of the SemRep onto a linguistic form. The 4 lexical construction instances (MAN, PUNCH, YOUNG, WOMAN) map nodes onto lexical items. The ADJ\_MOD construction instance translates the ‘mod’ semantic relation into sequential constraints on form (‘adj noun’ pattern). The PAS\_SVO construction instance builds on top of all those previous instances and translates the agent-action-patient frame into both sequential form order and required function words. How those construction instances combine their hypotheses to generate a linguistic form is illustrated in fig. 2.18.

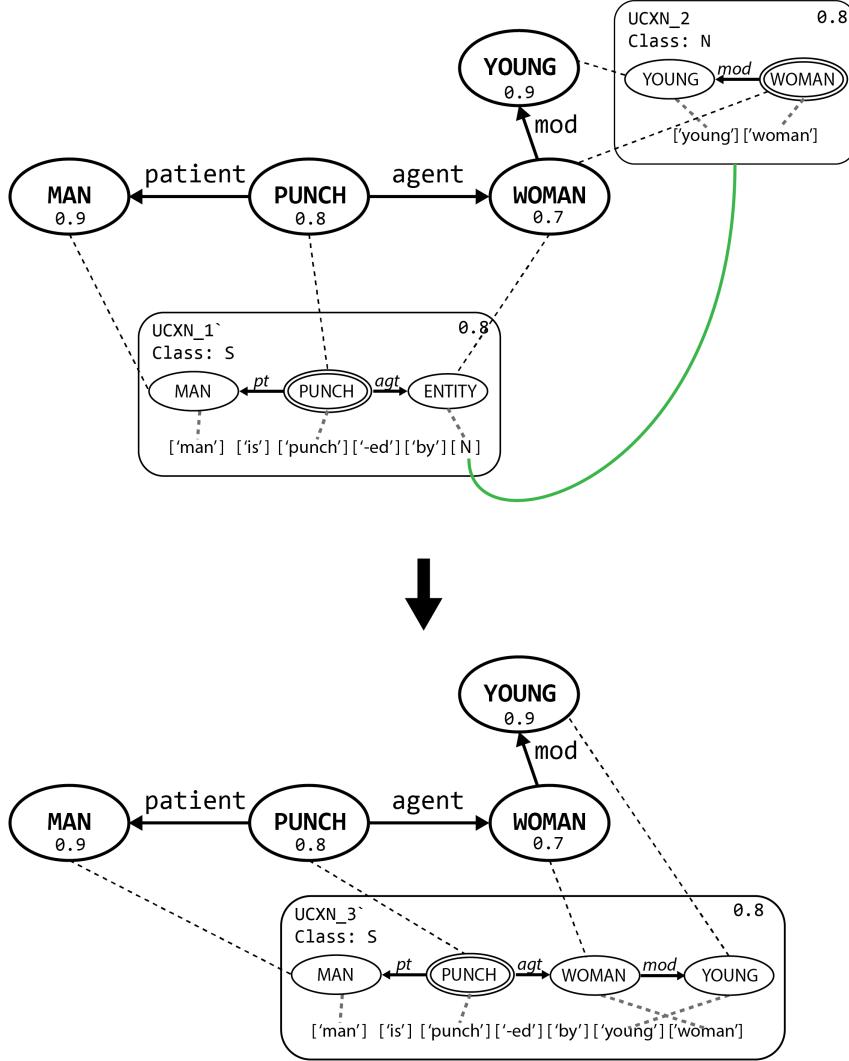


Figure 2.18: Construction instances unification example. Two stages of the unification process generate the equivalent construction instance for the assemblage  $A$  shown in fig. 2.17. (Top) Stage following the unification of all the lexical construction instances with the more abstract constructions they cooperate with. UCXN\_1 results from the unification of PAS\_SVO with both MAN and PUNCH instances. The subject and verb slot are lexicalized but the adjunct form is missing. UCXN\_2 results from the unification of ADJ\_MOD both with YOUNG and WOMAN instances. It has a fully lexicalized form ('young woman'). (Bottom) Stage following the unification of UCXN\_1 and UCXN\_2 resulting in UCXN\_3 =  $eq\_inst_A$ . All the construction instances forming the assemblage have been unified. The construction instance equivalent to the assemblage maps the entire SemRep onto a fully lexicalized form that can serve as an utterance output. Note: This figure shows an example of unification stages. However, to generate the final equivalent instance, the order of unification is irrelevant.

Those are the good-enough production constraints that result in the language production system interacting top-down with the attention system in order to optimize the visual parsing of the scene with respect to the state of the unfolding utterance.

Limitations imposed on the language processes in turn constrain and **shape** the dynamics of interactions between language and vision.

## 2.8 Preliminary Conclusions

This chapter outlined a novel computational cognitive model of language production in the context of the description of visual scene. The Schema Architecture Language-Vision InterAction cognitive model (SALVIA) provides an implementation that takes into consideration lessons derived from cognitive modeling theories, from conceptual theories of language production, from the known theories of visual attention, as well as from recent advances in cognitive linguistics and in particular in computational construction grammar.

The next chapter puts the model to work and provides simulations results that both validate its validity as cognitive model of language production (of visual scene description and of language in general), but also show how SALVIA can simulate and provide a novel computational interpretation of key psycholinguistic results.

Chapter 4 is then more specifically dedicated to the presentation in details of Template Construction Grammar as a computational construction grammar model of language production.

# Chapter 3

## From Gaze Patterns to Utterances: Simulating the Dynamics of Visual Scene Description

*“It seems perfectly reasonable to think that much, if not all, that is universal in human language is attributable to underlying cognitive structures and processes. Perceptual and linguistic sequences must, at some level, share a common representational (semantic) system and a common set of organizational (syntactic) rules, cognitive in nature.”*

Osgood

Where do sentences come from?

### 3.1 Introduction

This chapter presents a series of simulation results derived from the SALVIA model of language production. After having first provided computational counterpart to the conceptual examples used in ch. 2, showing in the process how SALVIA can handle the production of more complex messages than the ones discussed so far, the full model is used to simulate the process of scene description linking eye-movements and the verbal production of utterance. With this in place, the last section turn to the simulations of key psycholinguistic results and show how SALVIA offers an implementation supporting the analyses of Kuchinsky 2009. It shows that the view of a general flow of information from the visual processes to the grammatical processes with the saccadic sequence driving the linguistic sequence (visual guidance), only holds in certain experimental conditions. Other conditions reveal a feedback effect of the linguistic processes onto the visuo-attentional dynamics: having committed to a grammatical structure for an ongoing utterance, but lacking the necessary semantic information to fill in all the required role, saccades can be triggered towards the parts of the visual scene likely to contain the missing information (verbal guidance).

### 3.2 Input-Outputs

#### 3.2.1 Inputs

##### Semantic Inputs

The visual system of SALVIA can be bypassed by directly defining as input an incremental sequence of semantic information. As the model runs, the semantic information is retrieved at the pre-defined time, and the SemRep state of the Semantic WM is updated accordingly.

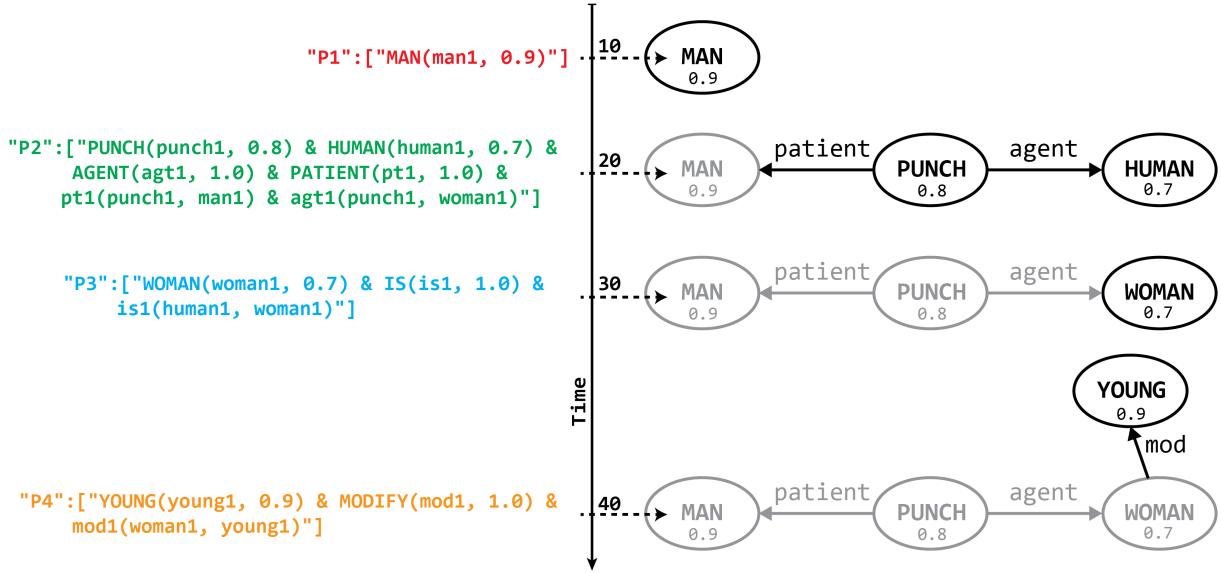


Figure 3.1: . A SemRep incremental input. This type of input bypasses the scene recognition system and the visual WM, directly providing to the system the incremental semantic content it should process as a basis to generate utterances (simulating only the incrementality inherent to the gathering of semantic content during visual scene parsing). (Left): The inputs are defined as simple propositions, each defining either concepts or relations. A value stipulate the rate at which the propositions should be interpreted and used to update the state of the SemanticWM. (Right) State of the SemanticWM updated each time new semantic content is received as input. The SemRep is built incrementally (at each time step, novel semantic content is shown is highlighted).

SALVIA can be run by providing directly an incremental semantic input to the SemanticWM. This allows to bypass the visuo-attentional system while retaining the key feature of the dynamic, incremental nature of the process of building the message to be linguistically conveyed. Figure 3.1 illustrates such an input (for a more in depth description, refer to (Barres, 2017)). The detail regarding how the incremental semantic input are defined is provided in Appendix C, Sec. C.1.5.

## Scenes

SALVIA can take as input visual scenes. Such scenes are defined as hierarchical network of subscenes each containing manually defined perceptual schemas associated with a spatial regions of the input (see also fig. 2.10). Subscenes and perceptual instances are directly associated with image areas and all are associated with a bottom-up saliency value

As the model runs, based on the state of the attention process (location, focus scale), the perceptual schemas that are under attentional focus are instantiated in the Visual WM, incrementally updating its state.

The example given here, compared to the one showed in fig. 2.10, illustrates the type of visual scenes used in psycholinguistic studies to analyze the relations between attentional and linguistic processes.

Those artificial scenes could be term “**minimal scenes**”. They contain the bare minimum amount of perceptual content to create a scene that is not merely a collection of objects. Such minimal scenes, if they are still useful for modeling purposes, limit greatly the amount of processing require to build a scene representation. Compare to a natural scene that can be conceptualize in a very large amount of ways, such scenes are very limited in the possibility of conceptualization they offer: all the entity and character are stereotypes and they only contain one main event. In many ways, they can be seen as pre-conceptualized scenes: the viewer is already given a conceptualized view of what a real scene would be. This justifies in part why in the present model, the conceptualization sub-system was kept simple. For such a minimal scene,

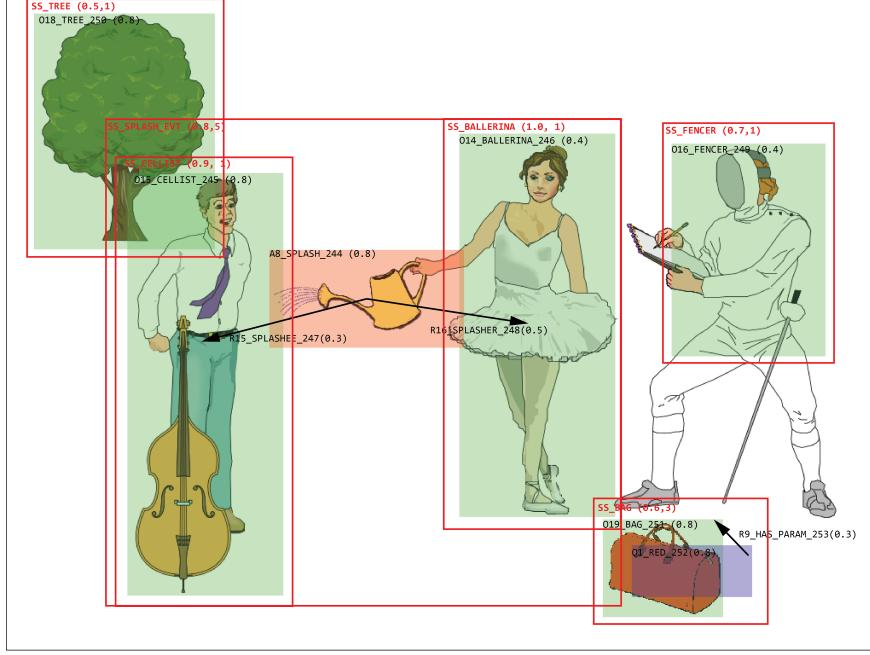


Figure 3.2: Example of input scene. Scene taken from (Knoferle and Crocker, 2006). This simple input was generated using the SceneBuilder (cf. Appendix B, sec. B.5). Arrows stand for perceptual relation schema instances while black boxes denote ENTITY, ACTION, or PROPERTY perceptual schema instances and the regions of the visual scene they cover. Red boxes denote subscenes, and the region of the visual scene they cover. Each element is associated with a saliency value.

the incrementally built scene representation (SceneRep) can be seen as very close to the SemRep that will conceptually express the scene content.

This justifies why, in many cases, when the spatial aspect of the regions covered by the perceptual schema instances does not play an important role, the scene representation will be simplified to bypass the actual spatial linkages to only keep the structural relations between subscenes (see below).

Similarly, it justifies why, in the case when the feedback from language processes onto visual attention are not considered, the incrementally received conceptualized perceptual representation, resulting from attentional parsing, can be simulated by a direct incremental SemRep input (see above).

Since this format is costly to generate and run, we will generally rely on a version of the scene format that abstract away from spatial anchoring while keeping the requirement that scenes are represented as hierarchical sub-scene structure, particularly adequate for the “minimal scenes” used in psycholinguistics experiments.

Scene inputs can bypass the perceptual schemas and be defined directly in terms of the semantic content the perception of each subscene will instantiate in SemanticWM. The input is then given in the same format as the one used for SemRep input that was described above. However, instead of stipulating the order in which the semantic content should be retrieved, a scene structure is stipulated. The scene structure is defined as a hierarchy of subscenes. Figure 3.3 provides an informal illustration of such input formats, using the case of a scene used by Kuchinsky. Here the semantic contents have been pre-assigned to subscenes (bypass conceptualization). The semantic information associated with a subscene is retrieved when it becomes the attentional focus (BU or TD directed).

Perceiving the scene is therefore based both on the bottom-up saliency of each subscenes but also on the how the subscenes are hypothesized to structurally organize into a complex cognitive scene representaiton. Cueing a subscene as well as top-down requests from the language system can directly focus the attention on a sub-part of the scene structure.

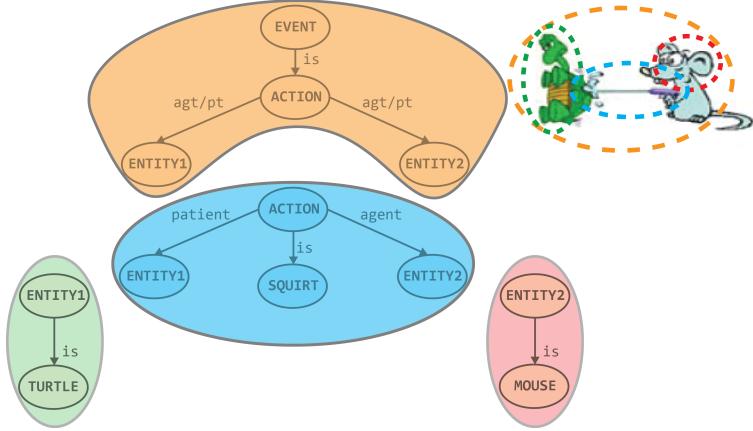


Figure 3.3: Illustration of a scene input. (Left) Subscenes and the semantic content that is retrieved when they fall in attentional focus. The organization from top to bottom reflects the hypothesized hierarchy between the subscenes in terms of how easily they are built (easiest at the top). In this case, the general and rather abstract event subscene is hypothesized to be most easily retrieved from the input (in absence of other factors such as top-down requests). Conversely, the subscenes perceptually defining the entities participating in the action-based event are assumed to be, relatively, harder to build. (Example of a scene with Easy event, Hard objects) (Right) The image regions they are linked to. (scene adapted from (Kuchinsky, 2009))

### 3.2.2 Outputs

Each subsystem that is part of the SALVIA model can be probed and generate outputs. With respect to language generation, the two main outputs are the utterance generated and the scene parse trajectory (in the case in which the scene input is provided).

**Utterances** are the first output of the model. The utterances generated by the model are defined as time stamped sequences of words (and occasionally bound morphemes).

**Fixation & Focus** If a full scene is provided, the scene parse trajectory consists in a time stamped series of gaze positions and focus attention window sizes (see fig. 3.4 for a simple example.)

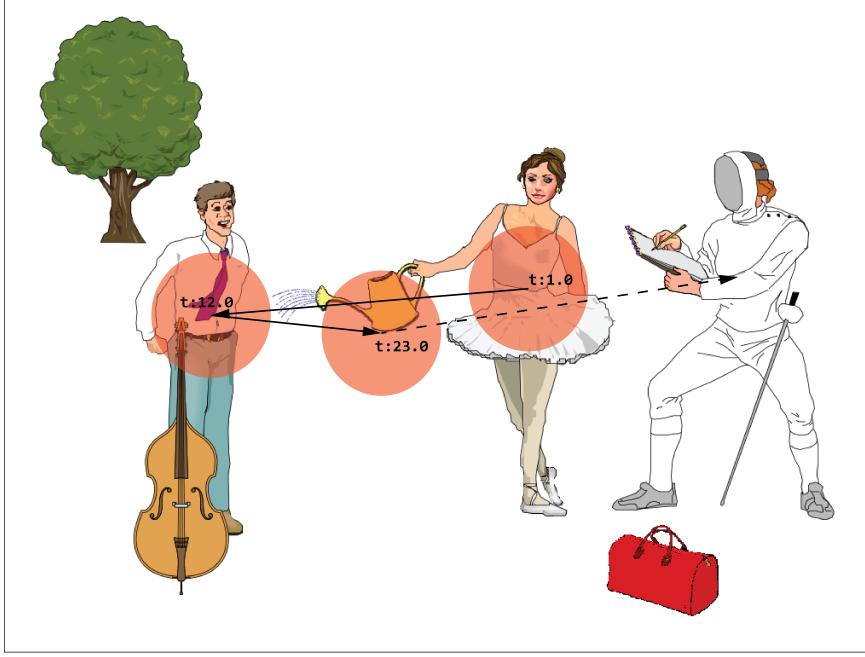


Figure 3.4: Example of saccade output (focus attention window not shown).

If a simplified scene input is used, the scene parse simply consists in the time stamped order in which the subscenes are attended to alongside the attention window defined in terms of restrictions of the attention within the scene structure graph.

### 3.3 Parameter Space

On the basis of Schema Theory, each WM in the system is associated with its own set of parameters. However, since the schema instances in Visual WM, Semantic WM, and Phonological WM do not form cooperation or competition links, their dynamics is much simpler to express and only depends on external inputs (cf. Appendix B, sec. B.1 for more details.)

**Parameter Space Problem in a Hybrid System** The question of the parameter space on which the SALVIA model can be instantiated is of course of great importance as one of the key questions the model has to address is whether or not one can link different areas of the parameter space with different types of observed linguistic behaviors. Hybrid systems such as the one presented here, because they combine analogical and symbolic representations in their operationalization, incorporate a challenge that is harder to directly quantify, namely the accuracy of the symbolic content incorporated in the model. In the case of the SALVIA model, the perceptual, semantic and grammatical knowledge are not learned but are given as input to the model. Clearly such knowledge contents cannot be analyzed in terms of the parameter space they span and remains in the realm of knowledge design until such time that the model will be extended to incorporate learning at all the levels that include symbolic representations. For this reason, it is crucial to understand that the present model exists as bridge between the conceptual work of experimentalists (including linguists that have outlined the existence of various forms of linguistic knowledge) and modelers whose role is to start putting those concepts to work.

## 3.4 From Incremental Semantic Representation to Utterances

### 3.4.1 Sanity Check

The first line of test for the model consisted in ensuring its capacity to properly generate linguistic outputs for a given semantic input in the absence of any constraints (competence test). This was carried by running the model on a series of SemRep inputs in which the semantic input was provided all at once or in the equivalent situation of Time Pressure >> Last input received time >> Tau\_GramWM.

For each of the input SemReps, that ranged in complexity from single entity (and modifiers), to single event, all the way to multiple events, was attached a set of ground truth sentences. Testing the model could therefore be compared to a translation testing. BLEU score was used and the model was able to score 100% on bigram BLEU score. It should be noted that this is a fairly lenient benchmark due to the inherent limitation of the TCG grammar used and of the BLEU score itself. This result should only be taken as a sanity check indicating that the model is able, from a pure competence perspective, to deliver what it was designed to deliver.

### 3.4.2 From Conceptual to Computational Example

The model received as input the same simple succession of semantic states as the one used in the conceptual examples used in this paper. First, no time pressure is applied.

The succession of states of the LinguisticWM (SemanticWM + GrammaticalWM) is given in Appendix B, sec. B.4.1

The model outputs: *[START](602)man is punch -ed by girl[END]* (time of utterance is indicated in parenthesis).

The temporal profiles of the construction instances' activation values are shown in Figure 3.5.

The dashed red line indicates the initial activation values of instantiated constructions (here there is no modulation of initial activation so all construction instances start with the same activation value). The MAN lexical construction is the first to be invoked in GrammaticalWM. Its activity builds up, driven by the activation it receives from the SemRep subgraph it maps onto. As the information about the action event is received, the PUNCH lexical construction gets activated while competition starts between the SVO and PAS\_SVO construction instance. Just before t=200.0, PAS\_SVO emerges as a winner.

When the semantic information about the woman agent is received, the two synonymous WOMAN lexical constructions are invoked and enter in competition. Meanwhile, as cooperation builds up between the lexical constructions and the PAS\_SVO construction, the latter gets an extra boost of activation and emerges as the structure that organizes the grammatical mapping. At around t=400, the symmetry between WOMAN instances breaks. The bottom dashed line indicates the value under which instances are pruned out of GrammaticalWM. At around t=600, both SVO and the loser WOMAN construction have been pruned. There is no more competition in the network.

A single assemblage remains. It is used to map the full SemRep onto the output utterance mentioned above. Following this step, the construction instances stop receiving activation from the SemRep instances that have been expressed (all of them in the present case) and therefore their activities start to decay; they will all eventually be pruned. However, if new semantic information were provided before the pruning occurs, i.e. during a time window proportional to the time characteristics of the GrammaticalWM, the old grammatical structures would still be available to cooperate with the new structures, influencing the continuity between utterances.

To illustrate this point, the model was then run with *time\_pressure* set at 200 (forcing the system to attempt to produce an utterance every 200 steps).

It outputs: *[START](208)man is punch -ed by (402) woman[END]*.

Here the system first produces a partial utterance at t=208. PAS\_SVO construction wins as it enables the language system to start expressing the semantic content, even though the information about the agent is not yet available. The system then pauses. When the nature of the agent becomes available, the grammatical processes, piloted by the PAS\_SVO instance, can smoothly incorporate the newly invoked lexical constructions and generate a single word utterance at t=400 that finishes the passive structure.

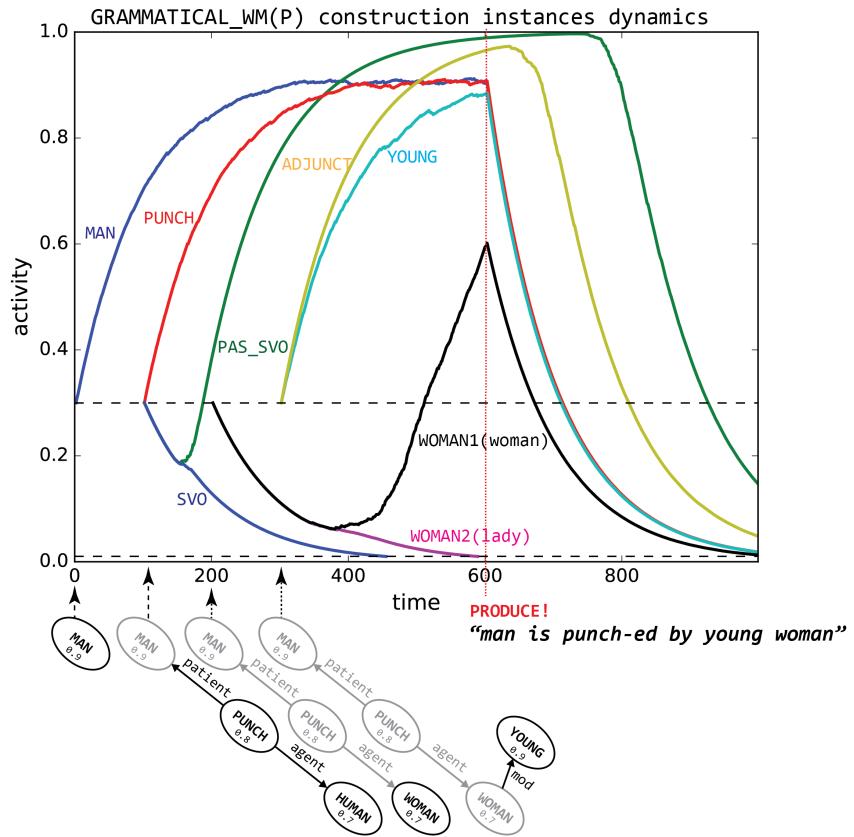


Figure 3.5: Construction instances activations for a simulation of language production based on the same succession of SemanticWM states as the one used as example in this paper. The bottom part of the figure indicates the state of the SemanticWM (semantic state update rate = 100). The new semantic content is highlighted.. The decay in activations that start at  $t=600$  corresponds to the fact that the production of the utterance “man is punch-ed by young woman” is triggered. (For a video see [victorbarres.github.io/media/SALVIA\\_p/states.mp4](https://victorbarres.github.io/media/SALVIA_p/states.mp4))

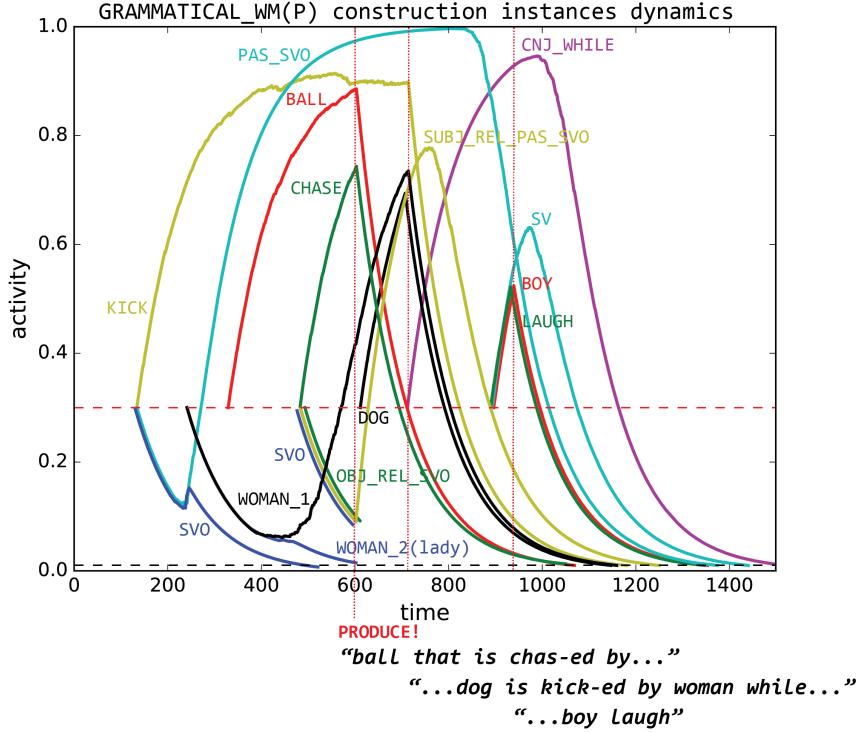


Figure 3.6: Construction instances activations for a simulation of language production based on the input described in sec. B.3, tab. B.6. Production required to start at  $t=600$ . (For a video see [victorbarres.github.io/media/SALVIA\\_p\\_complex/inst\\_activity.avi](http://victorbarres.github.io/media/SALVIA_p_complex/inst_activity.avi))

### 3.4.3 Increasing the Complexity of the Message: Impact of Time Pressure

To illustrate the behavior of the model on a more complex incremental semantic input and under time pressure, SALVIA was ran using the following input:

Fig 3.7 provides a snapshot of the (weakly interconnected) states of the Semantic and Grammatical WM, referred to as Linguistic WM. Displaying those states becomes quickly a challenge, however, it is worth noting that the model is able to provide seamless coordination between the two WMs. In the present case, competition remains in Grammatical WM (red links). The whole of the Grammatical WM state however, covers the entirety of the SemRep and could therefore be used, if required, to generate a meaning-form mapping and trigger the start an utterance.

The system is required to attempt to produce every 100 steps starting at  $t=600$ . The temporal profiles of the construction instances' activation values are shown in Figure 3.6. At  $t=601$ , SALVIA generates the utterance “ball that is chased by”. It has not yet received the semantic information regarding the nature of the agent of the chase action, but has already committed to the use of the passive voice (PAS\_SVO) construction. At  $t=712$ , SALVIA generates a smooth continuation for the preceding fragment of utterance “dog is kick-ed by woman while”. SALVIA had received the information that DOG is the agent of CHASE. The information about the KICK event and its participants (BALL, WOMAN) were already known but the choice of initial utterance delayed the expression of the action and the agent. At this point information about a concurrent event has already entered the SemRep and is captured by the conjunction “while”. Finally, at  $t=893$ , the system has gathered the semantic information regarding the nature of the concurrent event and once again smoothly continues the previous utterance by producing “boy laugh”.

The system is under time pressure to produce an output and therefore start producing utterances before the entirety of the semantic content has been retrieved. The utterance output is shown in tab. 3.1. Note that compared to the schedule of inputs shown in sec. B.3, tab. B.6, the two first utterances are generated before the last semantic input is received that stipulates the intransitive action performed by the BOY.

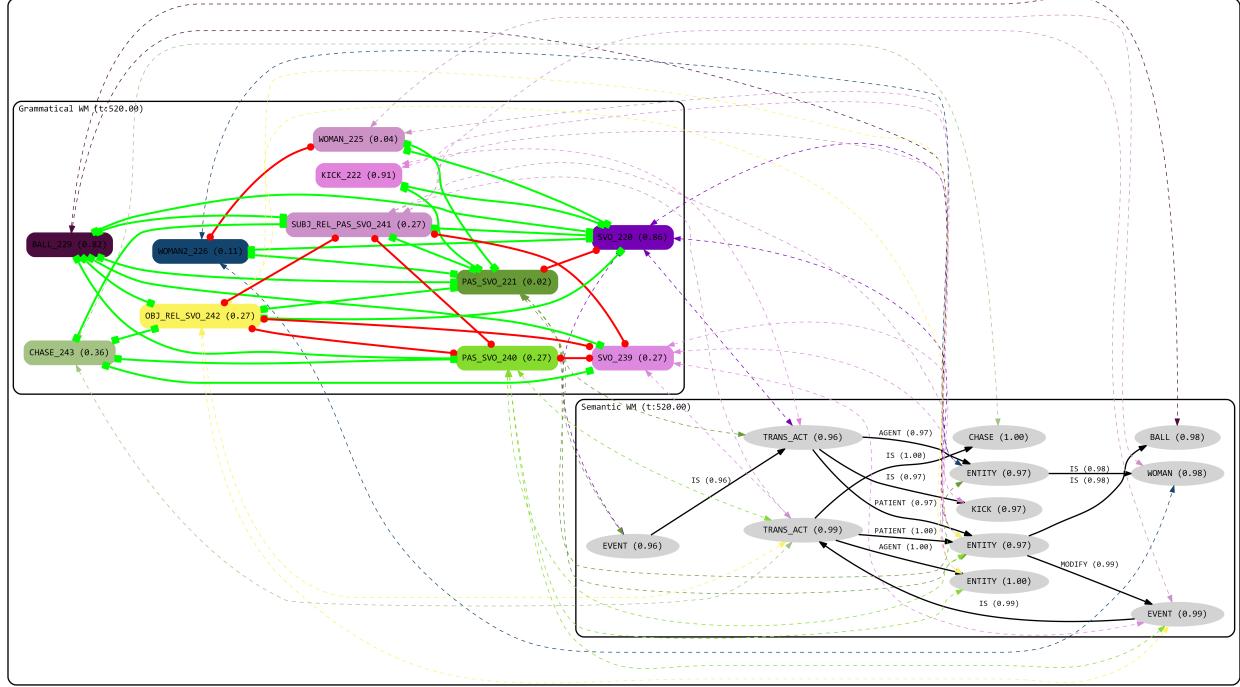


Figure 3.7: Snapshot of the state of Linguistic WM (Semantic and Grammatical WM) at  $t=520$ . Dashed lines between working memories represent cross WM activation links through which, in the present case, the Semantic WM instances can impact the activity values of the construction instances the invocation they are responsible for. This state of the Linguistic WM correspond to the one preceding the utterance “lady kick ball that is chase-ed by dog” at  $t=601$  (see tab. 3.1) (for video see [victorbarres.github.io/media/SALVIA\\_p\\_complex/states.mp4](https://victorbarres.github.io/media/SALVIA_p_complex/states.mp4)).

$t$	UTTER
601	“ball that is chase-ed by”
712	“dog is kick-ed by woman while”
934	“boy laugh”

Table 3.1: Utterances generated by SALVIA given the input described in sec. B.3, tab. B.6. Although the system, under time pressure, can produce fragmentary utterances, this does not prevent the succession of utterances to form a full well-formed sentence when taken together “ball that is chas-ed by dog is kick-ed by woman while boy laugh”

(Refer to Appendix B, sec. B.4.2 for a more detailed view of the simulation run)

### 3.5 Scene Description: Interaction between Visual Attention and Language Processes

#### 3.5.1 SALVIA: States

During the description process, each WM in SALVIA holds its own state composed of the (C2 network) of active schema instances currently relevant to the WM function. Fig 3.8 gives an example of those states at a given time. In all the WM the relation schema instances are represented as arrows.

The Visual WM contains a set of active perceptual schema instances that carry not only information regarding the perceptual meaning they hypothesize to hold, but also about the area of the scene that has

led to its instantiation. It can therefore re-orient attention towards this area if necessary, functioning also as a “deictic pointer”.

The state of the Semantic and Grammatical WM are directly reflecting the states that have been described in the conceptual models. They respectively hold concept schema instances and construction schema instances.

The phonological WM holds instances representing phonological/word-form content and their temporal relations that is here limited to the ‘next’ temporal relation (arrows) stipulating that one element should directly succeeds another during utterance production.

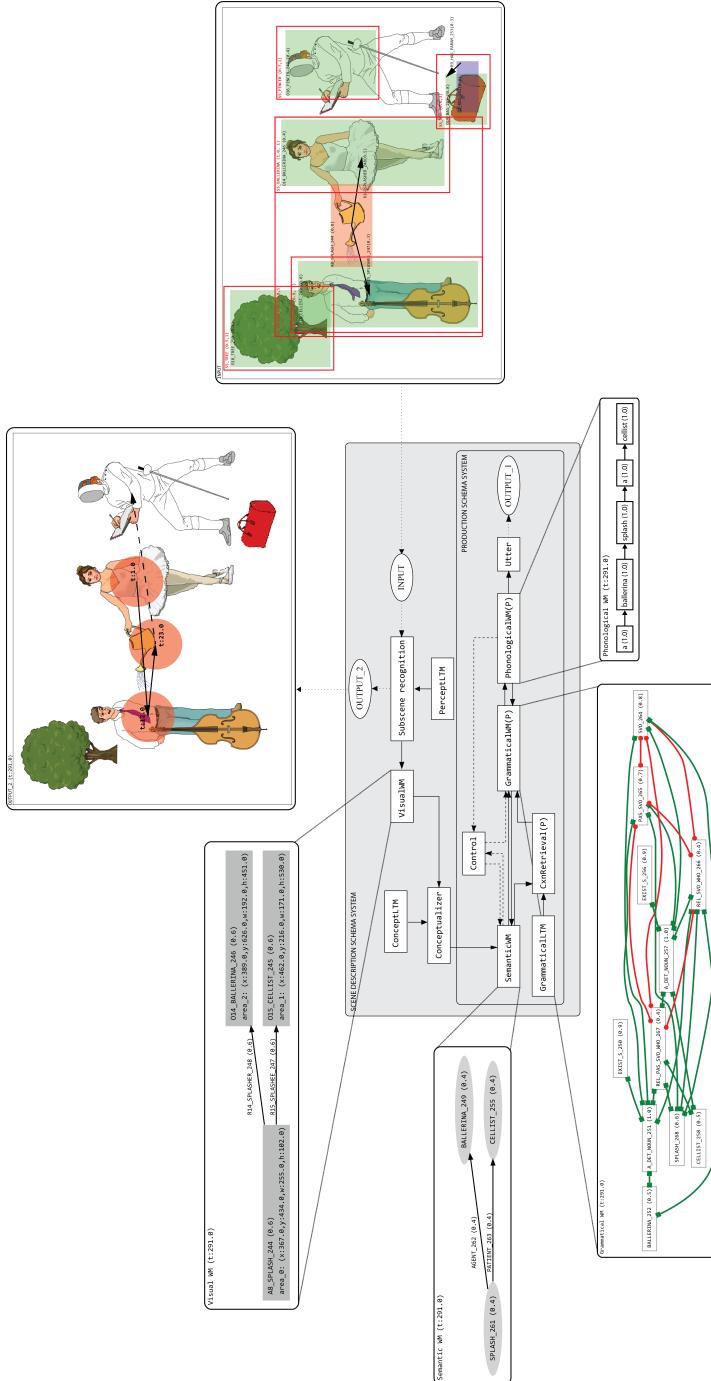


Figure 3.8: Snapshot of the various working memory states of the SALVIA model during the process of producing the description of a visual scene. The active schema instances forming the states of the Visual, Semantic, Grammatical, and Phonological working memories are zoomed in. The cross WM activation links are not shown for simplicity.

### 3.5.2 General Example: From Eye Movements to Utterance Production

The scene input used in this example is the same one as the one described in fig. 3.2.

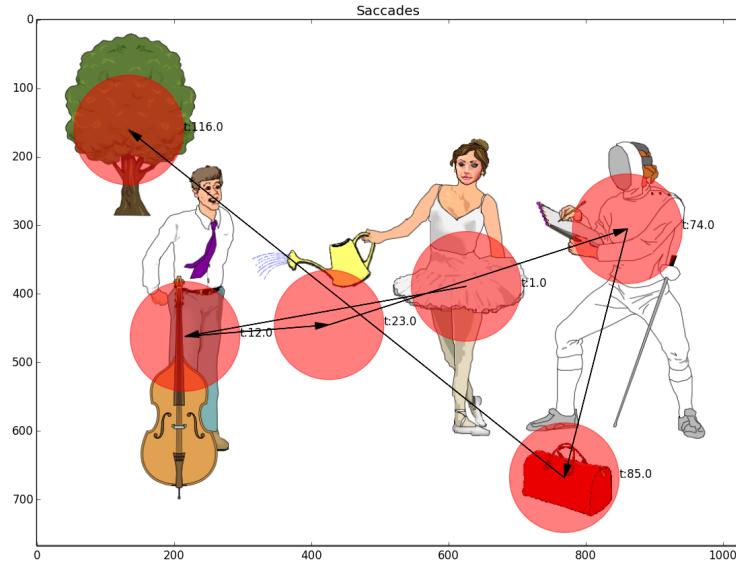


Figure 3.9: Model's saccades.

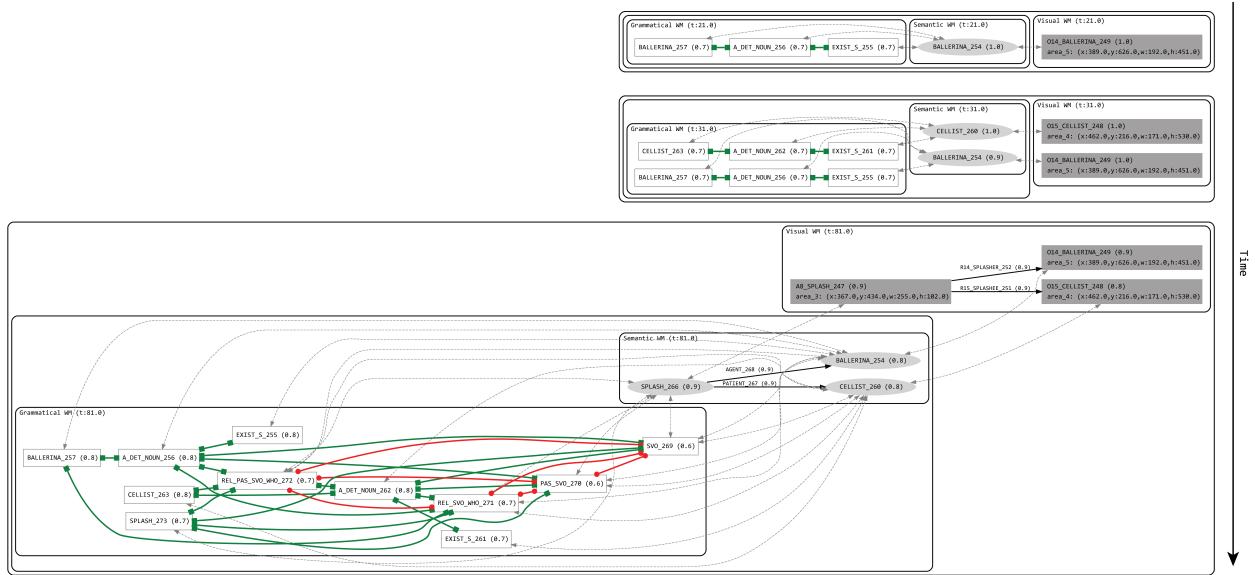


Figure 3.10: View of the interactions between visual, semantic and grammatical WM at the three first time points marking the updating of the at least one of the symbolic structure (C2 graph).

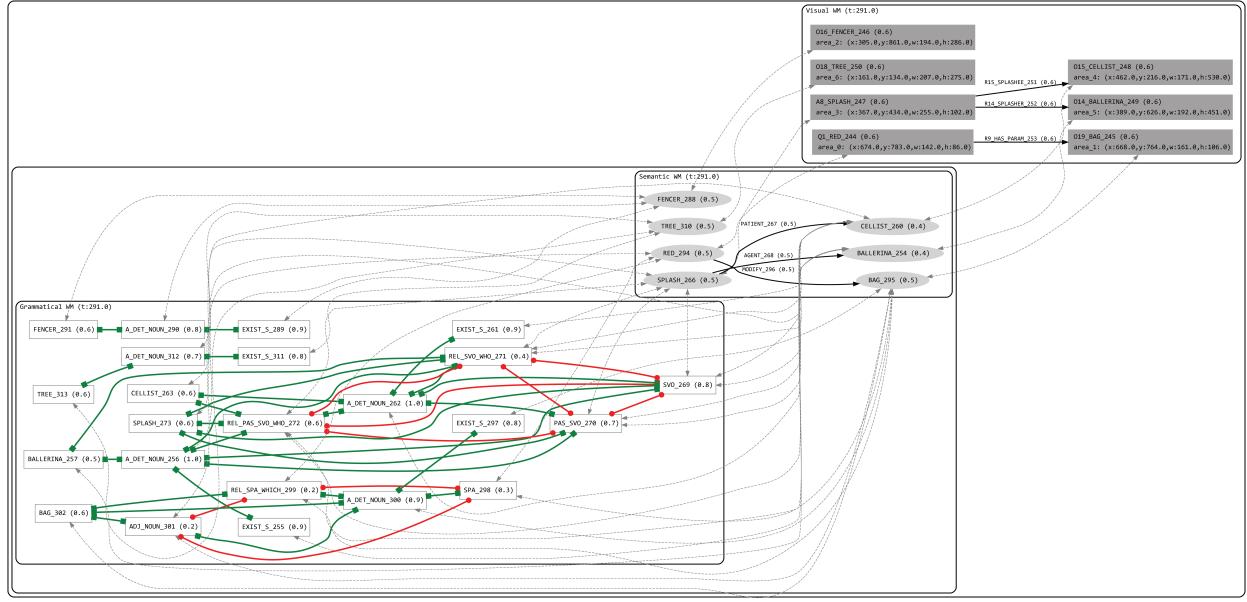


Figure 3.11: View of the interactions between visual, semantic and grammatical WM right before production is triggered (continued from fig. B.26)

The model outputs starting at  $t=310$  the utterances: “**a ballerina splash a cellist ... there is a bag ... there is a tree ... there is a fencer**”

(Refer to Appendix B, sec. B.4.3 for a more detailed view of the simulation run)

### 3.5.3 Saliency, Perceptual Guidance and Information Structure

The scene input used in these examples is the same one as the one described in fig. 3.2. However, the subscenes are limited to those involved perceptually representing the SPLASH action.

#### Case: Agent Fixated First

In this case, for the input scene, the SS\_BALLERINA subscene containing the agent has saliency 1, while the SS\_CELLIST subscene containing the patient has saliency 0.8. This triggers a bottom-up saliency preference towards inspecting the BALLERINA subscene first.

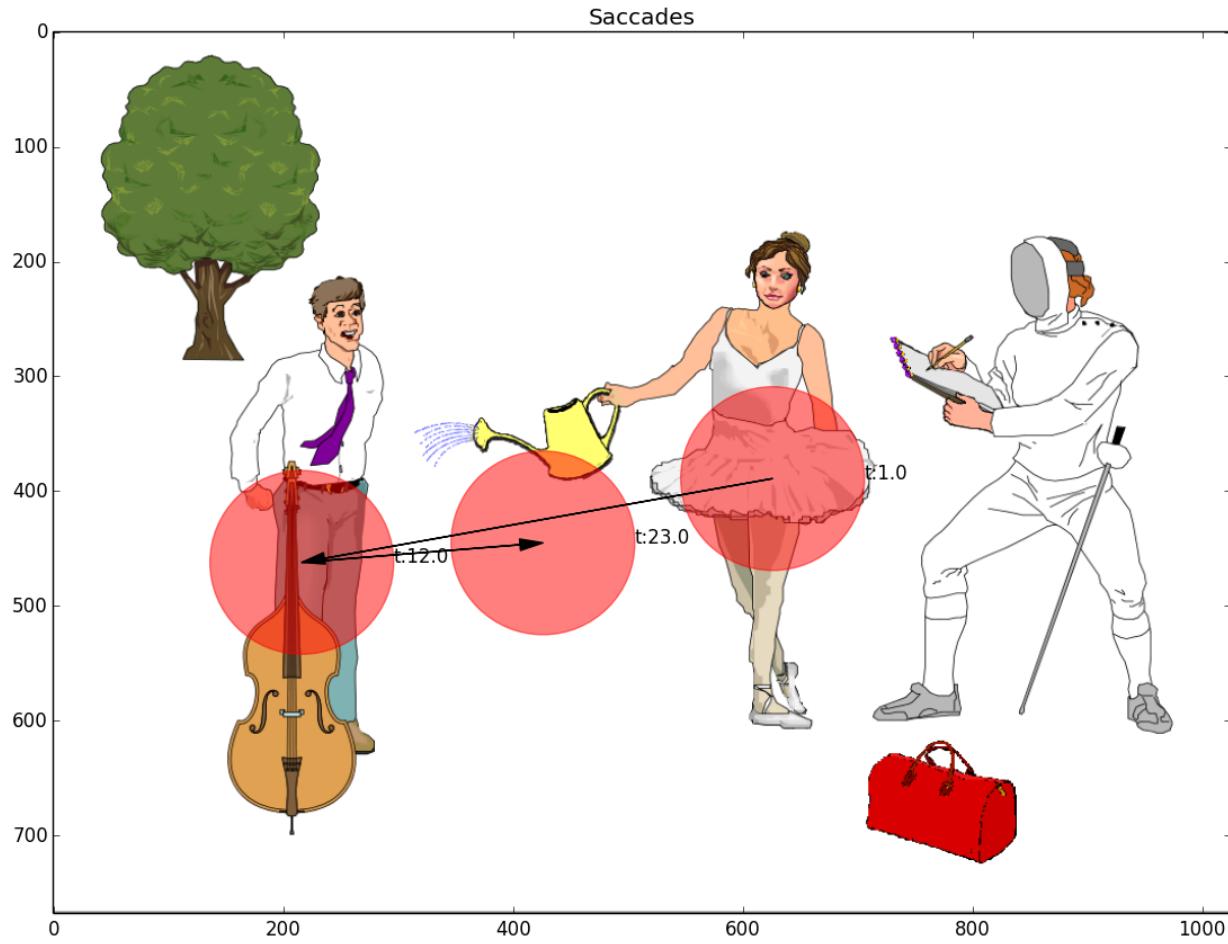


Figure 3.12: Model's saccades. The model inspects the BALLERINA subscene first, then the CELLIST and finally the SPLASH\_EVT subscene.

Fig. 3.13 displays the temporal activity levels of percept instances in Visual WM. The initial activation of the schema instances reflect the saliency of the subscene they belong to. O14\_BALLERINA is the first percept schema instance to enter visual WM followed by O15\_CELLIST with the later receiving a lower activation value. Note that because of the leaky integrator nature of the dynamics, the difference in activation level between agent and patient percept instances diminishes with time.

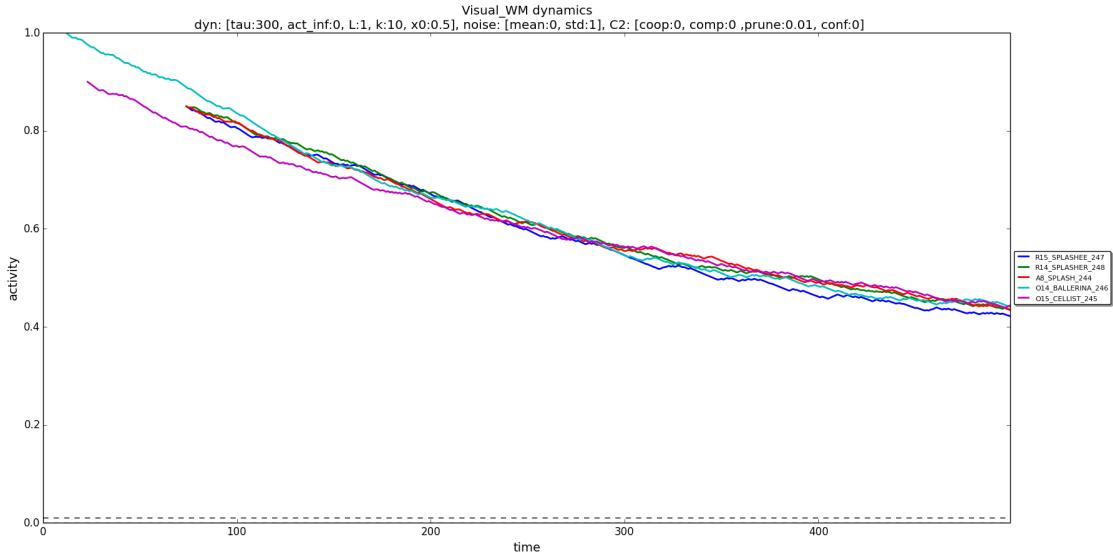


Figure 3.13: Percept schema instance dynamics in visual WM.

Fig. 3.14 displays the temporal activity levels of concept instances in Semantic WM. Following the timing in visual WM, the BALLERINA concept is instantiated first followed by the CELLIST concept. Their initial activation values reflect the activation values of the percept schemas they conceptualize at the time of instantiation. The difference in activation level between the BALLERINA and the CELLIST concept instance implements the information structure level of the SemRep with a higher activation value corresponding to a higher "focus" value.

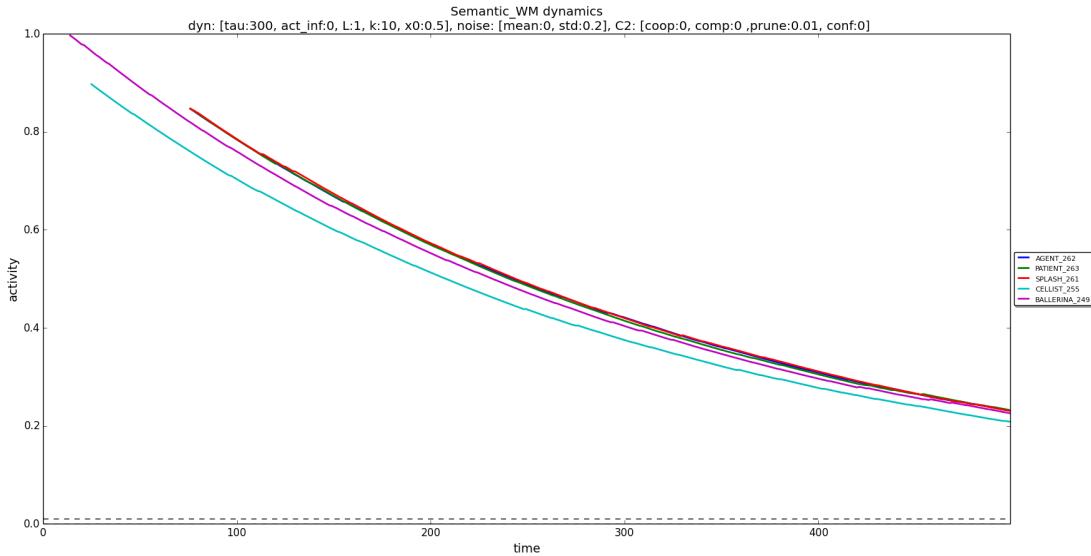


Figure 3.14: Concept schema instance dynamics in semantic WM.

Fig. 3.15 displays the temporal activity levels of construction instances in Grammatical WM. The main element to look at here are the two green line at the bottom of the graph. The top one represents the SVO cxn instance while the bottom one represents the PAS\_SVO cxn instance. In the grammar used for this

experiment, both constructions have the same preference value and should therefore start with a similar activation level. The difference in initial activation value results from the fact that the SemFrame of SVO matches the SemRep better since it favors an Agent focus. This initial boost results in SVO instances staying above PAS\_SVO during processing until production is triggered at  $t = 300$ . Consequently, the model will produce an utterance structured by the SVO cxn and therefore in active mode.

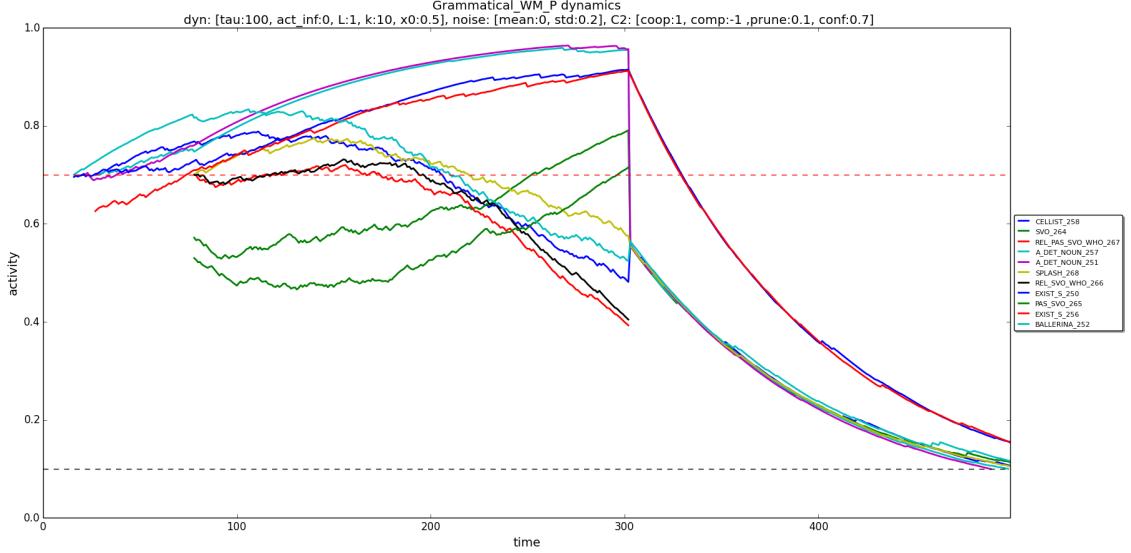


Figure 3.15: Construction instances activation in grammatical WM.

The model outputs starting at  $t=310$  the utterances: “**a ballerina splash a cellist**”  
(The reader should refer to Appendix B, sec. B.4.4 for a more detailed view of the simulation run.)

### Case: Patient Fixated First

In this case, for the input scene, the SS\_CELLIST subscene containing the agent has saliency 1, while the SS\_BALLERINA subscene containing the patient has saliency 0.8. This triggers a bottom-up saliency preference towards inspecting the CELLIST subscene first.

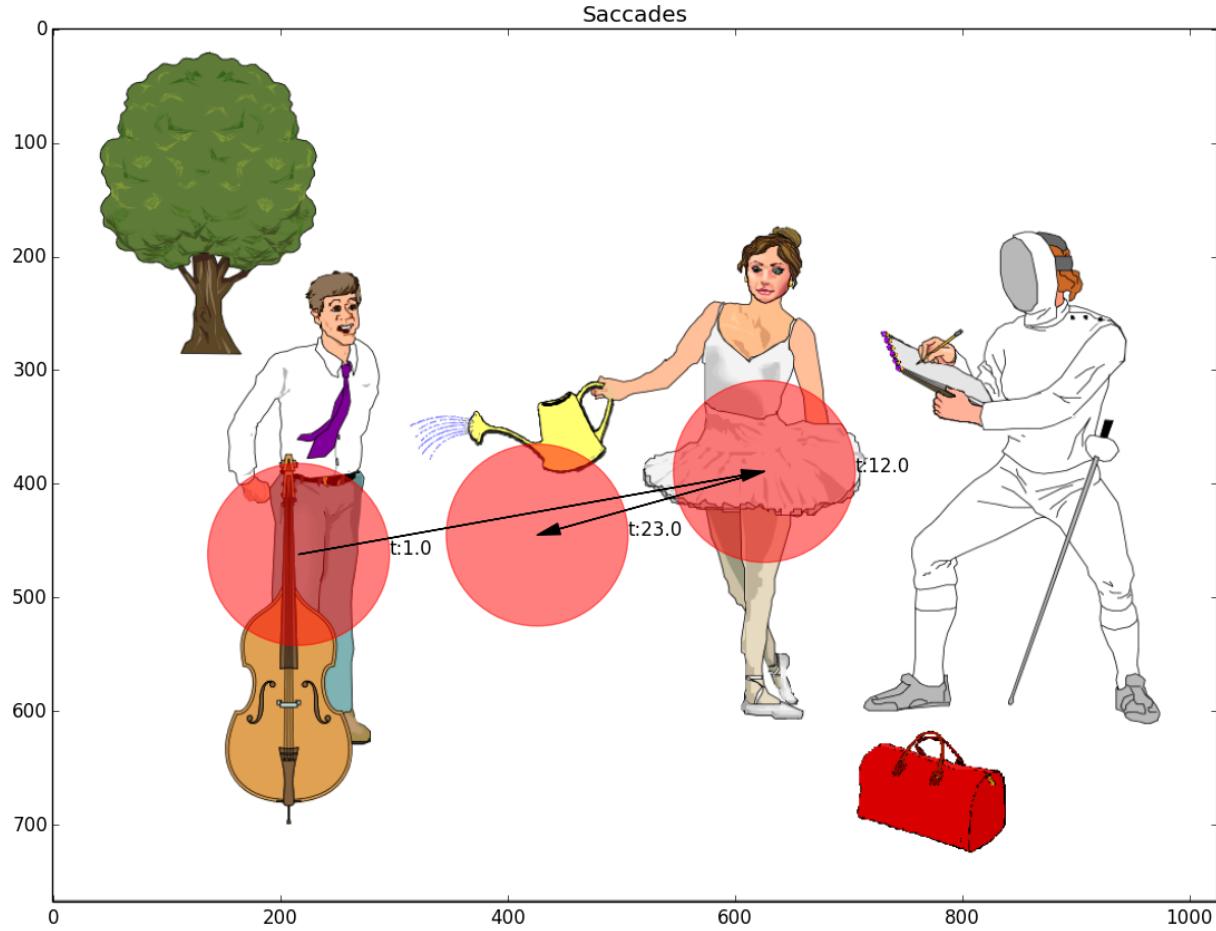


Figure 3.16: Model’s saccades. The model inspect the CELLIST subscene first, then the BALLERINA and finally the SPLASH\_EVT subscene.

The model outputs starting at  $t=310$  the utterances: “**a cellist is splash -ed by a ballerina**”

The figures equivalent to those above can all be seen in Appendix B, sec. B.4.5. They can be easily interpreted on the basis of the explanation given regarding the agent first case.

The following section will be dedicated to an in detail analyses of the interactions between task constraints, scene types, and dynamic characteristic of the system showing how SALVIA can shed a new light on results from psycholinguistics. Indeed, the next chapter will show that this apparently simple relation between initial gaze and grammatical voice choice (for example) is not straightforward.

## 3.6 Simulating Key Visual World Paradigm Psycholinguistic Results

### 3.6.1 Time Pressure and Utterance Fragmentation

#### Pilot Study: Good Enough Production Under Time Pressure

Given that time and incremental processing are of the essence in the question of modeling the language vision interface, incorporating time as a task variable through the intermediary of time-pressure allows the study of how the dynamics of the different processes (attention, apprehension, formulation, execution) and of their interactions is affected by changes in task timing.

In his initial research on developing TCG, Lee (2012) ran a pilot experiment, using complex scenes, during which subjects had to generate description of scenes with no restriction as to the linguistic form but under two different conditions: time-pressure and no time-pressure. The goal of this experiment was to look at the impact of time-pressure on the form of the utterances generated. In the ‘Quick task’ condition, participants were presented with a scene and asked to start verbally describing the scene extemporaneously after only a few seconds while they were the visual stimulus was still present . In the ‘Free task’ condition, participants pressed a button when they were ready to describe the scene, at which point the visual stimulus disappeared and the verbal description was recorded.

In both conditions, the semantic content of the description was similar but the ‘well-formedness’ of the utterances (and the arrangement of their components) differed. Table 3.2 provides an example of such description by two participants (KF and JI) under ‘Free task’ and the ‘Quick task’ condition respectively. The picture they were asked to described is shown in fig. 2.10.

Descriptions	
KF(Free Task)	JI(Quick Task)
a woman is kicking another woman in a blue dress in what looks like a boxing ring with many people watching the show	um there are two women one of them is kicking the other woman and sh- this looks like some kind of boxing match because they're in a ring and there are people watching them

Table 3.2: Example of the impact of time pressure on the quality of description utterances based on the scene shown in fig. 2.10

In order to capture the qualitative differences between produced utterances, those were analyzed according to two criteria : structural compactness (eq 3.1) and grammatical complexity (eq 3.2).

$$\text{Structural Compactness} = \frac{\text{Number of Core Words}}{\text{Number of Utterances}} \quad (3.1)$$

$$\text{Grammatical Complexity} = \frac{\text{Number of Embedded Structures}}{\text{Number of Utterances}} \quad (3.2)$$

Table 3.3 shows such analyzes for the two examples of Table 3.2.

Descriptions		
	KF(Free Task)	JI(Quick Task)
Analysis	[ <u><sub>u1</sub>woman</u> , <u><sub>kick</sub></u> , <u><sub>woman</sub></u> , ( <u><sub>e1</sub>blue</u> , <u><sub>dress</sub></u> ), ( <u><sub>e2</sub>in</u> , <u><sub>boxing_ring</sub></u> ), ( <u><sub>e3</sub>many</u> , <u><sub>people</sub></u> , <u><sub>watching</sub></u> , <u><sub>show</sub></u> )]	FILLER [ <u><sub>u1</sub>two</u> , <u><sub>woman</sub></u> ] [ <u><sub>u2</sub>kicking</u> , <u><sub>woman</sub></u> ] FILLER [ <u><sub>u3</sub>boxing</u> , <u><sub>match</sub></u> ] [ <u><sub>u4</sub>because in ring</u> ], [ <u><sub>u5</sub>people</u> , <u><sub>watching</sub></u> ]]
(CW, E, U)	(10, 3, 1)	(9, 0, 5)
<b>Structural compactness</b>	<b>10</b>	<b>1.8</b>
<b>Grammatical complexity</b>	<b>3</b>	<b>0</b>

Table 3.3: An example of description rating. (CW = Core Words, E = Embedded Structures, U = Utterances)

Based on these measures it is possible to look at the impact of time pressure on the quality of the descriptions produced. Lee concluded, based on the analyses shown in Table 3.4, that there was a *significant effect of time pressure on both structural compactness and grammatical complexity of the utterances produced*.

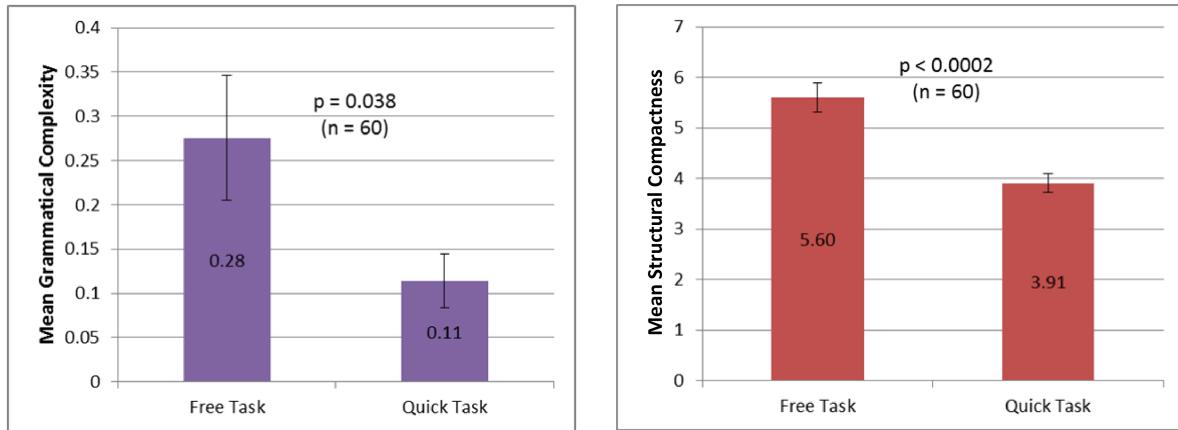


Table 3.4: Results from pilot experiment. (Left) Effect of time pressure on utterance grammatical complexity. Significant effect of time pressure on the grammatical complexity of the utterances produced (Free task: 0.28, Quick task: 0.11,  $p=0.038$ ,  $n=60$ ). Placed under time pressure, participants on average produce significantly less grammatically complex descriptions. (Right) Effect of time pressure on utterance structural compactness. Significant effect of time pressure on the structural compactness of the utterances produced (Free task: 5.60, Quick task: 3.91,  $p<0.0002$ ,  $n=60$ ). Placed under time pressure, participants packaged on average significantly less information in their utterances.

The reader is encouraged to refer to (Lee, 2012) for a full description of the empirical results.

### Simulating the Impact of Time Pressure on Utterance Quality

SALVIA was used to model the impact of time pressure on utterance quality in an attempt to simulate empirical results described above.

Parameter used were the following: Input rate = 100; Num restarts = 5; Time pressure = linspace(1,1000,10); Tau GramWM = 100; Tau SemWM = 1000; Max time = 2000; Total simulations =  $26*5*10 = 1300$

The ratio between input\_rate and GramWM\_tau was kept constant. In addition, the system was placed in a regime in which the SemanticWM can be considered constant at the order of time of GramWM processes ( $\text{Tau\_SemWM} >> \text{Tau\_GramWM}$ ). Tau\_GramWM was set to be in the same order of magnitude than the semantic input\_rate. Start produce was kept at 0. So in all case the first production takes place at  $t_0 = \text{time\_pressure}$ .

It is worth noting that the model only produces “core words” due to the simplicity of its syntax. This fits with the measure of utterance quality described in the previous section.

Tab. 3.17 show the simulation results. The simulations outputs regarding the impact of time pressure on syntactic complexity and structural compactness match well the empirical results even though the model was not tuned to do so. This suggests that the hypotheses embedded in the model’s architecture and dynamics capture, at least in part, some aspects of the architectural and dynamic constraints that shape the language-vision interactions. It is worth however to note that, rather than the direct matching of numerical values, the fact that the interactions between conditions fits those that were empirically observed might be a more reasonable successful comparison.

Importantly, the model offers new avenues for empirical studies. One of those stems from the fact that, in the simulations, the time pressure was varied continuously from high to low. This differs from the initial pilot experiment that only looked at two conditions (Quick or Free task). Future psycholinguistic empirical work should attempt to study the impact of time pressure on description qualities for a range of time pressures. This would provide a more robust data-set against which the dynamics of language-vision interactions could be computationally studied.

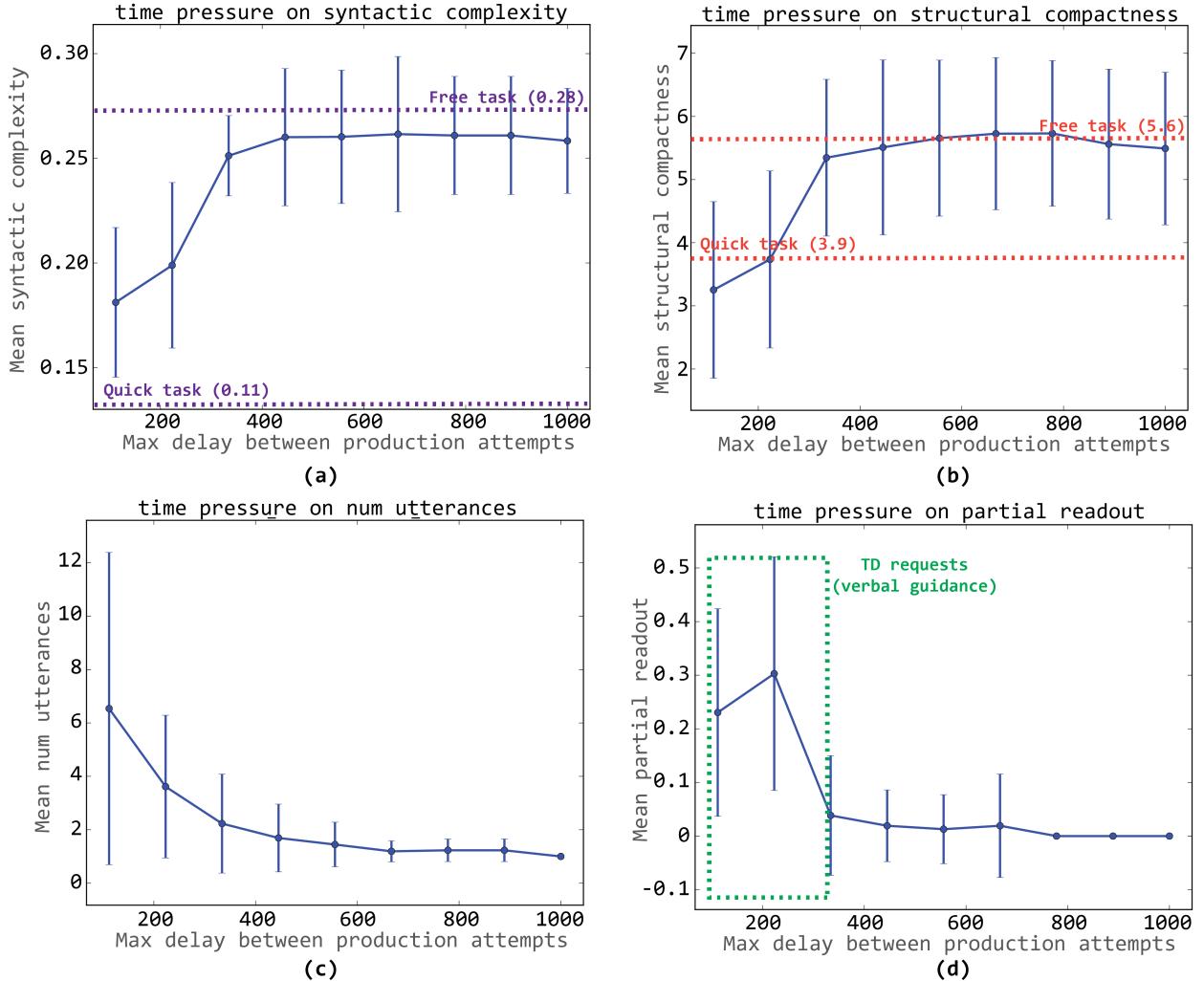


Figure 3.17: Simulation results. On the x axis is indicated, rather than time pressure, the max delay between production attempts for a more intuitive reading of the graphs (the higher the time pressure, the lower the max delay). Moving along the x axis, the system is in a state less and less constraint by time pressure. (a) and (b) Simulated effects of time pressure on both grammatical complexity and structural compactness. In both cases, dotted lines indicates the empirical values for both the “Quick case” and the “Free task”. (c) Effects of time pressure on number of utterances produced. (d) Proportion of utterances that trigger a partial readout, resulting in a TD requests for semantic information. The dotted box indicates the regime, under high time pressure, in which the system frequently produces partial readouts resulting in a general push toward verbal guidance of the visuo-attentional system in order for the production to be able to smoothly continue the utterance path it has engaged itself in.

### 3.6.2 Perceptual Guidance, Verbal Guidance, and Cognitive Thresholds

Psycholinguistics tend to use scenes that are much simpler than the ones that were used as a basis for the empirical studies described above. Those scenes could be referred to as “**minimal scenes**”. As was already mentioned in ch. 2, sec. 3.2.1, such scenes contain the bare minimum amount of perceptual content to create a scene that is not merely a collection of objects.

#### Attention Capture Effects as a Function of Scene Type

Kuchinsky has shown that, for a scene presenting a single transitive event, the level of difficulty of apprehension of the event and of the entities that it involves can explain in part the difference in results in the impact that Attention Capture Manipulation (ACM) has on the word order (Kuchinsky, 2009).

SALVIA will provide a quantitative account of Kuchinsky’s empirical findings while opening new avenue of research.

Scene content	<i>Easy objects</i>	<i>Hard objects</i>
<i>Easy event</i>	ACM - (Structural strategy)	ACM - (Structural strategy)
<i>Hard event</i>	ACM + (Incremental strategy)	ACM ?

Table 3.5: Different cases studied by Kuchinsky and the recorded impact of Attention Capture Manipulation (ACM on sentence grammatical structure. ACM-: No effect of ACM; ACM+ effect of ACM. ACM ?: unclear effect of ACM. *Easy event* refers to scenes in which there are direct perceptual correlates of the event taking place; *Hard event* refers to scenes in which there are no direct perceptual correlate of the event taking place, requiring further inspection of the scene content to infer the nature of the event taking place. *Easy objects* refers to scenes in which the object present are very easily recognizable both because of their nature as well as because of their high likelihood to appear in such scenes. *Hard objects* refers to the opposite situation. The notions of *easy* and *hard* compared across event and object should be taken as relative: Objects are overall easier to apprehend than the nature of the event they are involved in (*Easy objects, Hard event*) or the opposite (*Hard objects, Easy event*). See tab. 3.6 for examples of such scenes used by Kuchinsky.

If the event is easy to apprehend then initial bottom-up triggered attentional shifts will have no impact on the word order as the same perceptual event structure will be readily extracted independently of the sequence of eye-movements and will set the stage for the high level structuring of the description. However, if the precise nature of the participants is difficult to apprehend, feedback signals from the linguistic system might be triggered to gather more information about the event’s participants to allow for the retrieval of the appropriate lexical elements. The novel hypothesis here however, is that such feedback signals will be revealed only in the case of extemporaneous production under time pressure.

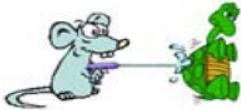
The case of extemporaneous production under time pressure has been studied by Bock and Griffin, but with three crucial caveats: first, the authors have discarded the produced utterance that were not fluent or did not conform to the active vs passive template. In the context of studying the dynamic interactions between active-vision and language production, a purely performance oriented analysis of situated language use, it is our contention that the impact of the interaction between temporal dynamics of attentional shifts in scene parsing and the temporal constraint stipulated by the task on the quality of the utterance produced (in terms of fluency, fragmentation, etc), should be an integral part of the modeling (see below for a more detailed analysis of this issue). The second caveat is that the scene used by Bock and Griffin might have been too simple to fully trigger an impact of time pressure on the visuo-linguistic interaction dynamics. Their result suggest that the subject were able to extract very quickly the general event information from the scene. Finally, an important aspect of extemporaneous production involves the fact that it involves generating utterances incrementally possibly prior to the time at which the full apprehension of the event, full conceptualization, and full grammatical formulation as taken place. Each word uttered therefore correspond to a decision point that commits the speaker to a some grammatical structure, which, as more perceptual information is gathered, need to be smoothly altered in order for the utterance to appear continuous (i.e. without correction, rephrasing, restarts etc.). In extemporaneous production under time pressure, “utterance continuity” becomes an important factor in determining the course of the visuo-linguistic interaction process.



(A) Easy event with easy characters



(B) Difficult event with easy characters



(C) Easy event with difficult characters



(D) Difficult event with difficult characters

Table 3.6: Example scenes used by Kuchinsky (2009), which are categorized into four types depending on the codability of the depicted events. The cueing effect was found only for type (B).

Time pressure	Utterance	Cued first	First TD request
Yes	(70)lady(160)[lady] talk to(221)[lady talk to] people	Yes	(160)patient(uncued)
No	(520)woman talk to people	Yes	None

Table 3.7: **Hard event, easy objects, agent cued.** Only under time pressure does the attentional cueing have an effect on the output. Input rate is set to 100. Active voice is assumed here to be always preferred to passive voice. **Time pressure:** stipulates whether or not time pressure was applied to the production system (Yes: time pressure = (input rate)/2, No: time pressure = (input rate)\*5). **Utterance:** Production output in the form ((t)'utter')\*, where (t) indicates at which time step 'utter' was produced. '[' ]' highlights utterance continuity. **Cued first:** has value 'Yes' if the first word uttered correspond to the semantic content of the perceptual areas that has been cued (here agent). **First TD request** time and nature of the first top-down attention-orienting request sent from the language system to the perceptual system.

### Perceptual vs. Verbal Guidance: a Function of Scene Types and Time Pressure

**Hard Event, Easy Objects** SALVIA's simulation of a case in which the participants are easy to identify but the structure of the event is difficult to extract while in Table 3.7.

The model is provided with a scene input described in fig. 3.18 (same conventions as those used in fig. 3.3). The scene is considered to be an hard event scene since there are no direct perceptual features marking the event-structuring action (while the participants are rather typical). Cognitive reasoning is required to recover the event identity.

Under time pressure, an **incremental attentional strategy** is chosen. This strategy is illustrated in fig. 3.19. The first TD request is sent to direct the system towards the uncued patient after the system has already processed the agent area towards which it was initially cued. In order, attention is directed towards (1) cued agent area, (2) action area triggering the retrieval of the event structure allowing for the sentential transitive active voice construction to start piloting the utterance production, (3) towards the uncued patient area, a choice that is piloted by the TD request from the Grammatical WM that needs information about the patient to finish the utterance.

In the high low time pressure case, all the semantic information has been extracted prior to the system having to generate an utterance. It uses the transitive active voice and does not need to generate top-down requests. The order in which the areas are attended to is irrelevant. A key aspect of the modeling results needs to be highlighted here regarding the fact that, in the low time pressure case, the cued element appears first in the utterance sequence. It should appear with (near) chance probability as subject or object. However, the cued element appears first due to the combination of active voice preference and **pre-conceptualization of event as "TALK"** (vs "LISTEN"). By pre-conceptualizing the event as "TALK", the cued entity (WOMAN) is bound to appear as agent in the semantic representation. If the active voice

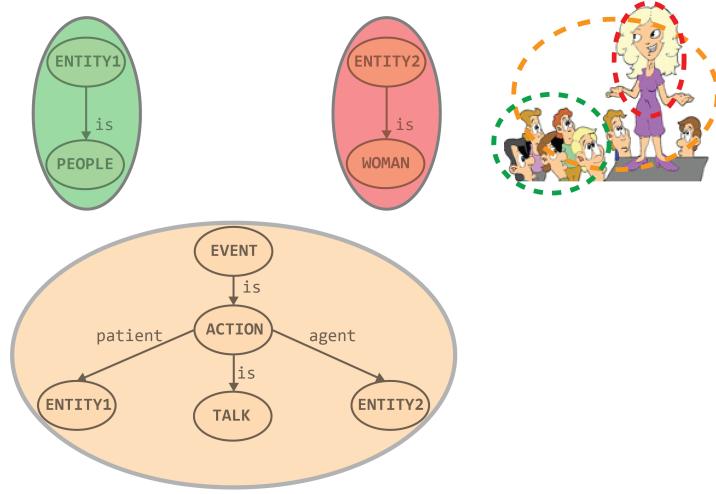


Figure 3.18: Illustration of a scene input. (Left) Subscenes and the semantic content that is retrieved when they fall in attentional focus. The organization from top to bottom reflects the hypothesized hierarchy between the subscenes in terms of how easily they are built (easiest at the top). In this case, the specific object anchored event subscenes are hypothesized to be most easily retrieved from the input (in absence of other factors such as top-down requests). Conversely, the subscenes that more abstractly structure the scene content, building action-based event level relations is assumed to be, relatively, harder to build. (Example of a scene with Easy object, Hard event) (Right) The image regions they are linked to. (scene adapted from (Kuchinsky, 2009))

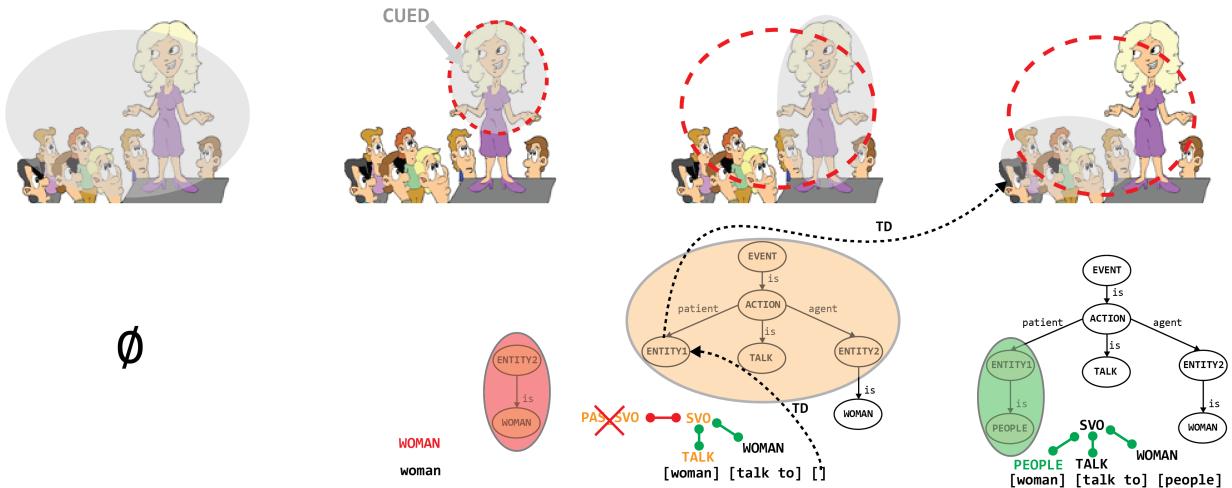


Figure 3.19: **Structure extension.** Process consistent with the incremental view (Gleitman et al 07, initial perceptual guidance). (High time pressure case) (grey: focus region, dotted-red: subscene perceptual structure) (Details in text)

Time pressure	Utterance	Cued first	First TD request
Yes	(312)mouse squirt at turtle	No	(70)agent(uncued)
No	(520)mouse squirt at turtle	No	None

Table 3.8: **Easy event, hard objects, patient cued** In both time pressure cases, the attentional cueing has no effect on the produced output. Input rate is set to 100. Active voice is assumed here to be always preferred to passive voice. **Time pressure:** stipulates whether or not time pressure was applied to the production system (Yes: time pressure = (input rate)/2, No: time pressure = (input rate)\*5). **Utterance:** Production output in the form  $((t)'\text{utter}')^*$ , where  $(t)$  indicates at which time step  $\text{'utter}'$  was produced. Parentheses highlight utterance continuity. **Cued first:** has value ‘Yes’ if the first word uttered correspond to the semantic content of the perceptual areas that has been cued (here agent). **First TD request** time and nature of the first top-down attention-orienting request sent from the language system to the perceptual system.

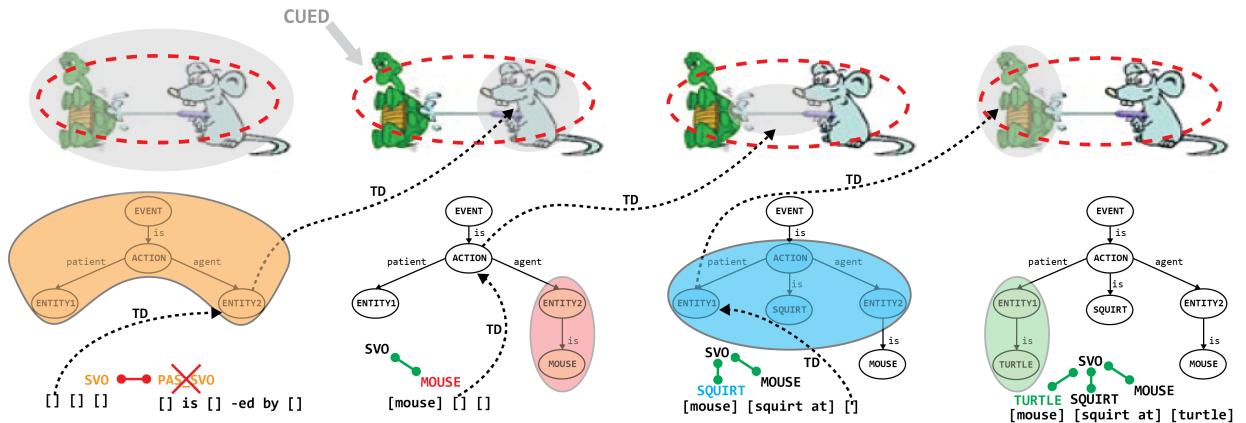


Figure 3.20: **Structure specification.** Process consistent with the structural view (Griffin et al. 00, linguistic guidance). (High time pressure case) (grey: focus region, dotted-red: subscene perceptual structure). (Details in text)

is favored, then, in absence of other constraints such as time pressure, it will be mapped as the subject of the utterance, appearing first.

This issue regarding the relation between conceptualization and language-vision interaction under time pressure will be addressed in discussion.

**Easy Event, Hard Objects** SALVIA’s simulation of a case in which the main structure of the event is easy to perceptually extract while the participants are difficult to identify is shown in Table 3.8.

The scene input used in this case is the one shown above (cf. Figure 3.3). The scene is considered to be an easy event scene since there are direct perceptual features marking the event-structuring action (while the participants are rather atypical).

Under time pressure a **structural attentional strategy** is chosen. This strategy is illustrated in fig. 3.20. The first TD request is sent to direct the system towards the uncued agent: since the event-action gist is available first, the transitive active voice construction is activated right away, and takes over the process of guiding the attention towards the agent area, whose information is required to fill the subject slot and therefore start producing utterances. The attention cueing towards the patient is overridden by TD structural constraints. In order, attention is directed towards (1) the general action area from which the semantic gist of the event (transitive action) is extracted, (2) agent area, following from the need to fill in the subject slot of the transitive active voice construction (3) action area, to fill in the verb slot in the construction, and finally (4) towards the cued patient area, a choice that is piloted by the TD request from the Grammatical WM that needs information about the patient to finish the utterance.

As in the previous case, in the absence of time pressure, all the semantic information has been extracted

prior to the system having to generate an utterance. It uses the transitive active voice and does not need to generate top-down requests. The order in which the areas are attended to is irrelevant.

**Comparison** Looking only at the case in which time pressure is applied, those simulations show how the structural strategy emerges in the case of scene for which the event is easier to apprehend than the objects: the quick apprehension of the event's structure sets the grammatical frames which governs from then on the attentional system. On the other hand, in the case of easy objects and hard event, the incremental strategy tied to perceptual guidance emerges during the early steps of the process: the grammatical processes and therefore the utterance are shaped by the patterns of attentional shifts until the structure of the event can be extracted.

In the absence of time pressure, the simulation finds no effect of pattern of attentional shifts appears on the output utterance. However this indicates neither a structural or an incremental strategy, rather, the impact of the order in which the information is received on the Grammatical WM processes becomes negligible in comparison to grammatical endogenous factors (e.g. grammatical preferences, lexical accessibility, priming effects, etc)

### 3.6.3 Saliency, Perceptual Guidance and Information Structure: Impact of Saliency on the Use of Active vs. Passive Construction.

In the previous section, preference for active voice was assumed and saliency of the various objects was only manipulated through the subscene structure that composed the input scene representations. SALVIA, however, offers a platform on which much deeper analyses of those issues can be carried out by expanding the parameter space considered. This is of course costly and, for the present work, only one result will be put forward, as a way to open the door for future simulations.

In order to better understand the relation between ACM and sentence structure, SALVIA allows to manipulate the relative preferences of SVO and PAS\_SVO constructions, the saliences of subscenes and of the perceptual schemas that compose them, etc.

The previous chapter (ch. 2, sec. 3.5.3) presented simulation cases in which changes of saliencies as well as differences in preferences between active and passive voice constructions, led to different state trajectory and ultimately different utterance outputs.

It would be worth investigating the joint impact of saliencies and voice preferences on description output, going further than what empirical ACM studies have done, possibly opening new avenues for psycholinguistic studies.

As a first preliminary result, SALVIA was run on a set of inputs corresponding to single transitive event, in which the saliency of the agent and patient varied over a given range. The semantic input was static (i.e. that only the saliency varied between inputs, not the order or timing at which they semantic information was received.)

The parameters for the simulations are: SVO preference = 1, PAS\_SVO preference = 0.7, Saliency range for patient and agent [0.5, 0.6, 0.7, 0.8, 0.9, 1.0], Random restarts = 30.

The summarized results of the simulations are shown in fig. 3.21

Such results should be taken to indicate that for now, this work only scratched the surface of the complexity of the interactions between the various parameters that can impact the SALVIA behavior and hence the observed patterns of relations between attentional and utterance sequences. It nevertheless offers a flexible framework in which such complex interactions can be quantitatively studied.

## 3.7 Conclusion

### 3.7.1 Summary of Results: SALVIA as a Cognitive Model of Language Production

SALVIA offers a novel Schema Theory based cognitive model of language production focusing: (1) on the dynamic coordination between an incrementally built message, tied here to the incremental attentional

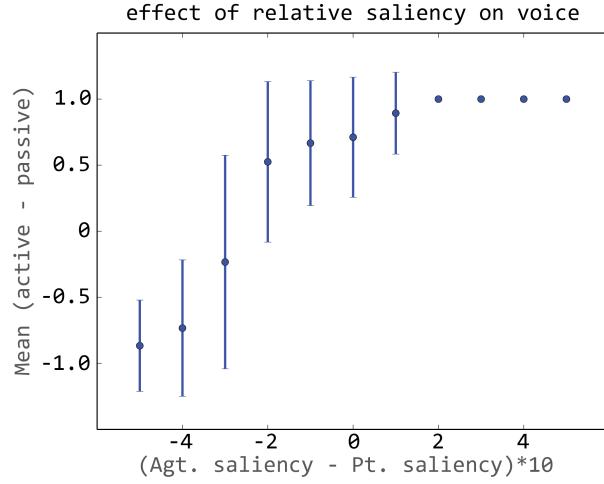


Figure 3.21: Transitive action (static). Given an input SemRep showing a simple and static transitive action involving one agent and one patient, the figure shows the prevalence of active voice and passive voice as a function of the saliency of the agent and the patient. A value of 1 signifies that only active voices are produced while -1 signifies that only passive voices are produced. As the saliency of the agent increases compared to that of the patient, the proportion of active voice descriptions increases monotonically. The negative value for the equivalence point ( $y=0$ ) reflects the lower preference value of the passive construction compared to the active construction.

Gleitman et al. 07	Griffin and Bock 00
Attentional processes ↔ Linguistic processes	Perceptual guidance ( $\rightarrow$ )      Linguistic guidance ( $\leftarrow$ )

parsing of a visual scene , and (2) on the role played in the visual scene description processes by language-driven feedback signals, insisting on the fact that vision-language interactions should be construed as part of a perception-production cycle as shown in fig. 3.22.

SALVIA should also be seen as a test case for the Schema Theory modeling methodology.

### 3.7.2 Summary of Results: SALVIA as a Psycholinguistic Model

Kuchinsky (2009) defended the hypothesis that the difference observed between the experiments of (Gleitman et al., 2007) and (Griffin and Bock, 2000), concluding respectively in the position re-summarized in tab. 3.7.2 can be explained by factoring in the nature of the scene.

SALVIA offered simulations that concord this hypothesis in the case of description under time pressure. However, for SALVIA, perceptual and linguistic guidance views correspond to *two extrema on a continuum*. It therefore provides a computational support for Kuchinsky's theory while extending the analysis to complex scenes where perceptual and linguistic guidance become intertwined.

TCG-SALVIA highlights how outcomes that have been attributed to specific symbolic processes could instead derive from variations in the temporal dynamics of the system's states and of their interactions. It shows how the quality of descriptions can derive from the interactions between task parameters (time-pressure) and the system's dynamics (Cooperative computation parameters, timing of BU and TD signals).

Many issues remain to be addressed by future work. The impact of Grammatical WM processes' characteristic time in the present model is tied to the preference for SVO active voice construction in two ways, a higher initial activation for SVO is assumed while the Semantic WM is considered to be quasi-static compared to Grammatical WM. It is legitimate to ask what happens if these conditions are dropped. In addition, the impact of various other factors on the description process should be investigated:

- Impact of idiosyncratic preferences between argument structure constructions.
- Impact of information structure on construction choices.

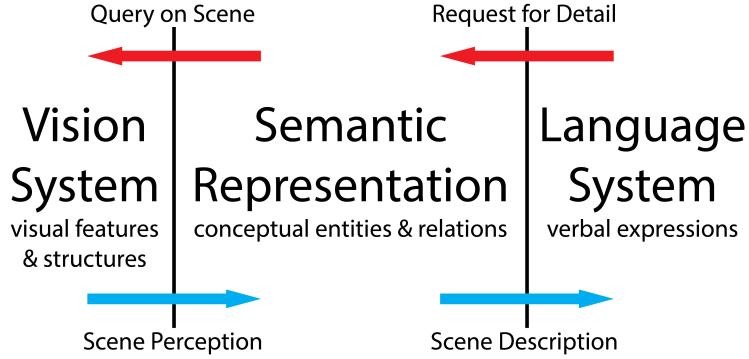


Figure 3.22: Vision-language as a perception-production cycle. SALVIA models the role played by visuo-attentional processes in incrementally extracting the perceptual meaning of a scene, meaning that is then conceptualized into a linguistic message dynamically mapped onto utterance forms (bottom arrows). But, importantly, it insists on the fact that this “feed-forward” pathway has to be supplemented by a “feedback” one through which the language production system can send requests to the visuo-attentional system in order to adapt the attentional processes to time dependent needs of the language processes (figure adapted from (Lee, 2012))

- Impact of grammatical semantics constraints (how agent-like is the agent ...)

SALVIA indirectly opened a discussion regarding the impact of the interplay between task constraints and language-vision interactions on conceptualization processes.

Fig. 3.23 shows a situation that differs from the one presented above in fig. 3.19. Whereas in the simulated case, SVO (active voice) was considered to be necessarily preferred to the PAS\_SVO (passive voice) for the description of the scene content, here both are considered. Compared to fig. 3.19, in the present case the PEOPLE are cued instead of the WOMAN. In this case, visuo-attentional system retrieves first the information about both the PEOPLE and the WOMAN before it tackles the question of the nature of the (hard) event they are involved in. As it attempts to retrieve perceptual information regarding the nature of the event, since PAS\_SVO and SVO are in competition, there are two possible outcomes depending on the winner. To describe those, it is necessary to abandon the idea of a pre-defined conceptualization of the action. Indeed, if SVO wins, placing “people” in the subject position with the active voice already selected and “woman” in the object position, then, the necessity to recover the missing semantic information required to specify the verb will trigger both visuo-attentional processes as well as *conceptualization process*. The action has to be conceptualized as LISTEN for the active voice to be used. In the converse situation in which the PAS\_SVO construction instance wins, the action will be required to be conceptualized as TALK.

The top-down feedback sent by the language system onto the visuo-attentional system necessarily need to also include information relevant to the conceptualization processes so that the perceptual content retrieved can also be construed in a way that smoothly fits in with the ongoing grammatical processes mapping meaning onto form.

### 3.7.3 Representation and Semantics of Complex Visual Scenes

It is worth interrogating the nature of the visual scenes used in the study of vision language interactions and how the type of scenes used could impact the analyses of the processes.

It is worth contrasting the examples of scene used in psycholinguistic studies of language production as shown, for example, in tab. 3.6, and a more realistic scene as represented by the photograph shown in fig. 3.9 (left). The scenes used in psycholinguistic experiments consist of a simple drawing of a unique event. It should be immediately clear that the very nature of those scenes will impact to what extent the various visuo-attentional processes will be involved in their analysis (but see (Knoeferle, 2016) for an overview of the more recent use of complex visual contexts in VWP experiments, which, however, are beyond the scope of the present paper).

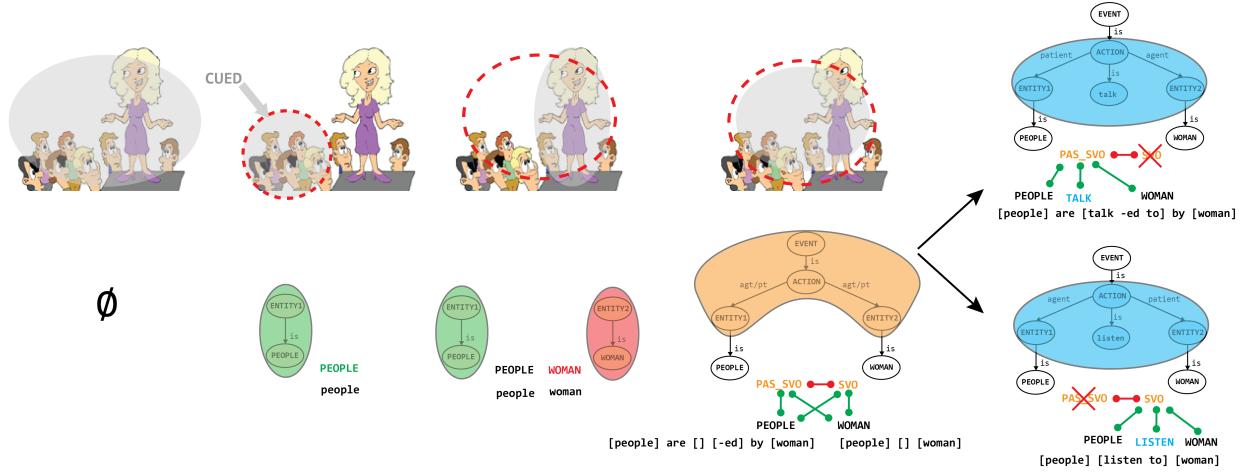


Figure 3.23: Impact of incrementality on conceptualization. (Details in text)

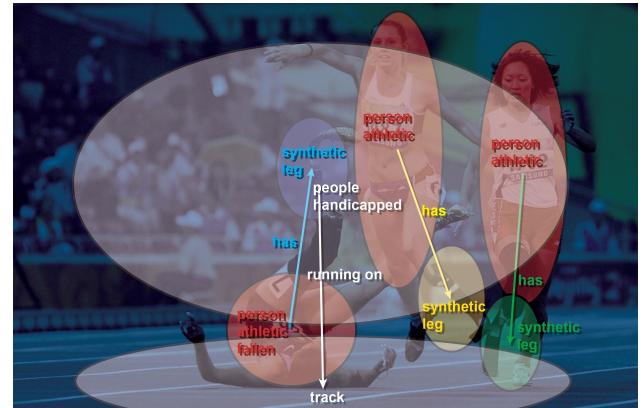


Table 3.9: Left) A complex scene. (Right) A example of scene representation (SceneRep) that could be used to represent the perceptual content of this scene.



Figure 3.24: Photo taken of President Obama by Chief Official White House Photographer Pete Souza. The photo was accompanied by the caption: “President Barack Obama jokingly puts his toe on the scale as Trip Director Marvin Nicholson, unaware to the President’s action, weighs himself as the presidential entourage passed through the volleyball locker room at the University of Texas in Austin, Texas, Aug. 9, 2010. (Official White House Photo by Pete Souza)” (<https://www.flickr.com/photos/obamawhitehouse/4921383047/in/photolist-95Yjre-8uTnHR>)

What perceptual information should the cognitive system apprehend at each time during the processing of a visual scene in order to support the formulation of a description? This question finds a relatively direct answer in the case of the simple drawings since those represent an already conceptualized/idealized version of a real scene, that can be wholly apprehended through very few fixations.

The scene shown in fig. 3.9 already presents a much more complex situation. It can be analyzed as potentially containing multiple sub-scenes, each with its own complexity. In addition, following the first few saccades, the subject might have extracted some information regarding the perceptual meaning of the scene: e.g. this is a scene of a track and field running race. But the perceptual representation might need to be revised once the viewer realizes that one of the the runner is missing a limb, leading to check the other runners, noticing that they too miss a limb. Influenced by the intrinsic property of the image content, the task, the state of the viewer etc., the attentional system will incrementally build the relevant sub-scenes from which are derived the semantic content to be formulated. Those sub-scenes end up forming their own complex network that represent the current outcome of the scene apprehension process (Itti and Arbib, 2006).

It is worth noting that even in the case of a photograph, we are still far from the complexity of a naturalistic task such as asking a subject to describe what is you are seeing at a random moment of her day.

We have propose with SALVIA a treatment of visual scene parsing, representation, and interaction with the language system that does not fall prey to some of the main oversimplifications that one could derive by using simple drawing as the scene prototype. The model does not treat scenes as sets of visual elements but takes into account the possibility of a cognitive representation made of a hierarchy of anchored subscenes. However so far the treatment of the visual input has essentially not been integrated to the model and much work remains to be done in order to be able to handle scenes such as the one showed in Figure 3.24. So far no model integrating vision and language comes close to being able to generate the kind of caption that human readily produce and understand as the one associated with this official white house photography. We nevertheless contend that in order to progress in our understanding of the human language system in its contextual use, it is worth assuming the availability of visual computation outputs that the vision models cannot yet deliver, so that progress can be made jointly by the language and vision computational communities towards the better understanding of high-level cognitive representation of visual scenes.

# Chapter 4

## Template Construction Grammar (TCG): Formalism for Dynamic Grammatical Processing of Incrementally Built Semantic Representations.

*"If one's goal is to "naturalize" semiolinguistics structures, one has to account for them as a special kind of emerging Gestalt. A key consequence of this conversion of paradigm is to abandon the requirement that models of natural syntactic structures be formal (algebraic, combinatorial, etc.) Indeed, in natural sciences, the mathematical structures used for modeling an empirical phenomenal realm have nothing to do with any "ontology" of this realm. Their scope is to provide appropriate computational tools for reconstructing phenomena. It is therefore a deep epistemological mistake to believe that natural languages have necessarily to be modeled using formal languages."*

Petitot

Morphogenesis of Meaning

### 4.1 Introduction

Template Construction Grammar (TCG) is a novel implemented computational construction grammar framework. It is part of a more general effort to develop a neurolinguistic model of vision-language interactions and follows the tenets of Schema Theory as a cognitive-level brain modeling philosophy (Arbib, 1989). Its main focus is to provide a framework to model the brain's capacity to seamlessly and dynamically coordinate the multiple sub-systems involved in situated language use. This chapter presents the details of the formalism and processing supporting TCG in the context of language production (A more informal presentation of TCG can be found in (Barres, 2017).)

The nature of the cognitive processes at play during the generation of visual scene descriptions has been investigated by psycholinguistic experiments based on the Visual World Paradigm (VWP): the subject is asked to verbally describe a visual scene while her eye-movements and utterances are recorded. The relations between those two temporal sequences (saccades and words) reveal complex dynamic interactions between three cognitive systems: visual, semantic, and grammatical, each having its own internal dynamic behavior (Knoeferle and Crocker, 2008). As visual information is actively gathered through attentional exploration of the scene, this information is readily used to update the semantic representation (message) to be linguistically

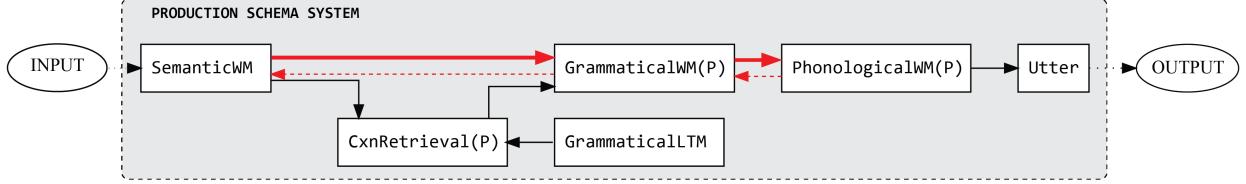


Figure 4.1: Language production sub-system of the Schema Architecture Language-Vision InterAction model (SALVIA). Each box corresponds to a system with arrows indicating message passing. The core of the system lies in the articulation and temporal coordination of the two main working memory systems: SemanticWM (message) and GrammaticalWM (grammatical processing). (See main text for details)

conveyed in a description. The state of the grammatical processes, mapping meaning onto verbal form, are constantly updated to adapt to changes in the semantic state.

Schema Theory offers a top-down counterpart to the bottom-up neural network modeling approach. It focuses on the adaptive and dynamic nature of the interactions between distributed computational units, respecting the computational style of the brain.

Template Construction Grammar (TCG) provides a new chapter in the application of schema theory to language (Arbib et al., 1987). As a construction grammar (CxG), it assumes that constructions, defined as a meaning-form mappings, are the only units of grammatical knowledge. Building on the VISIONS-Schema System model of visual scene interpretation (Draper et al., 1988), TCG plays a central role in the Schema-Architecture Language-Vision InterAction model (SALVIA) where it operationalizes the grammatical processes at work in the dynamic and adaptive translation of visually extracted, incrementally built semantic representations into online utterance production (see Ch. 2).

Template Construction Grammar is part of a growing number of computational construction grammar frameworks, all sharing the same goal to model situated language use and its relation to non-linguistic cognitive systems. Being based on Schema Theory, TCG is uniquely placed to serve as bridge between the computational CxGs with a classic A.I. focus (Fluid Construction Gramar (Steels, 2011), Embodied Construction Grammar(Bergen and Chang, 2005; Feldman, 2010) and the neural network implementations (Dynamic Construction Grammar (Dominey et al., 2009; Hinaut et al., 2015; Hinaut and Dominey, 2013).

TCG is first and foremost designed to fit the requirements posed by the question of the dynamic interactions between visuo-attentional and language processes tackled in the context of brain theory. Its goal is not offer a CxG analysis of a particular language. It insists on the need to develop models that accommodate the requirements of distributed computation. Anchoring the model in vision-language interactions offers a fruitful terrain to explore those computational issues. TCG however is not a priori limited to generating visual descriptions.

TCG as a grammatical processing model has been used to simulate key psycholinguistic results regarding the interactions between visual scene attentional parsing and utterance characteristics (Lee, 2012) (see Ch. 2).

## 4.2 System-Level View

### 4.2.1 A Schema-Theoretic Model of Language Production

TCG supports the online grammatical processes ensuring the flexible coordination of an incrementally built message and the ongoing production of utterances that reflect whole or part of the current semantic content to be conveyed. Fig. 4.1 presents the integration of this process within the Language Production sub-system of the Schema Architecture Language-Vision InterAction model (SALVIA). In what follow, I will describe the TCG processes as part of the language production system of SALVIA. The suffix “(P)” in fig. 4.1 indicates that the systems are linked to language production. The TCG framework does not assume a priori the symmetry of processes between production and comprehension (see (Barrès and Lee, 2013) for a discussion of the comprehension processes).

The model takes as INPUT the specification of a temporally unfolding message content which incrementally updates the semantic representation hosted by the semantic working memory system (SemanticWM).

The grammatical working memory system (GrammaticalWM) builds on top of the semantic representation by applying the appropriate constructions to build a mapping from meaning to form (Those constructions are retrieved by the CxnRetrieval system from a grammatical knowledge stored in the grammatical long term memory (LTM) system (GrammaticalLTM)). Those two working memory systems hosts time dependent states and processes. The main challenge for TCG is to dynamically and adaptively handle their interactions. The phonological working memory system (PhonologicalWM) simply hosts the current state of word sequences that have already been chosen as the basis for an utterance. Those are posted as OUTPUT. The remainder of the paper details those systems and processes as they relate to TCG.

#### 4.2.2 Schema Theory and Cooperative Computation: What You Need To Know

The full neurocognitive computational framework of Schema Theory (ST) is presented in sec. 1.4 and in Ch. A. For the reader that wished to skip those details, the key tenets of this theoretical foundation of the present modeling work or briefly restated here.

At a cognitive level, schemas represent portions of knowledge (declarative or procedural). They are organized into schema networks that form the state of long term memory systems (LTMs), each defining a type of knowledge over given domain. A LTM is always linked to a Working Memory (WM) in which the knowledge it stores is put to use. Once a schema is deemed relevant to the current state of the computation, it is invoked in WM in the form a schema instance. Each instance represents a hypothesis offering a partial solution to the problem the WM attempts to solve. It carries an activation value that indicates the degree of confidence associated with its hypothesis.

Cooperative computation (C2) fuels WM processes. Instances compete and cooperate, respectively forming inhibitory competition links (comp\_link) and excitatory cooperation links (coop\_link). At each time the whole set instances and C2 links (coop\_links and comp\_links) form a C2 network. The dynamic system it defines governs the temporal trajectories of the instances' activation values. Cooperating instances form assemblages, each corresponding to a potential way to compose instances in order to generate a solution. Schema Theory prescribes that instances corresponding to hypotheses that support each-other engage in cooperation while those that correspond to contradictory hypotheses compete. The precise process through which instances organize into a C2 network however is specific to each WM sub-system. (For more details on the formalism of Schema Theory please refer to sec. 1.4 and Appendix. A).

The use of Cooperative Computation is a core step in building computational cognitive models. It has been shown to adequately capture the known properties of cognitive operations (McClelland, 1993) (see discussion in sec. 1.4).

In what follows, I will present in order: the semantic representation format , the TCG constructions, the process by which the construction instances are invoked, and what governs the creation of competition and cooperation links governing the C2 dynamics.

### 4.3 Incremental and Dynamic Semantic Representation (SemRep)

The rather simple formalism of the SemRep has already been presented in Ch. 2, sec. 2.3.

In SALVIA, a Conceptual LTM (not shown in fig. 4.1) defines a network of concept schemas. It forms a repository for a semantic network model of world knowledge. In the current implementation the semantic network is a hypernym-based (IS\_A) taxonomy. Concept schemas are limited to five types: EVENT, ENTITY, ACTION, PROPERTY, and RELATION.

Beyond concepts, EVENT, ENTITY and ACTION can be used to define FRAMES whose behavior is in line with typical frame semantics. Although for now the world knowledge model does not contain frame knowledge, frames can be used to significantly open the semantic expressiveness of the model. In this chapter the full SemRep representation will be shown including the FRAME nodes which were hidden, for simplification purposes, in the previous chapters.

Conceptual schema instances are invoked in Semantic WM to form a Semantic Representation (SemRep). Since all the conceptual relations are binary, the SemRep is conveniently expressed as a labeled (not necessarily connected) directed graph: edges correspond to RELATION, while nodes correspond to FRAMES, EVENT, ENTITY, ACTION, or PROPERTY concept schema instances. No cooperative computation is implemented within Semantic WM (i.e. the semantic message does not contain any conflict).

The limitation to these classes of semantic concepts derives from the fact that the model is designed as part of (although not limited to) a model of visual scene description in which a core requirement of semantic contents is that it is always tied to incrementally generated perceptual representations. For this reason, elements that can be more directly tied to percepts have been the focus, leaving aside in the process a large chunk of key elements that should enter into the definition of semantic representation. The model could be supplemented with some of those elements (e.g. negation, quantification etc.). But those should be added if a process to link them to some cognitive processes (be it one tied to visuo-attentional processing or other) can be outlined as their support.

As was discussed in previous chapters, at each time step, the SemRep can be updated, modifying the content of the message that has to be expressed (fig. 2.15, 2.16, 2.17). Incrementality takes place both through updating the semantic graph structure and through the activation value dynamics of the conceptual schema instances that compose the SemRep graph.

The goal of grammatical processing using TCG is to generate a flexible grammatical structure articulating the incrementally built SemRep and the production of utterances.

## 4.4 Grammatical Processing

### 4.4.1 Template Construction Grammar: Language Representations as Templates of Meaning-Form Mapping

We propose Template Construction Grammar (TCG) as the basis for a schema theory model of grammatical processing. TCG, as a computational construction grammar, builds on the insights of more complex symbolic models (Embodied Construction Grammar and Fluid Construction Grammar) (Steels, 2011; Feldman, 2010; Bergen and Chang, 2005). TCG however significantly reduces the complexity of the semantic and grammatical representations tackled in order to better focus on the use of the constructions as language schemas engaging in C2.

It is important to keep in mind the distinction between the construction format (defined here) and its use as basis to define language schemas which taken together form the state of grammatical knowledge (described below).

#### Construction

A *construction*  $Cxn$  is defined as the tuple

$$(Class, SemFrame, SynForm, SymLink)$$

where:

- *Class* represents the general grammatical category the construction belongs to.
- *SemFrame* (Semantic Frame) represents the meaning pole.
- *SynForm* (Syntactic Form) represents the form pole.
- *SymLink* (Symbolic Links) represents the symbolic linkages between form and meaning elements.

The elements that compose a construction are defined as follow:

#### Class

*Class* of a construction is defined as a string  $c \in \mathcal{C}$ , where  $\mathcal{C}$  is the set of all possible construction classes. This set is user defined and need not reflect the canonical syntactic classes.

## SemFrame

*SemFrame* is defined as a labeled directed graph  $G = (N, E)$  with  $N = \{N_i\}$  set of labeled nodes, and  $E \subseteq N \times N$  set of labeled edges. Nodes stand for *concepts* while edges stand for *conceptual relations*. This structure of the SemFrame mirrors that of the SemRep.

The SemFrame graph representation is enriched to convey features such as *Head* and *Focus*, which are both boolean values. The node defined as a *Head* defines the semantic class of the construction, in much the same way *Class* defines the grammatical class of the construction. The *Focus* feature carries pragmatic information (information structure (IS)) which is directly co-expressed with the more classically semantic content represented by the SemFrame graph. Following the notion of IS, the Focus node is associated with the information that is pragmatically highlighted by the construction. For example, if a transitive and a passive construction have an almost identical SemFrame, they differ in terms of their Focus (which falls on the agent and on the patient respectively).

## SynForm

*SynForm* is defined as a *list* ( $f_i$ ) such that  $\forall i, f_i \in \Sigma_f$ , where  $\Sigma_f = \text{Word\_forms} \cup \text{Slots}$ .

*Word\_forms* is the set of word forms (or lexical forms), linked to Phonological knowledge of the system (which won't be detailed here.).

*Slots* play a key role as variables (phonologically empty form) that need to be filled by the form of another cooperating. Slots are defined as elements of  $\mathcal{C}^N$  and are noted  $[c_1, c_2, \dots]$  where each  $c_i$  is a construction class. These impose class-based restrictions on the constructions can be provide their missing form content. The list format of the SynForm is essentially equivalent to a representation of syntactic feature limited to *next* temporal relations, where the  $\text{next}(f_1, f_2)$  simply imposes the syntactic constraint that  $f_1$  directly follow  $f_2$  in the temporal unfolding of the utterance. This focus on temporal sequence syntactic constraints is of course limited, but sufficient for our purposes.

## SymLink

*SymLink* (or *SL*) defines a partial mapping between SemFrame nodes and SynForm forms.  $SL : N \in \text{SemFrame} \rightarrow f \in \text{SynForm}$ . As *SL* is partial, not all nodes are mapped onto a form, not all forms have an antecedent in the SemFrame, however, each slot must have a unique antecedent node by *SL*. Symbolic links establish the coupling between semantic and syntactic features. The fact that some nodes might not belong to  $SL^{-1}(\text{SynForm})$  reflects the fact that some constructional semantic information might be packaged in a way that is not directly symbolically mapped onto the form pole (although it is implicitly mapping through the general association of the SemFrame with the SynForm within a construction, cf. IN\_COLOR construction in fig. 2.11). A main simplification made by TCG is that all the SemFrame edges (conceptual relations) are assumed to be symbolically represented in the SynForm (there are no slot/unbound variables symbolically associated with edges in TCG framework).

## Template

SemFrame, SynForm, and SymLinks, taken together form the cxn template defining the meaning-form mapping. Figures 4.2 and 4.3 provide a few example.

## Preference and Group

Two optional features *Preference* and *Group* can be added to the constructions to allow difference in processing treatment. *Preference* defines a scalar value that captures usage preferences (e.g. derived from frequency of use) and during processing modulates the initial activation value of construction schema instances. *Group* defines construction subsets (e.g. lexical and grammatical constructions) that can then be processed differently. There is no a priori limitations to using this grouping. Groupings could be defined based on usage or other empirical results.

From this construction formalism, it is possible to define a Grammar.

## Grammar

A grammar  $\mathcal{G}$  is defined as a set of constructions  $\{Cx_n\}$  (at times referred to as a *constructicon*). Constructions in a grammar can span all levels of linguistic representations (in particular including word-level as well as argument structure level and even multi-clause level constructions). As a construction includes a SemFrame which is defined in terms of concepts, a construction, and by extension the whole grammar, is necessarily defined in relation to a conceptual knowledge.

The model does not impose a particular content for the grammar and offers the option to write and test new grammars using simple json format.

### A Few Examples

Figures 4.2 and 4.3 present a few construction examples illustrating those features. Each construction is assigned a class. If for simplicity the classes used here are similar to the classic syntactic classes, there is no a priori constraint on the number or nature of those classes. Following the main tenets of cognitive linguistics focusing of language in use, linguistic knowledge is not divided into components (phonology, syntax, semantics, and pragmatics), rather any construction can potentially cut across all those strata. For this reason, constructions ran the gamut from lexical constructions (e.g. WOMAN\_1, WOMAN\_2) all the way to argument structure constructions (e.g. PAS\_SVO). Double circle SemFrame nodes mark head nodes. Filled nodes in the SemFrame mark nodes that imply a link to a referent (in the present case, a referent is often a referent in the visual world, but this is not an intrinsic limitation of the system). The link to referents plays a role in triggering co-reference resolution (used in comprehension). Dashed lines indicate symbolic links between form and meaning.

Fig. 4.2 highlights a few “lexical” constructions.

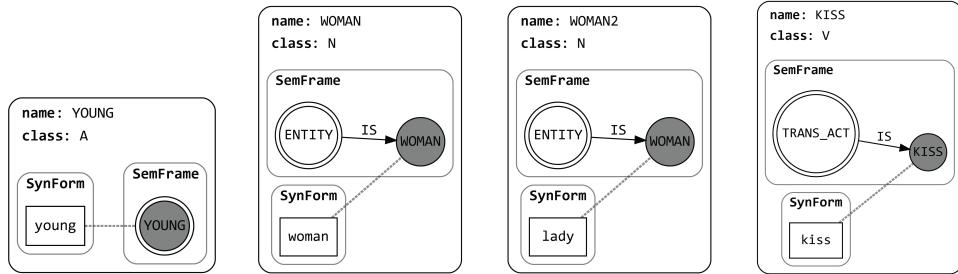


Figure 4.2: Example of “lexical” constructions.

Fig. 4.3 highlights a few abstract constructions ranging from discourse level constructions to noun phrase level constructions.

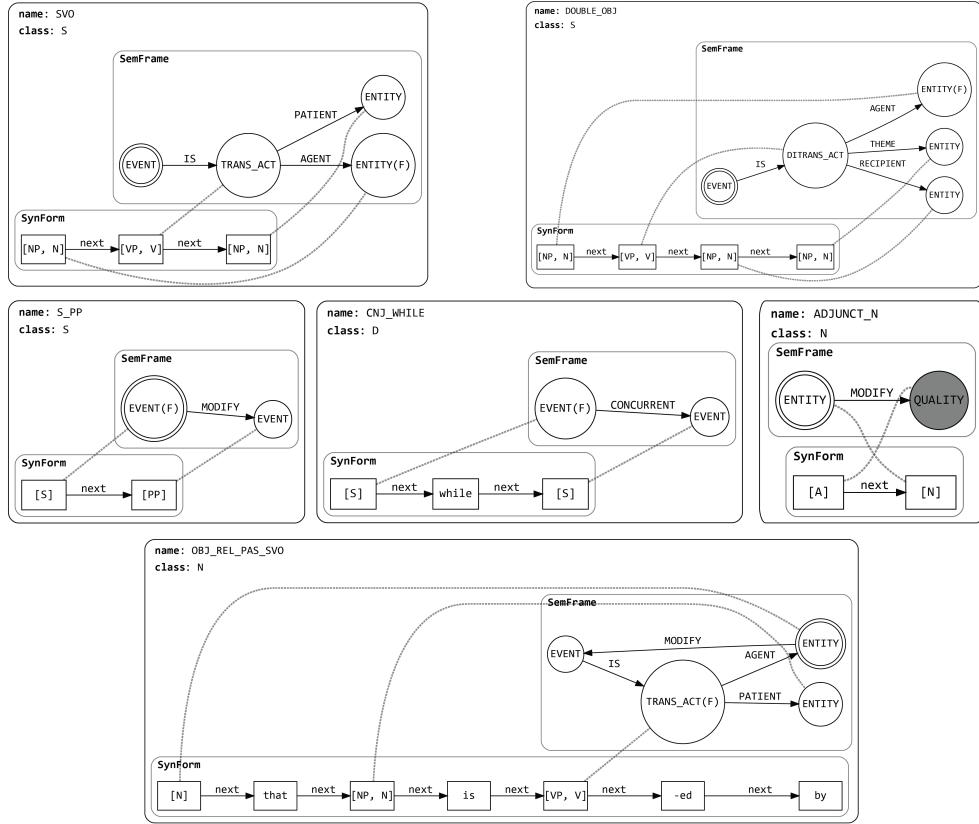


Figure 4.3: Example of abstract constructions

The reader can find in Appendix G the list of most of the abstract constructions that were used.

#### 4.4.2 Language Schemas

##### Construction Schema

A language schema or *construction schema* defines a functional unit of grammatical knowledge. The construction schema is defined as a tuple

$$(Cxn, act^0)$$

where  $Cxn$  is a construction as defined above, and  $act^0 \in [0, 1]$  is a scalar value used to define the initial activation value when an instance of the schema is invoked.

##### Grammatical Knowledge

Although schema theory hypothesizes that long term memory (LTM) should be represented as a schema network, TCG in its current version simply models Grammar LTM as the set of all construction schemas defined based on the grammar:

$$GrammaticalLTM = \{(Cxn_i, act_i^0); Cxn_i \in Grammar\}$$

. Future work will need to follow in the footsteps of Fluid Construction Grammar that has made use of a dynamic priming network to simulate the temporal evolution of the state of grammatical knowledge Wellens and Steels (2011). In the current implementation, for each construction  $Cxn_i$ ,  $act_i^0$  is set based on an LTM defined  $act^0$  and on the construction's preference  $p_i$ :  $act_i^0 = p_i.act^0$ . If no preferences are defined for the constructions, all the preferences are automatically set to 1.

### 4.4.3 Dynamic Grammatical Processing of Incrementally Built Semantic Representations

#### Incremental Instantiation of Construction Schemas

Incrementally, when new SemRep nodes or edges (i.e. conceptual schema instances) are invoked in Semantic WM, constructions schema whose SemFrame semantically match (SemMatch) a SemRep subgraph that contains some new elements are invoked as instances in Grammatical WM (cf. fig. 2.15 and 2.16). A semantic match between a SemRep subgraph and a construction schema SemFrame indicates that the construction expresses in its form, at least in part, the semantic content of this subgraph and is therefore a candidate hypothesis for participating in the mapping of the SemRep onto a linguistic form in Grammatical WM.

There they dynamically interact through competition and cooperation to yield stable assemblages, each representing a possible organization of a mapping between meaning to form.

The goal of the cooperative computation (C2) is to orchestrate the seamless incremental dynamic in which the construction schema instances are engaged in self-organizing and built meaning-form mappings.

#### Grammatical WM State

At each time step, the state of the Grammatical WM is defined by the construction schema instances that are currently active as well as by the cooperation and competition links that they have established and that governs the cooperative computation (fig. 4.5).

Resulting from its invocation into its associated working memory, a construction schema instance is defined as a tuple:

$$(id, Cxn, a(t), covers)$$

where:  $id$  is a unique identifier (since multiple instance 'tokens' can be invoked from the same schema 'type'),  $Cxn \in Grammar$ ,  $a(t) \in \mathbb{R}$  is the activation value, and  $covers : g \in SemRep \rightarrow SemFrame$  is a subgraph isomorphism that links a subgraph of SemRep to the instance's SemFrame.

Following the invocation process (see below), this mapping keeps a trace of the semantic representation that the construction instance translates into a (possibly incomplete) linguistic form template. It also establishes cross WM links (c.f.  $Ext_i$  in Eq. 4.2)

Figure 4.5 provides an example, taken from a simulation, of both the state of the Semantic WM (center graph, concept instances forming a semantic representation - SemRep) and of the Grammatical WM (Construction schema instance graph). It mirrors the informal example that was presented in fig. 2.12. Here the message to be expressed is that of an EVENT involving a single TRANSITIVE ACTION (PUNCH), that in turn involves two participant ENTITY: A WOMAN playing the agent role, and a MAN playing the patient role. Construction schema instances enter in cooperative computation, resulting in a dynamic coordination being put in place between Semantic and Grammatical WM in fig. 4.1. Construction instances attempt to dynamically map the semantic content of the message to convey onto a linguistic form by forming a C2 network. Note that here the cross WM links between a construction schema instance and the concept schema instances it covers are not shown (but the reader can easily look at the SemFrame of the various construction instances and see what part of the SemRep they contribute to map onto form.)

### 4.4.4 Construction Schema Instantiation: SemMatch

Constructions in TCG adopts a coupled structures approach to constructional representational format (SemFrame SynForm linked through SymLinks), approach common to most Computational Construction Grammars (CompCxG) formalisms: ECG and FCG used coupled feature structures one for meaning pole and one for form pole with coupling ensured by co-references of variables across meaning and form pole structures (Chang et al., 2012). In the case of production, the meaning pole defines the part of a whole semantic structure that a construction can express in its form pole.

The guiding principle in designing construction schema processing algorithms consists in the necessity to handle the incremental nature of the semantic representation, an issue that is sidestepped by all the other CompCxG frameworks. Contrary to those other endeavors in which incrementality and online processing is not the focus, an approach in which the search for a solution to the meaning-form mapping operates following

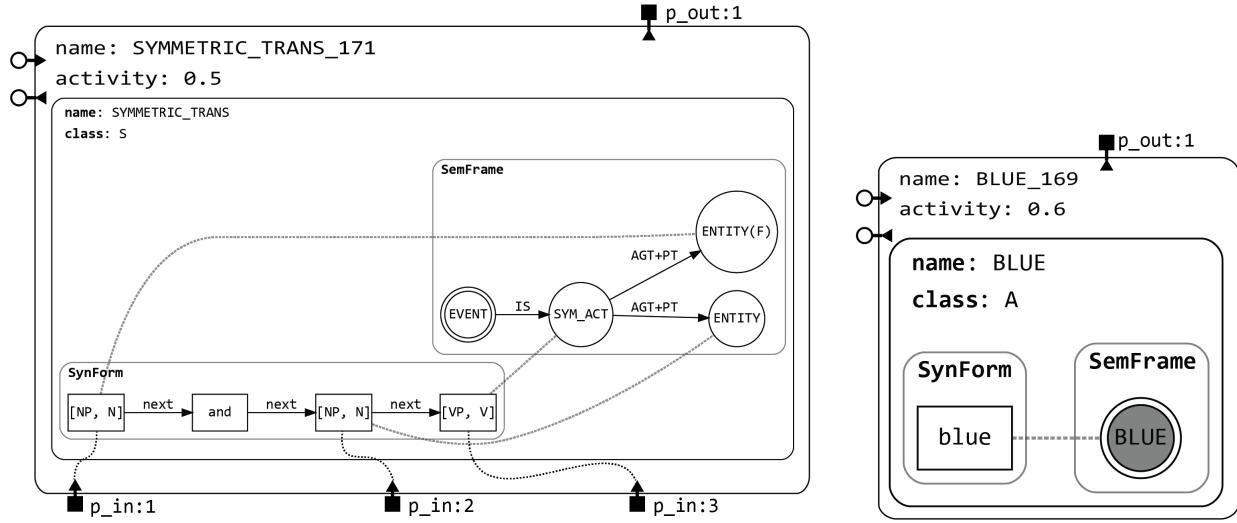


Figure 4.4: Construction schema instances. As for all schema instances (cf. Appendix A, fig. A.2), a construction schema instances is defined by some declarative content derived from a knowledge schema (the construction as defined in TCG format), an activity value  $a(t)$  reflecting the relevance at each time of the construction instance as useful in forming a meaning to form mapping and that is determine through C2 dynamics, a set of input and output ports through which the instances can create cooperative assemblages. Each construction instance is defined by a single output port and one input port for each slot element in the associated cxn SynForm. (Left) Example of construction instance derived from the SYMMETRIC\_TRANS construction (involved in sentence structure such as “the man and the woman shake hands.” (involve symmetric action). The instances has three input port corresponding to the 3 slots of its SynForm. Those will serve to build cooperation links with other instances that can provide the constructional content necessary to fill the slots. It has one output port that can cooperate and link to multiple input ports of construction instance that use its template to fill in their missing details. The two ports on the left side correspond to activity input and output port. Here the activity of the construction is 0.5 and reflects its relevance to the current grammatical processes. (Right) Example of a simple lexical construction instance derived from the BLUE construction. Note that this construction instance does not have any input port since it does not require inputs from other construction instances to fill in its details but only links, across WM to the SemRep instances.

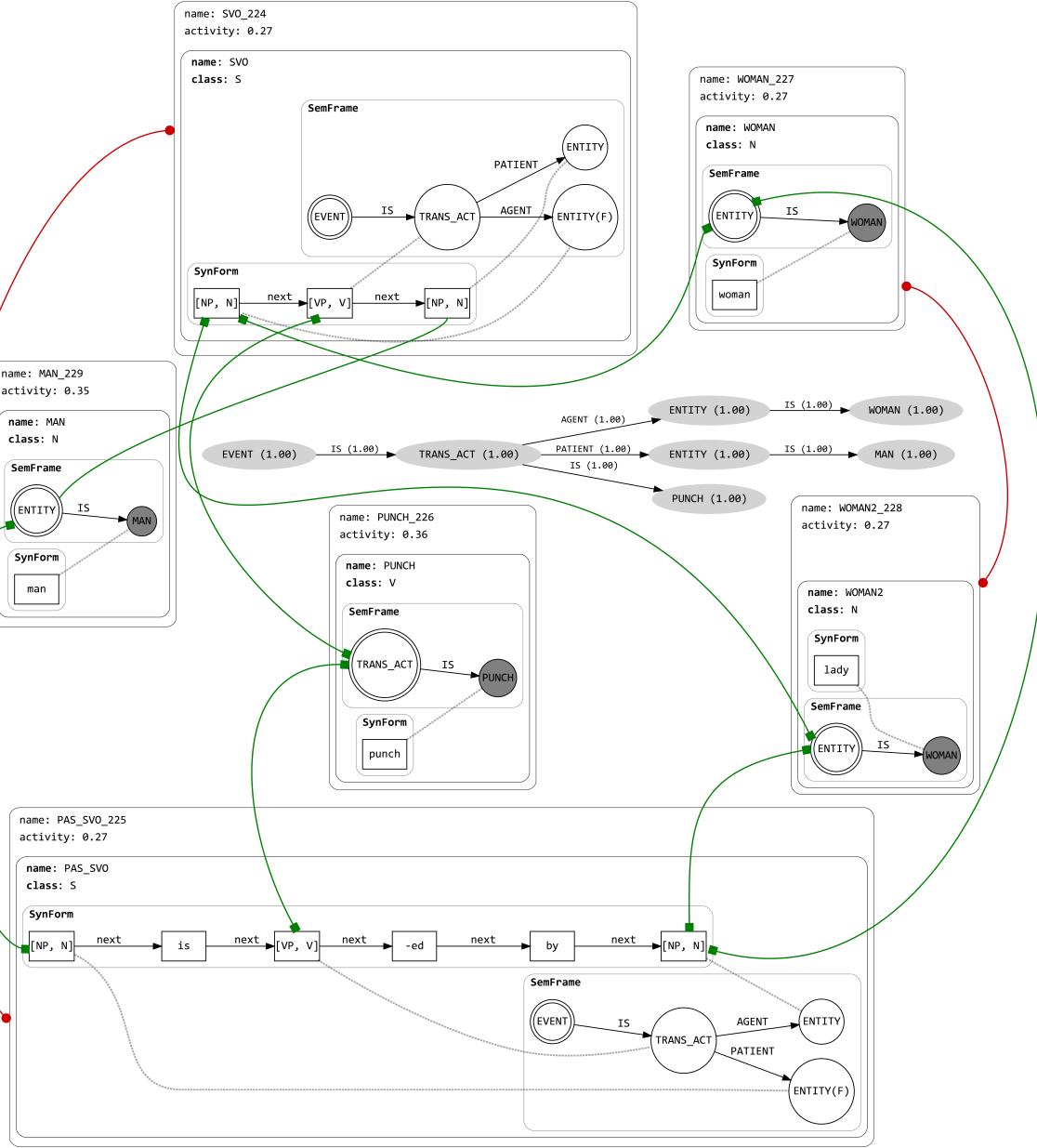


Figure 4.5: Dynamic coordination between Semantic and Grammatical WM. Concept schema instances form a semantic representation graph (SemRep) at the center (Semantic WM's state). Construction schema instances (boxes) are shown forming C2 network (Grammatical WM's state, green cooperation, red competition). The dashed lines linking constructions to SemRep represent the portion of the SemRep for which each construction provides a partial meaning-to-form mapping hypothesis. This figure presented the simulated output of the system corresponding that mirrors the informal example shown in fig. 2.12

a search tree algorithm cannot be used, it is ill suited to model the adaptive coupling of the Grammatical WM's state to the time dependent state of the semantic representations in Semantic WM: Each modification of the SemanticWM state would require updating the whole search tree. Moreover a tree search approach does not fit what is known of the principles that support cognitive processes (captured by Schema Theory).

SemMatch adapts to the TCG focus on incrementality one of the key operations that used by formalisms defining the processing going from an input to an output as the accumulation of knowledge sources forming a network of constraints and in particular by unification based language formalism Shieber (2003) (and for CompCxG FCG, ECG<sup>1</sup> fall in this category). All of those systems need to define a unification test that, upon attempting to apply a new source of knowledge, returns the set of variable bindings for which a legal unification can be performed (Knight, 1989). In TCG however, this operations governs construction schema instantiation.

If the SemFrame of a construction schema is considered to be a network of free variables with constraints (on their domain (concept) and their relations (shape of graph)), then SemMatch essentially consists in finding, the set of valid bindings of those variables onto the SemRep values. This is implemented as finding a labeled sub-graph isomorphism.

A construction schema is invoked as an instance in grammatical working memory (GrammaticalWM) each time its semantic pole (SemFrame) matches a part of the SemRep. Such matching indicates that the construction stored in long term memory is relevant to the ongoing process of building a linguistic expression of the message.

Invocation can be divided into two sub-processes:

- *SemMatch* process checks the applicability of a construction schema as a meaning-to-form mapping given a SemRep. Given the graph structure used as a basis for semantic representations, *SemMatch* is defined as a sub-graph matching algorithm with additional label matching. Labels correspond to concepts and their semantic matching is tied to the associated model of conceptual knowledge. In fig. 4.6, the shape of the PAS\_SVO construction schema's SemFrame graph matches the SemRep graph of concept schema instances active in Semantic WM (in general a construction schema's SemFrame only matches a sub-graph of the SemRep): they are isomorphic and their labels match. The labels of the SemRep are equal or hyponyms of their corresponding labels in the SemFrame, e.g. HUMAN is a hyponym of ENTITY given the model of world knowledge. Given a construction schema, *SemMatch* returns the set of all the sub-graph isomorphisms from the SemFrame of the construction to the SemRep (the SemFrame of a construction schema can match multiple SemRep's sub-graphs). The SemMatch can be also expanded so that it returns, in addition to returning a Boolean value reflecting the applicability or not of a construction, an additional continuous reflecting the quality of each matching (it is done by defining a distance metric on the labels). (See Appendix.C, alg. 2)
- *instantiate\_cxn* uses *SemMatch* output to generate the construction instances that will be invoked in GrammaticalWM: given a construction schema and SemRep, each sub-graph isomorphism generated by SemMatch results in the invocation of a construction instance (c.f. fig. 4.6). The initial activation value of the instance is defined by the schema in LTM but can be modulated by factors such as preference (see above). (See Appendix.C, alg. 3)

Search for sub-graph isomorphisms is known to be quite a costly operation. However, the relative small size of the graphs the model is dealing with ensures that the problem remains tractable (Typically, at each time a SemRep contains less than a dozen nodes and SemFrames are usually limited to a few nodes.).

In fig. 4.1 the construction invocation process is handled by the construction retrieval sub-system (Cxn-Retrieval(P)).

The semantic match (*SemMatch*) requirement functions as a filter. The Conceptual LTM is represented as an input to *SemMatch* since the conceptual knowledge is necessary for one of the two *SemMatch* step. A match between a SemRep subgraph (S) and a SemFrame requires first graph isomorphy, but also matching between the corresponding conceptual content of the nodes and edges of S and SemFrame. This latter step requires access to conceptual knowledge since, for example, it allows for semantic matching between a concept and its hypernyms, etc.

---

<sup>1</sup>ECG does not handle production

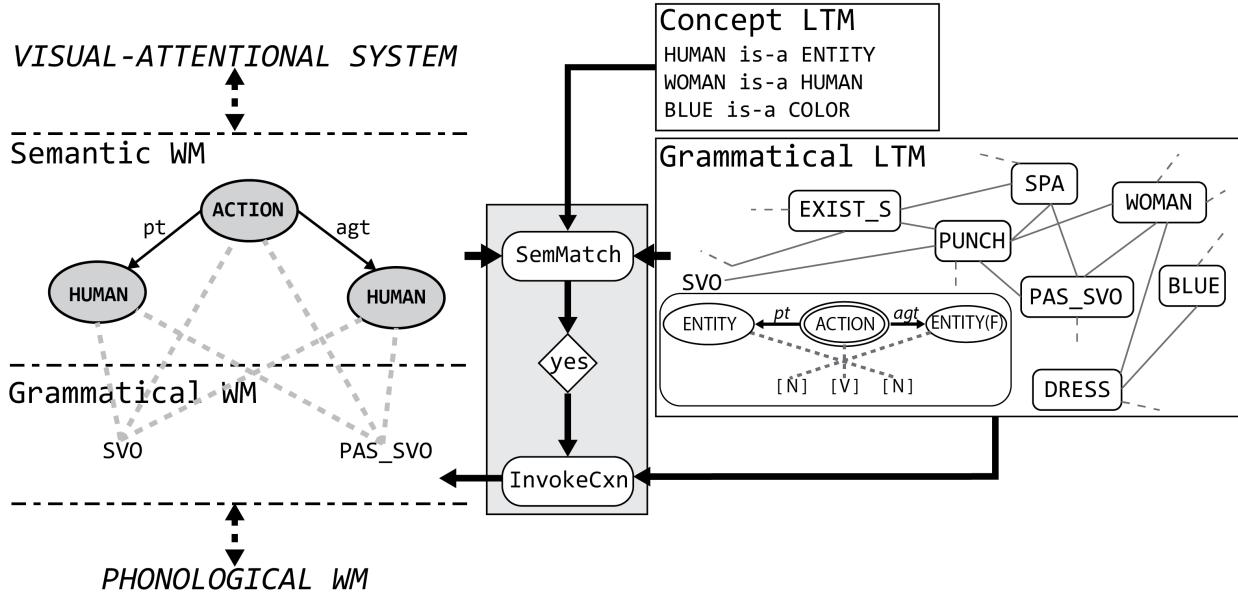


Figure 4.6: Construction instantiation: SemMatch process. High level view of the construction instance retrieval and invocation process. In this simplified view, the SVO and PAS\_SVO construction's SemFrames are isomorphic to the SemRep graph (ENTITY is a hypernym of HUMAN). Those isomorphisms are determined by SemMatch which results then in the invocation in GrammaticalWM of an instance of each of those construction schemas, each covering the SemRep graph it is isomorphic to (in this case, the whole graph, dashed grey lines, coverage of edges is omitted for clarity).

Using *Group* features, the system can require that certain constructions be invoked first (e.g. lexical construction before argument structure constructions).

In fig. 4.1 the construction invocation process is handled by the construction retrieval sub-system (Cxn-Retrieval(P)).

#### 4.4.5 Cooperative Computation (C2): Match

TCG needs an algorithm to decide, based on the state of the Linguistic WM (SemanticWM: SemRep + Grammatical WM: Set of active construction instances) how a newly invoked construction schema instance can contribute its meaning-form mapping hypothesis, entering in cooperation with other construction instances to improve the existing meaning-form mappings. As it rests on a self-organizing decision process (search), TCG also needs to build competitions between construction schema instances that carry incompatible hypothesis.<sup>2</sup>

Match structures the relations between construction instances in a way that allows the system to bypass tree search and adopt the C2 approach to cognitive modeling: Match establishes cooperation links where cooperation is possible, and competition links where instances hypotheses are incompatible, a situation which, in a search tree, would result in branching.

The constructional information carried by active instances is not merged onto the Linguistic WM state<sup>3</sup>. The possibility of cooperations are symbolized by the cooperation links, while the incompatibilities (that would trigger a branching in the search tree) are symbolized by competition links.

The goal of the Grammatical WM consists in incrementally building mappings to express the semantic content of the SemRep (itself built incrementally) in a linguistic form. Construction schemas that correspond

<sup>2</sup>The state of both the Grammatical and the Semantic WM are involved at this stage since for each construction instance, the free variables of their SemFrame of are now bound (to the SemRep subgraph they cover). Match can also be seen as an adaptation of a unification test this time applied between a construction instance and the state of the Linguistic WM.

<sup>3</sup>In this TCG differs from Fluid Construction Grammar

to relevant meaning-form mapping hypotheses are invoked in Grammatical WM (see above) where they enter in cooperative computation (C2).

Each construction instance active in GrammaticalWM carries a mapping hypothesis of a portion of the current semantic representation onto a linguistic form. Cooperation emerges between two constructions whose mapping can be composed to generate a new mapping covering a larger portion of the semantic content, or refining the mapping. Competition, on the other hand, is triggered when two constructions represent incompatible mapping hypotheses.

Each construction instance carries an activation value, whose initial value is modulated by the preference value stored in the schema, representing the idiosyncratic usage preferences of the speaker (to which can be added a factor reflecting the quality of the semantic match). They organize into a C2 network, whose dynamics defines at each time step the values of the instances activation values. If a construction instances activation value falls below a given threshold, the instance is pruned out of the Grammatical WM. The C2 network is therefore intermittently reshaped following either the invocation of new constructions instances or the pruning of construction instances that lost' the competitions in which they were involved.

C2 links are built based on the Match operation. Two instances that do not overlap in their coverage of the SemRep do not form any C2 link. Informally, if two instances overlap in their SemRep coverage, the core constraint is that one of the constructions (child) needs to provide a SynForm that can (partially) fill in the missing form information of the other construction (parent).

If a construction instance  $C_i$  is defined as a mapping from a subgraphs  $S_i \subset SemRep$  onto a linguistic form  $f_i = C_i.SynForm$ , only if two constructions instances  $C_1$  and  $C_2$  are in such relation that  $S_1 \cap S_2 \neq \emptyset$  will they enter in cooperation or in competition since they overlap on the semantic content they map onto linguistic form. The cooperation and competition process therefore only need to concern itself with construction instances whose meaning poles map overlapping subgraphs of the semantic representation.

Given two constructions instances  $C_1$  and  $C_2$  with their overlap  $O = S_1 \cap S_2 \neq \emptyset$ :

- If  $O$  contains a semantic relation (i.e. an edge in the semantic graph) then the constructions necessarily compete. This stems from a particular limitation of the TCG formalism: a given construction always symbolically fully map the semantic relations it covers onto its syntactic form. The mapping is implicit (the symbolic links only map SemFrame nodes onto the SynForm), and the use of slot in the SynForm (thought of as unbound form variables) only allows constructions to left under-determined the formalization of conceptual entities and not relations<sup>4</sup>.
- If  $O$  only contains semantic entities (i.e. nodes in the semantic graphs). In this case, in the current formalisms, two cases emerge.
  - Both constructions lexicalize this node (as a phonetic form) in which case the constructions compete.
  - At least one construction does not lexicalize the node (i.e. only map the node onto a slot), in which case the construction will either cooperate or not form any link. The conditions under which cooperation occur are detailed below.

Again, the distinction between the two cases reveals the asymmetry in treatment of semantic entities and relations (nodes and edges) in TCG. A SemFrame semantic relation (edge) is necessarily translated into the form of this construction. A SemFrame semantic entity (node) however, may or may not be lexicalized as a phonological form by the construction, as reflected by the fact that the SynForm can include respectively *slots* or *word\_forms*.

### Matching constraints

Figure4.7 shows portions of two constructions,  $CNX_1$  and  $CNX_2$  that overlap on a semantic node  $N_{sem}$ . Assuming that those constructions are not already in competition (condition (0), i.e. they do not overlap also on a semantic relation (see above)), we can define the syntactic and semantic constraints that will govern the potential creation of a cooperation link (shown here potentially linking  $N_2$  to  $F_1$  (for simplicity only the case of  $CNX_1$  as parent and of  $CNX_2$  as child is shown). Syntactic and semantic constraints each contain a type and a Boolean constraint.

---

<sup>4</sup>Further development of TCG could lift this constraint.

## Syntactic Constraints

(Syn1) states that link should target a form of type *Slot* in the parent construction, i.e. that the parent construction does have a symbolic link mapping the semantic node to the syntactic form and that the parent construction does not define a phonological form for the semantic node.

(Syn2) states that the child construction class needs to be included in the set of classes that can be associated with the slot in the parent construction.

## Semantic Constraints

(Sem1) states that the link should originate from a SemFrame node of type *HEAD* in the child construction<sup>5</sup>.

(Sem2) states that the concept of the child's semantic node fits the semantic requirements of the parent construction's semantic node. Currently, the fit is decided based on whether or not the concept type of the child semantic node is subsumed by the concept type of the parent's semantic node. However, this is a simplification that serves as a binary proxy for a constraint that should take into account the distance between child and parent concepts.

(Sem2) is the semantic counterpart to the (Syn2) constraint. While Syn2 imposes that the child's construction class fits the parent's construction's slot class constraints, Sem2 imposes that the concept of the child semantic node fits the semantic requirements of the parent semantic node. For a child construction to be able to contribute its content to a parent construction by linking to one of its slot, there are both syntactic class and semantic constraints. This reflects an important aspect of construction grammar: syntax and semantics are not disjoint. The example of the verb alternation contrast between the *Theme-object construction* “**X V Y in/on Z**” (Cxn1) and the *Goal-object construction* “**X V Z with Y**” (Cxn2) illustrates this point. They are both abstract argument structure constructions but their semantic requirements for the verb slot V vary.

- (1) “John pours water in the plant”
- (2) \*“John pours the glass with water”
- (1') \*“John filled water in the glass”
- (2') “John filled the glass with water”

Cxn1 accepts “pour” but not “fill” in its verb slot, while the reverse is true for Cxn2. Both share the same syntactic class, and are very close in term of general conceptual/world knowledge. But they differ in the fact that “pour” only specifies the path to the object and not the change of state of the object (if the glass has a hole at the bottom, one can pour water in it and not fill it, or one can pour water in an already full glass), “fill” only specifies the change of state, not the manner by which it is achieved (one can fill a glass with water by dipping it into a pool of water or by pouring water into it). This difference between the two concepts “pour” and “fill” reveals the difference in semantic constraints associated to the V slot in Cxn1 and Cxn2.

One could consider the class of a construction plus the concept of its HEAD to form a syntactico-semantic class for the construction that is the one that matters when the construction is used in relation to other construction.

I keep the separation between rather syntactic-like class and the semantic HEAD SemFrame node concept to clearly indicate the importance of construction based semantic constraints on the building of construction assemblages (for a slot, the semantic constraint comes from the SemFrame node it is linked to). I have discussed elsewhere how constructional semantics can be participate in “light semantic” operations (as opposed to “heavy semantics” operations when the conceptual knowledge is put to full use) that should be dissociated, from a neurolinguistics perspective and in the case of comprehension, from more classically syntax like operations (Barrès and Lee, 2013).

---

<sup>5</sup>The HEAD node in a construction represent the semantic type of the construction. For now the implementation only allows for endocentric construction, but the inclusion of exocentric constructions is being investigated

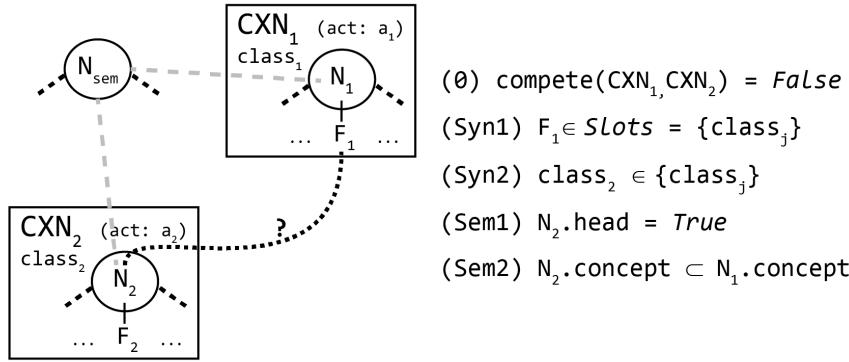


Figure 4.7: Syntactic and semantic constraints on match yielding a cooperation link. Syn1 and Sem1 are obligatory constraints (type constraints) while Syn2 and Sem2 are qualitative constraints. ((Syn1) includes the assumption that a symbolic link exists between  $N_1$  and the SynForm of  $CXN_1$ .)

Using the definition of  $d_{syn}$  and  $d_{sem}$  below, (syn2) can be restated as  $d_{syn}(cxn_2, F_1) > 0$  and (sem2) can be restated as  $d_{sem}(N_2, N_1) > 0$ . This formulation highlights the fact that future implementation of TCG will move away from (syn2) and (sem2) as obligatory constraints and consider those as qualitative constraints where  $d_{syn}$  and  $d_{sem}$  will be treated as more classic distance functions, and will return analogical values reflecting the quality of the match between syntactic and semantic features. These metrics will then be used to impact the strength of the cooperative links created between constructions, allowing for finer grain grammatical dynamics.

$$d_{syn}(cxn, slot) = \begin{cases} 1 & cxn.class \in slot.classes \\ 0 & \text{otherwise} \end{cases}$$

$$d_{sem}(N_1, N_2) = \begin{cases} 1 & N_1.concept \subseteq N_2.concept \\ 0 & \text{otherwise} \end{cases}$$

*match* returns a categorical value *match\_cat* defining the nature of the relation between the mapping hypotheses put forward by *inst<sub>1</sub>* and *inst<sub>2</sub>* respectively<sup>6</sup> (See Appendix.C, alg.4).

$$match\_cat = \begin{cases} 1 \equiv \text{COOPERATION} \\ 0 \equiv \text{NO RELATION} \\ -1 \equiv \text{COMPETITION} \end{cases}$$

In the case of cooperation, *match* also returns the relevant data to generate the cooperation links between the two instances, as well as, for each link a *match\_qual* value that reflects the result of constraints (Syn2) and (Sem2). As mentioned above, for now those are categorical constraints, but the system is already set up to handle them as qualitative constraints.

The *comp\_link* algorithm is called by *match* to check the competition related constraints for a given point of semantic overlap between two construction instances. There is competition if both construction instances propose a phonetic form for the point of semantic overlap. Importantly, in TCG there is the possibility for a SemFrame node not to be symbolically link to any element of the SynForm. In this case, the node is considered to be implicitly fully mapped onto the form (e.g. in the IN\_COLOR construction). (See Appendix.C, alg. 5). Note that the case of an overlap on an edge (which leads to competition) is, for reason of simplicity, directly handled by the *match* algorithm.

The *coop\_link* algorithm is called by *match* to check the cooperation related syntactic and semantic constraints for a given point of semantic overlap between two construction instances. If a link is found, it returns both the link as well as the *match\_qual* value. (See Appendix.C, alg. 6).

<sup>6</sup>Note that  $match(inst_1, inst_2) \equiv match(inst_2, inst_1)$ .

The SemMatch process is exemplified in Figure 4.8. The bottom example shows two constructions that overlap on the WOMAN SemRep node. However, in PAS\_SVO, the SemFrame node that covers WOMAN (ENTITY) is linked to a slot and therefore linguistic information to express the semantic content of WOMAN is missing. WOMAN\_1 also covers the WOMAN SemRep node. In addition, it can serve to fill in the slot in PAS\_SVO since: WOMAN\_1 has a class that matches the class requirement of the slot (N) (SynForm requirement), WOMAN node in the SemFrame of WOMAN\_1 is semantically compatible with the ENTITY node in PAS\_SVO (SemFrame requirement) to which the slot is symbolically linked, and finally WOMAN in the SemFrame of WOMAN\_1 is a HEAD node. Match therefore results in the creation of a cooperation link between the two constructions (green) that link WOMAN\_1 to PAS\_SVO through the relevant slot for which WOMAN\_1 provides the missing phonological content.

The top-left figure presents a situation largely similar to the previous one with the exception that in this case, both constructions associate the SemFrame node that covers WOMAN with a phonological form (and not a slot). Here, this represents the case of two synonymous lexical item competing to express a concept<sup>7</sup>. In this case Match creates a competition link between the two constructions instances.

The last example, on the top-right, presents the case of two argument structure constructions that overlap on a subgraph and not only on a single node. This necessarily results in competition since, as we have mentioned above, it is implicit in the formalism of TCG that edges of the SemFrame are symbolically represented in the SynForm (i.e. there is not equivalent of slot” variables for edges).

The cooperative computation dynamic within the Grammatical WM is that prescribed by Schema Theory as described above. External activation is received by the construction instances from the Semantic WM on the basis of the cross WM links defined above (see above, fig.4.4).

As an example, in Figure 4.5 WOMAN\_227 (“woman”) and WOMAN\_228 (“lady”) construction schema instances compete as synonymous lexical constructions. At the more abstract level of argument structure/voice: PAS\_SVO\_225 and SVO\_224 instances compete as they both build on top of the same portion of the SemRep but express the agent-patient semantic roles in different ways in their SynForm.

Each construction instance active in Grammatical WM carries a mapping hypothesis of a portion of the current semantic representation onto a linguistic form. Cooperation emerges between two constructions whose mapping can be composed to generate a new mapping covering a larger portion of the semantic content, or refining the mapping. Competition, on the other hand, is triggered when two constructions represent incompatible mapping hypotheses.

C2 links are created incrementally: each time a new construction instance is invoked it is matched against the ones that are already active in the Grammatical WM. As it is shown in ch. 2) there exists a deep relation between incrementality at the visuo-attentional, semantic, and grammatical level (see fig. 2.15 and fig. 2.16).

Construction schema instances that represent incompatible meaning-form mapping hypotheses enter in competition through mutually inhibitory links. Two construction schema instances that can cooperate, each serving as a context favoring the use of the other, enter in cooperation through excitatory links. The links therefore reflect rule like symbolic processes that have defined the relationship between construction schema instances (see sec. C.1.2).

## C2 dynamics

We build on the theory of cognitive level hybrid models that have operationalized symbolic processes or representations manipulations using dynamical systems, in particular using (localist) cognitive networks (Graded Random Adaptive Interactive (non-linear) Networks (GRAIN) (McClelland, 1993) and/or Parallel Distributed Processing (PDP) approaches (Rumelhart and McClelland, 1986)), or Cooper & Shallice’s schema theory (a different version of schema theory focusing on the organization of motor control)(Cooper and Shallice, 2000, 2006; Cooper et al., 2005). As mentioned above and in the previous chapter (see sec. 1.4 and Ch. A), the general cooperative computation approach follows the McClelland’s principles of cognitive theory. The main difference is that in our Schema Theory setting, the C2 network is not predefined but change topology as new instances are invoked while some are pruned out. However, the justification put forward by McClelland for the modeling of cognitive dynamics remain valid.

---

<sup>7</sup>Although it can be claimed that no two constructions are synonymous (principle of no synonymy (Goldberg, 1995)), this only holds for an idealized speaker, not at the level of performance of individual speakers that we consider.

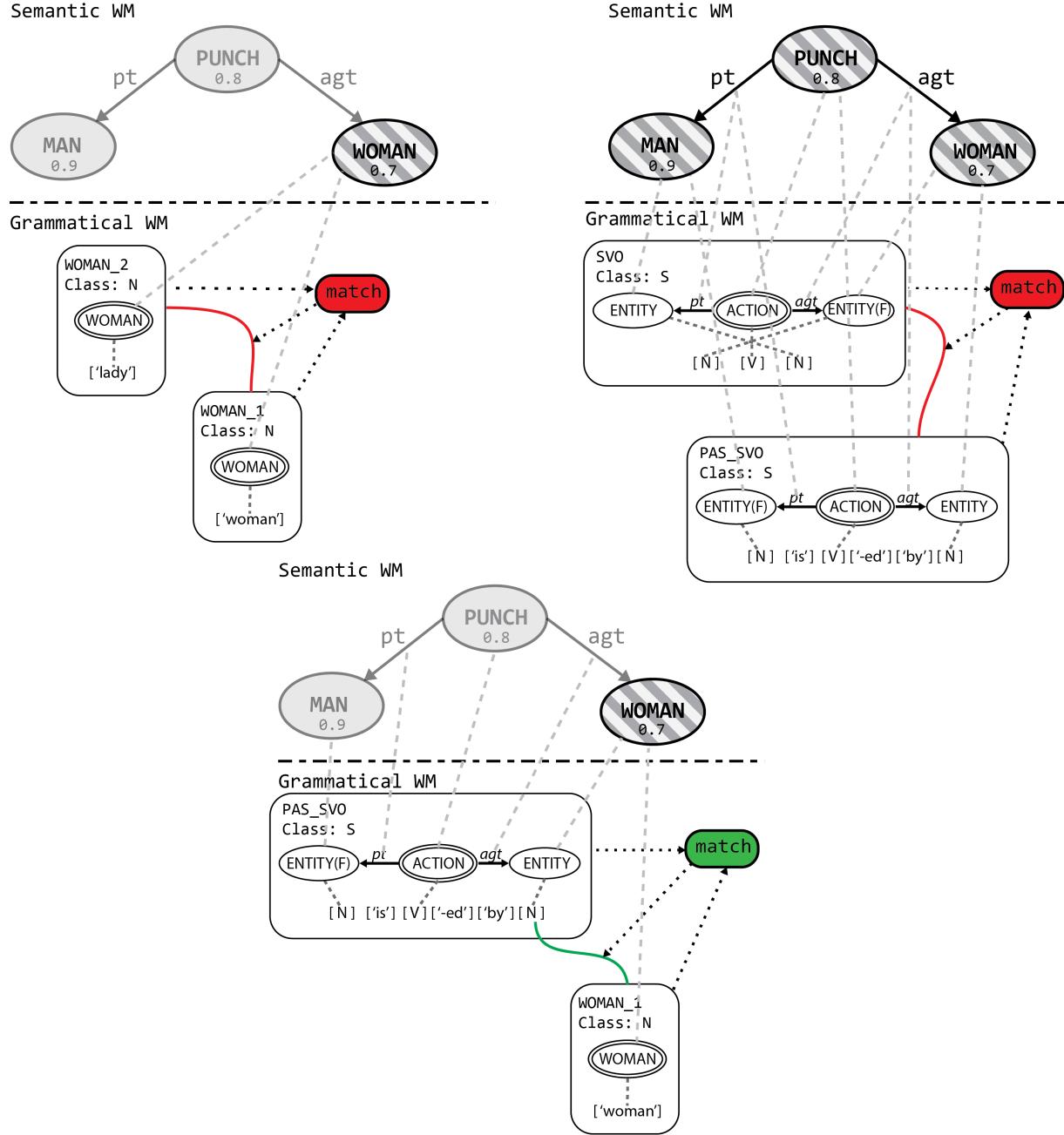


Figure 4.8: Examples of matching outcomes between construction instances. Highlighted part of the SemRep are covered by both constructions. Dashed lines across WM indicates the relations between the constructions' SemFrames and the SemRep. The Match process takes two constructions as input and generates as output either a cooperation link (green) or a competition link (red). The case in which no link is created is not shown (case in which constructions express subgraphs). (See main text for details)

All schema instances have continuous valued activation levels. They form a dynamic system of interacting activations, each instance functioning as a leaky integrator, building on the Leaky Competing Accumulator (LCA) model (Usher and McClelland, 2001, 2004; Bogacz et al., 2007; Tsetsos et al., 2011).

Cooperation links are excitatory while competition links are inhibitory. The weight of those respective links is fixed and defined by the Grammatical WM. This type of competition has been often used to perform “contrast enhancement” (Grossberg, 1982), favoring the dynamic emergence of a “figure” on a “background”, the figure here being a construction instance assemblage that fulfill the goal of the Grammatical WM. It suppresses instances with weak activities and the amount of inhibition each instance receives reflects how much alternatives are supported (relative weighing process).

In the current implementation, competition links are always symmetric (bidirectional with similar weights) while the cooperation links are asymmetric (unidirectional).

The sigmoid function  $\sigma$  is used as non-linearity with its shape fixed by  $\sigma(-\infty) = 0$ ,  $\sigma(\infty) = 1$ , and two parameters:  $\sigma(0)$  that defines the activity of the schema instance at rest, in absence of any input, and  $\sigma(0)'$  that can control the steepness of the sigmoid (influencing the sensitivity of the activity to small inputs). The full equations are given in Appendix.

The convergence of the C2 dynamic towards a single solution (or a stable state in the case of a stochastic system) cannot be guarantee since the system is not necessarily symmetrical in its connections and the topology of the network changes<sup>8</sup>.

For a construction schema instance  $i$ , active in a WM as part of C2 network, its activity  $Act_i^t$  is updated following a leaky integrator equation:

$$Act_i^{t+1} = \alpha Act_i^t + (1 - \alpha)\sigma(Input_i^t + noise^t) \quad (4.1)$$

with  $\alpha$  defining the characteristic time of the WM system  $\alpha = (1 - \tau^{-1})$ ,  $\sigma$  the logistic function, and with a Gaussian noise  $noise^t \sim \mathcal{N}(0, noise_{std})$

$Input_i^t$  is defined as:

$$Input_i^t = w_I \left\{ \sum_{k \in comp(i,k)} w_{comp} \cdot Act_k^t + \sum_{j \in coop(i,j)} w_{coop} \cdot Act_j^t \right\} + \sum_{e \in ext(i)} w_e \cdot Ext_{(e,i)}^t \quad (4.2)$$

$Ext_{(e,i)}^t$  represents activation that an instance  $i$  receives from outside the working memory by subsystem  $e$ .

Here, the competition, cooperation, and external weights are taken to be the same for all instances within a WM<sup>9</sup>.

$w_{ext}$  balances the strength of internal and external activation inputs.  $w_{comp}$  balances the strength of competition and cooperation. The parameters of the logistic function  $\sigma$  are chosen so that, in addition to  $\sigma(\infty) = 1$  and  $\sigma(-\infty) = 0$ ,  $\sigma(0) = Act_{rest}$  the activity in the absence of input. The remaining degree of freedom can be used to set  $\sigma(x_0)' = \sigma_0'$  in order to define the steepness of the logistic function. In this case the dynamics of the leaky integrator is defined by the parameters  $(\alpha, A_{rest}, \sigma_0', w_{comp}, \{w_e\}, noise_{std})$ . In addition,  $\theta_{prune}$  defines the pruning threshold. A constructions whose activation values falls below  $\theta_{prune}$  is pruned out of working memory. Each WM system sets its own set of parameters.

## Construction Instances Assemblage

Through the process of competition and cooperation, construction instances generate construction assemblages, each representing a potential (possibly partial) self-organized program to translate the message (SemRep) into a phonological form.

At each time step, a constructions assemblage  $A$  can be defined as a set of cooperating construction instances  $A = (Insts, Coop\_Links, act)$ . The hypothetical meaning-to-form mapping it represents is associated with its own activation value  $act$  derived from that of the assemblage component instances and that reflects its relevance as a meaning-form mapping solution. In the current implementation  $act = mean(inst_i.act), inst_i \in Insts$ . We take the mean rather than the sum of the activation values of the constituent instances in order

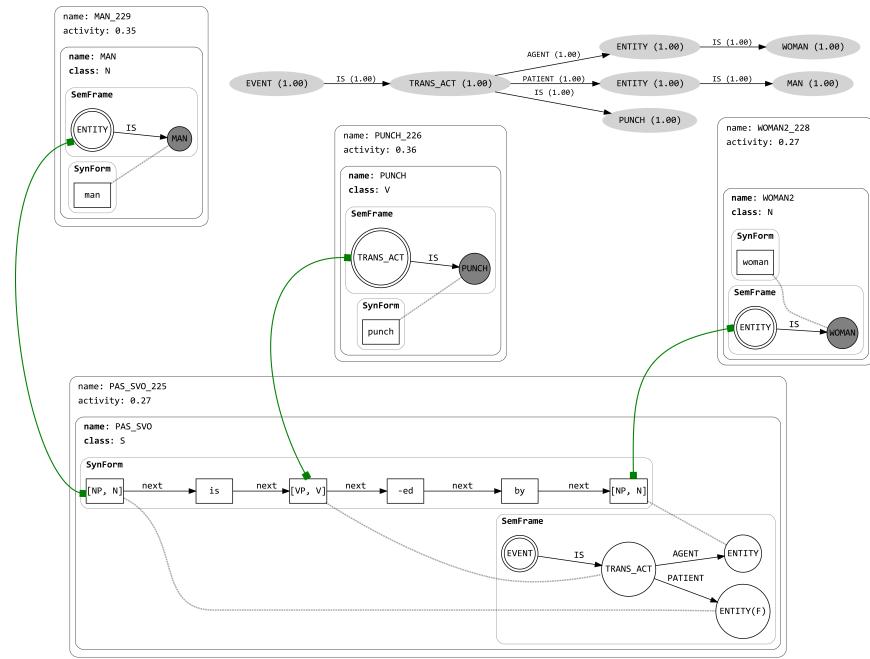


Figure 4.9: State of the Linguistic WM (SemRep graph, top-right) and Grammatical WM (construction instance competition-cooperation graph) corresponding to a later computational state than the one described in fig. 4.5. Here there are no more competition links between construction instances (all the competitions ended with the losing instances being pruned out of the Grammatical WM.) Only a single cooperative assemblage of construction instances remain that maps the SemRep onto the linguistic form “man is punh-ed by woman” once the SynForm of the lexical constructions are used to “fill in” the missing information in the slots they are associated with in the argument structure passive construction instance (PAS\_SVO).

not to favor assemblages that contain more instances.

Looking back at fig. 4.5 it appears that lexical constructions WOMAN and WOMAN2 compete as synonymous lexical constructions. At the more abstract level of argument structure/voice: PAS\_SVO and SVO compete as they both build on top of the same portion of the SemRep but express the agent-patient semantic roles in different ways in their SynForm. At the early point in time shown in fig. 4.5, the competitions have not yet impacted the relative activation values of the construction instances. Fig. 4.9 display the WM states once all the competitions have resulted in one of the competing instance emerging as the winner. SVO and WOMAN\_2 lost the competition and have been pruned out, we are left with a single construction instance assemblages, corresponding to the use of the passive voice and the lexicalization of WOMAN as “woman”.

It is worth noting that the system’s dynamics does not guarantee convergence, and in particular does not guarantee convergence to a state without competition or convergence to a state that is only composed of a single assemblage. If multiple options remain, when forced to choose, the system employs a winner-take-all strategy to select the assemblage with the highest score .

(see also Appendix.C, fig. C.1)

### Assemblage Unification

Cooperation links within an assemblage define unification links through which the coupled semantic and syntactic constraints defined by construction instances can be unified<sup>10</sup>.

An important consequence of the relation between cooperation links and unification is that an equivalent construction instance can be associated to any construction assemblage. That is to say that a construction assemblage can be itself considered as a construction instance, mapping meaning onto form. Indeed, for an assemblage  $A = (Insts, Coop\_Links, act, score)$ , the equivalent instance  $eq\_inst_A$  is defined as:

$$eq\_inst_A = \bigsqcup_{L \in Coop\_Links} (L_{parent}, L_{child})$$

where  $\bigsqcup$  here denotes the unification operation, and where  $L_{child}$  and  $L_{parent}$  correspond to the instances connected by the cooperation link L.

We will note  $\sqcup^{eq}()$  the operation that generates the unification based equivalent instance from an assemblage  $A$  i.e.

$$eq\_inst_A = \sqcup^{eq}(A)$$

Figures 4.10, 4.11 and 4.12, show unification steps based on the example of the final winning assemblage presented in fig. 4.9. Those provide a simulation based counterpart to the informal examples presented in ch. 2, fig. 2.18 (although here the final state considered is slightly simpler than in the informal example since the SemRep considered does not include the YOUNG modifier for WOMAN.)

#### 4.4.6 Generating Form

When an utterance forms is to be generated, the winner assemblage is selected, the constructions instances are unified and the form of the resulting meaning-form mapping correspond to the form that is sent to the Phonological WM as the basis for generating the utterance (cf. fig. 4.12).

When the system is required to generate an utterance, the winner assemblage is selected, the constructions instances are unified, and the form of the resulting meaning-form mapping is sent to the Phonological WM as the basis for generating the utterance. The Phonological WM plays an important role as the system

<sup>8</sup>For fixed symmetric networks on the other hand a monotonic Lyapunov function exists, ensuring convergence (Hopfield, 1982).

<sup>9</sup>This condition can be relaxed and is not a strong requirement of Schema Theory.

<sup>10</sup>The process of unification will not be described in details here as it can be easily understood from an example or in reference to the classic unification operation between feature structures (Knight, 1989; Shieber, 1986). A few elements are worth being noted. The unification procedure used here is non-destructive and yield a new construction instance without altering the original instances. Unlike the classic unification procedure, the unification operation is not commutative (since in the cooperation links are not symmetrical linking an output port to an input port). However, the unification operation is associative and commutative and therefore the order in which the unifications are performed is irrelevant.

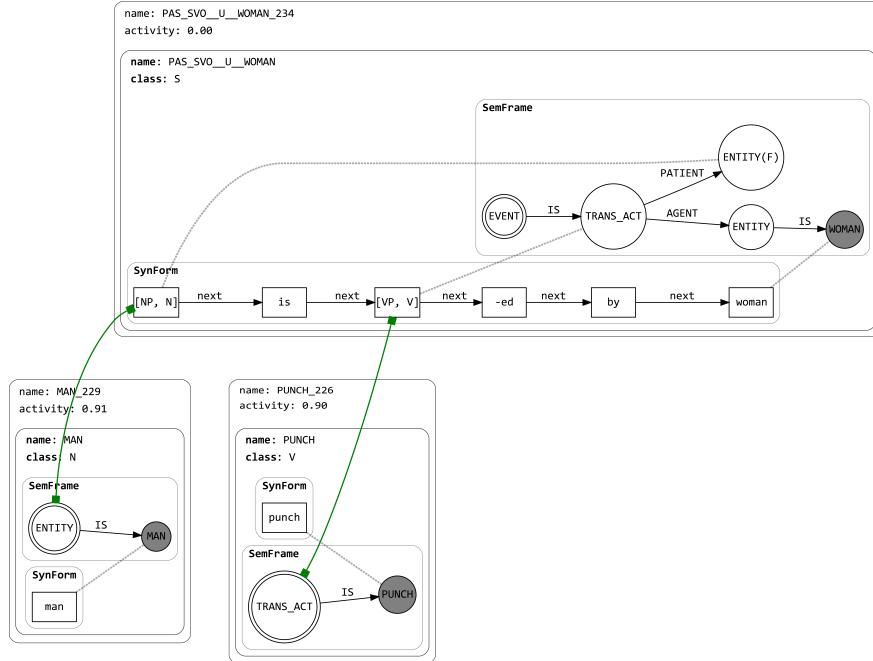


Figure 4.10: Construction instances unification example: Starting with the assemblage shown in the assemblage A shown in fig. 4.9, pairs of constructions instances are iteratively chosen to be unified. The unification process is commutative and associative so the order in which those pairs are chosen does not matter. Once unified, the pair of instances (C<sub>XN1</sub> and C<sub>XN2</sub>) are replaced in the assemblage by their unification product (noted C<sub>XN1\_U\_XN2</sub>). Here the PAS\_SVO construction instance has been unified with the WOMAN construction instance, with the SynForm of WOMAN “woman” providing form content for the last element (slot) of the PAS\_SVO SynForm. Similarly, the SemFrame of WOMAN is unified with the SemFrame of PAS\_SVO: the nature of the agent ENTITY is now precised conceptually as a WOMAN. The resulting construction PAS\_SVO\_U\_WOMAN defines a meaning-form mapping that composes the mapping carried by PAS\_SVO and WOMAN, composition that is constrained by the variable binding established by the cooperation link.

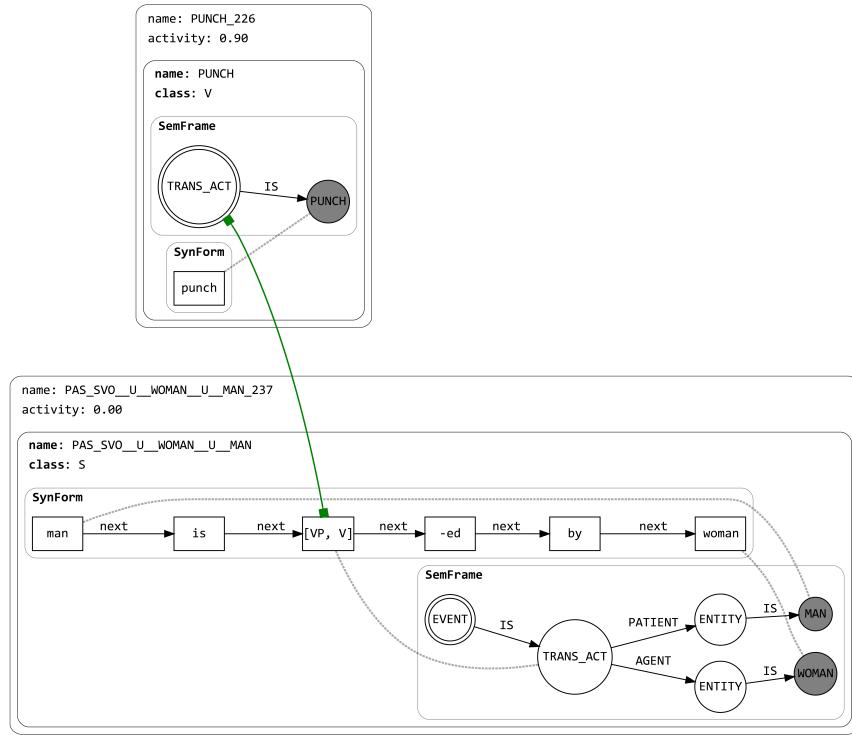


Figure 4.11: This shows the next unification step following the one shown in fig. 4.10. MAN construction instance has been unified with the PAS\_SVO\_U\_WOMAN construction.

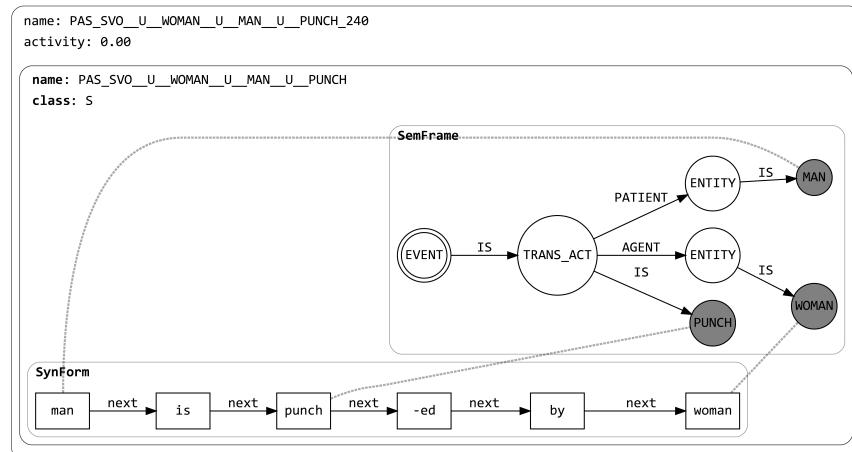


Figure 4.12: This presents the final stage of the unification process. All the construction instances forming the assemblage have been unified. The construction instance equivalent to the assemblage maps the entire SemRep onto a fully lexicalized form that serves as an utterance output: “**man is punch -ed by woman**”.

might be required, in order to continue the incrementally production of utterances, to take into account the phonological content of previous utterances.

Importantly, in the context of SALVIA, if the form still contains slots (a result of missing semantic information), only the portion of the utterance that precedes the first slot is sent to the phonological working memory, while, the system will now use top-down signal to the visual system to try and recover the missing semantic information that will allow it to continue the utterance it started (verbal guidance of visual attention).

When an assemblage has to be selected to generate an utterance, only those that allows for the language production system to start the production are considered (i.e. those whose associated form associated does not begin with slot). Construction instances receive activation both from within the Grammatical WM (through the C2 process) but also from the Semantic WM (from the concept instances that form the SemRep subgraph they express). This functional connectivity across WMs plays an important role in ensuring the dynamic coordination of the semantic and grammatical states: both WMs form a loosely coupled system that can be considered, at a coarser scale, as a single Linguistic WM. When a construction instance is used to generate a form content, the SemRep instances it as contributed to express are marked as expressed' and the functional connections between those and the construction instance are terminated: the construction instance does not receive any external activation from them.

#### 4.4.7 Linguistic WM

Working memories are sub-systems that maintain active information as long as it is relevant to the ongoing behavior. In a system-of-systems approach that consider WMs to be themselves in time-dependent functional interactions (with activation signals flowing from one WM to another), the nature of the WMs to consider is first necessarily a matter of granularity. If Grammatical and Semantic WM are treated as separate systems, at a coarser level, they can be subsumed into a more general Linguistic WM. However, the usefulness of this subsumption (or of the dual decomposition) depends on the (possibly time varying) strength of the C2 interactions *between* WMs with respect to the strength of the local C2 interactions *within* each WM. This won't be discussed further in this work, but is a well known property and corresponding epistemological problem associated with complex systems (see for example the analysis the notion of near decomposability by Simon (1977, 1962))

The Linguistic WM could also be seen as a multi-layer working memory going from a layer that contains the SemRep all the way to a layer that contains only abstract argument structure constructions. Each layer builds on top of the ones below in a way similar to that used in VISIONS.

### 4.5 Good Enough Production of Utterances: Speaker and Task Relevant Parameters

Much work on language comprehension has by now outlined the necessity to understand the comprehension process as solving a satisfying problem: finding an interpretation for an utterance that is good-enough for communication to succeed while satisfying the constraints defined by the current task as well as by the system itself. To this “good-enough comprehension” principle (Ferreira and Patson, 2007) the TCG framework proposes that should be added a “*good-enough production*” principle: the output of the language production system corresponds to a good-enough solution to a given task. Whether or not fluency and well-formedness are the overarching constraints depends on the task at hand. TCG within a language production model (fig. 4.1) accounts for the fact that the processes can function at various regimes and can be impacted by task-related requirements.

**Assemblage score** Alongside its activation value, an assemblage is assigned a score *score*. The score of an assemblage is introduced to account for modulations of the qualities of the meaning-form mapping desired that can translate in difference in utterance production style. For each assemblage four criteria are taken in consideration when computing the score: the assemblage activation value ( $v_{act}$ ), amount of semantic information covered ( $v_{sem}$ ), length of the associated form ( $v_{form}$ ), utterance continuity value ( $v_{cont}$ ).

In the current implementation,  $v_{sem}$  is function of the size of the SemRep graph covered by the assemblage. It is defined as the sum of nodes and edges in the semantic graph covered by the assemblage (or similarly, the number of nodes and edges in the SemFrame of the assemblage equivalent instance, see below).  $v_{form}$  reflects the length of the utterance generated by the assemblage. It is defined as the number of word\_forms in the form generated by the assemblage (or similarly, the length of the SynForm of the assemblage equivalent instance, see below).  $v_{cont}$  accounts for how much the form associated with the assemblage smoothly overlaps and continues an already produced utterance. (Proper definition given below).

With the exception of  $v_{act}$ , all the other values, when computed, are always normalized  $v_i \leftarrow \frac{v_i}{v_{max}} \in [0, 1]$ .

**Speaker style** Four *style parameters* define the weights associated with each of the assemblage scoring criteria  $\overrightarrow{w_{style}} = (w_{act}, w_{sem}, w_{form}, w_{cont})$ , with the constraint that  $|w_{style}| = 1$ .<sup>11</sup>

The score of an assemblage is defined as:

$$score = \overrightarrow{w_{style}} \cdot (v_{act}, v_{sem}, 1 - v_{form}, v_{cont})^T \in [0, 1] \quad (4.3)$$

Varying the value  $\overrightarrow{w_{style}}$  associated with the grammatical working memory results in changes in the style of utterance produced. For example,  $v_{form}$  appears as  $1 - v_{form}$  in the scoring equation so that a higher  $w_{form}$  style parameter value pushes the system towards generating shorter, more semantically compact, utterances.

**Time constraints** One set of parameters defines the characteristics of the dynamics taking place both within and between WMs. In doing so, the core temporal behavior of the model with respect to incrementally received inputs is set.

We propose that those be supplemented by two other sets reflecting constraints on the GrammaticalWM dynamics.

Two key task-related parameters simulate the impact of time pressure on utterance production:  $t_{time\_pressure}$  and  $t_{start\_prod}$ .  $t_{time\_pressure}$  constrains the model to try to produce an utterance at each  $\Delta_T = t_{time\_pressure}$  intervals. Crucially, the system has to do so whether or not all the required perceptual and semantic information has been gathered, and also whether or not the state of the GrammaticalWM has converged to a unique solution (no more competition).  $t_{start\_prod}$  corresponds to an offset time with the first production time  $t_0$  defined as  $t_{start\_prod} + \Delta_T$ . Following  $t_0$ ,  $t_n = t_{n-1} + \Delta_T$

**WM capacity constraints** For each WM, a WM\_capacity parameter stipulates the number of instances that a WM can sustain without loss of information. This parameter captures potential idiosyncratic differences between speakers and task related effects (WM overloading by competing tasks). Importantly, it can also serve as a proxy to simulate the degradation of WM performances following lesions to the neural systems that supports them. If the number of instances in a WM reaches its maximum capacity, instances are pruned out until the WM load within capacity range. Instances that have already been used in the process of generating outputs are pruned first (this nevertheless has an impact as it prevents those from remaining for a while active, although at a low activation level, readily available if their process is required again). In addition, if the maximum capacity is reached, the system is forced to attempt to produce an utterance.

**Control system: Defining and applying the task and speaker related parameters.** The control system is used to store the task and speaker related parameters. Based on those it provides high level task-related control signals to the system. One of its role is to allow for the algorithmic implementation of the processing of experimental designs requirements such as that to wait until a signal is received before starting to produce a description.

The state of the Control system is defined as  $(t_{last\_prod}, unexpressed\_sem, produce, mode)$ .  $t_{last\_prod}$  keeping track of the time at which the last utterance was produced. The PhonologicalWM sends a signal to update this value each time it outputs a new utterance.  $unexpressed\_sem$  is a boolean value updated based on the input received from the SemanticWM that is *true* only if there are semantic elements that have not yet been expressed in an utterance.  $produce$  is a boolean value that is *True* only if  $t >= t_{start\_prod}$

---

<sup>11</sup>In the current model this parameter is taken to be time independent, but further developments should investigate the possibility to define adaptive scoring policies based on the varying requirements of the communicative task.

, *unexpressed\_sem*, and  $(t - t_{last\_prod}) > t_{time\_pressure}$  are all *True*. *mode*, in the present case, will always be set as 'production'. In the more general model integrating production and comprehension, the Control system also pilots the modality switches.

**Phonological WM & Utterance continuity principle** At each time step the Phonological WM feeds back the sequence of word forms composing its state to the grammatical WM. Based on this signal, the grammatical WM can compute for each assemblage, , when asked to produce a linguistic form, an utterance continuity value. Given an assemblage  $\mathcal{A}$ , given sequence  $phon\_form = (f_i), i \in [0, n]$  defined by  $form\_readout(\mathcal{A})$  and given the sequence  $S = (s_j), j \in [0, m]$  of word forms in PhonologicalWM, *continuity* is simply defined as the length of the longest subsequence common to  $phon\_form$  and  $S$  that contains  $f_0$  and  $s'_m$ . For example, if  $phon\_form = ['a', 'boy', 'kick', 'a', 'ball']$  and  $S = ['a', 'boy']$ , *continuity* = 2 while if  $phon\_form = ['a', 'ball', 'is', 'kick', '-ed', 'by', 'a', 'boy']$ , *continuity* = 0.

**Input** When the system is not used as part of the SALVIA architecture, inputs can be defined as incremental semantic representation that stipulates when and how the state of the SemanticWM should be updated (See Appendix.C sec. C.1.5 for details).

**Output** Utterances generated by the model are defined as time stamped sequences of words (and occasionally bound morphemes). The interaction of the system parameters and time pressure (and task parameters in general) impacts the dynamics of the language processes yielding qualitatively different types of utterances ranging from well-formed sentences efficiently packaging the semantic information to short disfluent utterances with little grammatical complexity.

The focus of TCG on online incremental processing enables the exploration of the impact of the dynamics of constructional processes on the quality of utterance production.

## 4.6 Conclusion

The TCG computational approach to construction grammar places at its heart the challenge to model the human brain's capacity to dynamically coordinate two concurrent incremental processes, one generating a message and the other organizing its mapping onto a linguistic form.

In the context existing frameworks of computational construction grammar (CompCxG), we believe that TCG can serve as a bridge between the full-fledged symbolic CompCxG model with large scope (Fluid Construction Grammar (Steels, 2011), and Embodied Construction Grammar) (Feldman, 2010) and the neurally implemented models that work from the bottom-up, tying directly language processing to the existing neural architecture (Dynamic Construction Grammar) (Hinaut et al., 2015). By basing its design philosophy on the Schema Theory approach to brain theory, TCG attempts to learn from the former in developing its symbolic formalism while keeping in contact with the focus on time and dynamics that are at the heart of the latter.

Comparison with other computational frameworks highlights the two main challenges that TCG faces. Scale: How the C2 dynamics scales with the size of the grammar remains to be studied. In particular, the amount of redundancy in the grammar and therefore the ratio of competition to cooperation in the network can potentially have profound impacts on the system's behavior. Semantic expressiveness: the SemRep format was designed not for its expressiveness but to enable the study of how incrementally built semantic representations can be processed online. It will be necessary to enrich it. The challenge is to always do so while preserving the dynamic nature of the operations that build and process the SemRep.

Belonging to a general effort to develop computational neurolinguistic models, TCG is destined to be embedded within schema architectures that simulate what is known of the organization of the neural architecture of the language system. A first consequence is the requirement that the model be able to simulate the degradation patterns observed in aphasics patients (Barrès and Lee, 2013). Another core departure from other CompCxG model lies in that TCG does not assume a priori a symmetry between production and comprehension processes but considers the nature of those relations in the human brain to be an empirical question.

# Chapter 5

## SALVIA: Toward a Neuro-Cognitive Model of Normal and Agrammatic Language Comprehension

*Geschwind:* ‘Do you know what a leopard is?’

*Patient:* ‘Yes.’

*G:* ‘Do you know what a lion is?’

*P:* ‘Yes.’

*G:* ‘The leopard was killed by the lion. Which animal died?’

*P:* ‘I don’t know!’

Neurologist Dr. Norman Geschwind interviewing an agrammatic aphasic.

### 5.1 SALVIA as Neurally Informed Model of Comprehension

In order to go from Schema Theory to Neural Schema Theory, it is necessary for schema-theoretic models to make direct contact with *neural data*. The SALVIA model of language production focused on defining a cognitive architecture that both preformed computations in a way that is coherent with what is known of the brain’s operating principles, but was also built on the basis of empirical *behavioral data* for which it also provided computational interpretations.

In the spirit of schema theory that prescribe to build models incrementally, incorporating new function, refining other by tackling an always wider range of empirical results, this chapter initiates a double movement in the development of SALVIA. Rather than to continue extending the production model in the same direction, the research moves forward by looking at the modeling challenges from a new point of view, resulting from two orthogonal perspective shift:

1. The work will now focus on developing a SALVIA model of **language comprehension**.
2. This move toward comprehension will be accompanied by a focus on neurolinguistic data and in particular on those provided by **aphasiology**.

This chapter offers a conceptual extension of SALVIA into a model of language comprehension: SALVIA\_c<sup>1</sup>. It particularly focuses on two related empirical challenges:

---

<sup>1</sup>the suffixes ‘c’ and ‘p’ will be used when necessary to distinguish between the comprehension and production processes respectively

1. Incorporating the role that world knowledge plays in the incremental generation of semantic representations during comprehension.
2. Doing so in a way that accounts for the comprehension performances of agrammatic aphasic patients.

### 5.1.1 Comprehension Patterns of Agrammatic Aphasics

*Agrammatic* aphasics are patients suffering from brain lesions that result in the deterioration of their capacity to speak in a grammatically correct fashion. Their disfluent speech production patterns and agrammatism have historically been closely associated with Broca's aphasia, although there is no one-to-one link of symptoms with specific lesion sites (and in particular Broca's area). In contrast to a relatively unimpaired capacity to use the correct content words to carry out their message, agrammatic aphasics tend to omit function words, verbal inflections, etc. Caramazza and Zurif (1976) were among the firsts to show that agrammatic aphasics could also be impaired in their capacity to make use of syntactic cues during language comprehension. They found that Broca's aphasics were no different than normal subjects when asked to match a picture with canonical active sentences such as "the lion is chasing the fat tiger", but were no better than chance for center-embedded object relatives such as "the tiger that the lion is chasing is fat". However, performances of Broca's aphasics was restored to the level of normal subjects for object relatives when world knowledge cues were available to constrain the sentence interpretation as in "The apple that the boy is eating is red". This latter result led the authors to hypothesize a neuropsychological dissociation between two comprehension processes: a "heuristic" system based primarily on world knowledge information and an "algorithmic" system relying mainly on syntactic information. Sherman and Schweickert (1989) replicated the experiment while controlling for the possible combinations of syntactic cues, world knowledge plausibility and, importantly, picture plausibility.

Since this seminal work was published, it has been shown that agrammatic production does not necessarily entail agrammatic comprehension and that the comprehension performances of agrammatic aphasics appear quite heterogeneous. Moreover, the very notion that agrammatism reflects the impairment of an identifiable function of a syntactic system (as in the Trace Deletion Hypothesis of Grodzinsky (2000)) is strongly challenged by the diversity of comprehension performances. In their meta-analysis of 15 studies published between 1980 and 1993 that reported agrammatic aphasics' comprehension performances on sentence-picture matching tasks and included contrasts between active and passive constructions, Berndt et al. (1996) found that the 64 unique data sets (for 42 patients) could be clustered into three groups of approximately equal size, each reflecting a distinct comprehension pattern:

1. Only active constructions are comprehended better than chance;
2. Both active and passive constructions are comprehended better than chance;
3. Both structures are comprehended no better than chance.

So far none of the theories linking agrammatism to a specific deficit in syntax processing has been able to account for this variety in performances (for a discussion of the possible role that group selection played in generating this variety see (Berndt and Caramazza, 1999; Zurif and Piñango, 1999)). Rather than conclude that agrammatism does not constitute a useful neuropsychological syndrome for the understanding of the neural and cognitive structure of the language system (Caramazza et al., 2005) we suggest that this diverse set of lesion-behavior data points provides a good target for a new neurocomputational approach (including in particular the fact that Broca's aphasic patients, for whom the lesions tend to be localized in the left-anterior cortex, seem to display only the second pattern of comprehension (Grodzinsky et al., 1999), but we leave the problem of the neural anchoring of the model for subsequent work. However, it remains a target for our work rather than an achievement of this chapter.

Importantly the second conclusion of Caramazza and Zurif (1976) regarding the role world knowledge plays alongside syntax is largely admitted as a non-controversial empirical fact confirmed by subsequent studies (Ansoll and Flowers, 1982; Kudo, 1984; Saffran et al., 1998; Sherman and Schweickert, 1989). We seek, then, to extend SALVIA to provide a model of language comprehension that includes the possibility of selectively impairing various aspects of grammatical processing while leaving world knowledge processes involved in comprehension relatively unimpaired.

### 5.1.2 Light and Heavy Semantics

We cannot yet provide a comprehensive explanation of the heterogeneous performances of agrammatic aphasics but do show how a dynamic, schema-based model can be used as a tool to study some key aspects of these data sets. The focus of SALVIA on the language-vision interface offers a platform well suited to simulate sentence-picture matching tasks. We adopt a two-route approach to comprehension, with a world knowledge route that may be left more or less intact while lesions are performed on a grammatical route. But first we discuss the necessity, following both our construction grammar approach and neurophysiological evidence, to distinguish theoretically between the roles of two different types of semantic constraints on the comprehension system.

We saw in ch. 4 how TCG constructions combine both form (SynForm) and semantic constraints (Sem-Frame). This operationalizes the core tenet of construction grammar that syntax and semantics are not dissociated into two different theoretical components (Croft and Cruse, 2004). But the empirical results reviewed above demand that we distinguish the world knowledge preserved in agrammatic aphasics from construction-related semantic constraints. In (Barrès and Lee, 2013) we thus coined the terms **heavy semantics** and **light semantics** for world knowledge and construction-based (grammatical) semantics, respectively.

World knowledge, as we saw in previous chapters, represents a source of information that plays a pervasive role in both visual scene and language comprehension and is heavy in terms of content since it spans motor and perceptual schemas but also conceptual abstract knowledge that we can acquire through the very use of language. Such knowledge of agents, objects, actions and more abstract entities that gets richer as we interact with the physical and social environment, contrasts with the light semantic content of constructions which develops through experiences of patterns of language about agents, objects, actions and more. The latter may vary from the highly abstract (as in noun versus verb providing a language-dependent syntactic elaboration of the semantic categories of objects versus actions) or strongly linked to sensory or motor experience as illustrated by the example of the IN-COLOR construction (see fig. 2.11).

For us, light semantics reflects this construction-related categorization, more or less abstracted from world knowledge in a usage-based language-laden way. It is “light” because only a few semantic features matter, and it cannot be refined and enriched by interacting with the world beyond the bounds set by a given language (although of course performances vary). Theoretical distinctions have been proposed by others that are closely related to ours (for example see Levin, 1993; Mohanan and Wee, 1999; Pinker, 1989). Those however stem from linguistic analyses, while the light and heavy semantics distinction emerges from considerations related to neuropsychological data and computational brain theory. It will be the role of future work to analyze the relations between these different theoretical perspectives and in particular to discuss how to bridge between approaches like ours that go from the brain-up and those that work from language-down.

Kemmerer (2000a) reported a word-picture matching task that required discrimination between 3 verbs that differed only on the basis of semantic features relevant from a grammatical point of view such as “spill”, “pour”, and “sprinkle”. One subject performed poorly on this word-picture matching task while performing well in a grammaticality judgment task involving the same verbs “slotted” into constructions that matched or not in terms of the “grammatical” semantic constraints e.g., “Sam spilled beer on his pants” vs. \*“Sam spilled his pants with beer” (see ch. 4). Two other patients showed the opposite pattern of performance. This double dissociation has been replicated for the semantic constraints associated in English with prenominal adjective order (“thick blue towel” vs. \*“blue thick towel”) (Kemmerer, 2000b; Kemmerer et al., 2009), those associated with the un-prefixation of verbs (buckle-unbuckle vs. \*boil-unboil) (Kemmerer and Wright, 2002), and for the body-part possessor ascension construction (“Sam hit Bill on the arm” vs. \*“Sam broke Bill on the arm”) (Kemmerer, 2003). These empirical results, bringing a new light on the grammatical impairments that can result from brain lesions, demonstrate the need for our model to account for the possibility of selective impairments of heavy and light semantics in language comprehension.

This distinction between light and heavy semantics (grammatical semantics and world knowledge) derived from empirical neuropsychology results, insists on the fact that, as one develops a neuro-cognitive model anchored in construction grammar, the typical tenets of cognitive linguistics might not fit the endeavor.

“A pivotal theoretical issue is the relation between meaning and grammar. (...) The central claim of cognitive grammar [is] that meaning and grammar are indissociable.” (Langacker, Structural

Syntax: the View from Cognitive Grammar.)

When it comes to brain modeling, the relations between meaning and grammar have to be carved, not by declaring them two separate components as in the generativist approaches, but on the basis of the data that neurolinguistics provide. The notion of world knowledge representations and use will be treated here in a rather cavalier way. Ch. 8 will kick start a discussion of the complex theoretical and empirical challenges that the endeavor of unveiling “meaning” and “meaning processing” in the brain carries.

## 5.2 Dynamic Interactions of World Knowledge and Linguistic Information during Language Comprehension

While remaining centered on the question of modeling the language-vision interface, SALVIA is extended to account for the role of world knowledge during comprehension. This change of perspective places the semantic representation (SemRep) in the position of output of the comprehension system. Meaning and its role in language use remains at the heart of the modeling concerns of SALVIA which therefore focuses on the generation of form-meaning mappings rather than on the generation of a parse trees (cf. (Vosse and Kempen, 2000) for an example of a related and detailed cognitive model that focuses on syntactic parsing).

Moreover the SemRep emerges through cooperative computation between two concurrent and parallel processing routes, using linguistic information and world knowledge respectively. The general architecture of the SALVIA comprehension model is described in fig. 5.1.

### 5.2.1 Lessons from Neurolinguistics: A Two-Route Model for the Processing of Linguistic Inputs

In the work on scene description, the SemRep was generated and dynamically updated by the visual system, with construction assemblages controlling the flexible generation of utterances corresponding to all or part of the SemRep (see ch. 2).

In the SALVIA comprehension model, on the basis of the empirical results mentioned above, we enrich this dynamics by allowing the SemRep to be built and updated not only by the vision system, but also by two routes processing input utterances in parallel. These two routes, shown in fig. 5.1, correspond to (1) the world-knowledge or heavy semantic route (2) the grammatical route.

This idea of parallel routes has also found support in both psycholinguistic and neurolinguistic models. While it had been proposed that the architecture of the language system organizes the linguistic processes serially with syntax being processed first, yielding a syntactic tree from which meaning can be derived (Frazier and Fodor, 1978; Friederici, 2002), recent empirical studies tend to favor a “multi-stream” view of the language system (Osterhout et al., 2007) in which comprehension is the result of parallel processing pathways (with usually one of which is more semantic in nature and related to world knowledge while the other is more syntactic) interacting only at given interfaces (e.g. for models based on ERP data and related to the issue of the “semantic P600” see(Bornkessel and Schlesewsky, 2006; Kim and Osterhout, 2005; Kos et al., 2010; Kuperberg, 2007), for eye-tracking experiments during reading see (Vosse and Kempen, 2009) and for models based on direct comprehension tests see (Ferreira, 2003)). This data show that the emergence of such parallel processing routes does not simply play a role in situations in which the aphasic subject has to make use of as many cues as possible in order to compensate for deficits in grammatical processes. Those parallel routes are at play in online processing for normal language user.

As a consequence of this architectural choice, a given utterance input is fed incrementally and in parallel to a grammatical route (G) that updates the SemRep through the creation of a construction schema assemblage in Grammatical working memory (Grammatical WM) and to a World Knowledge route (WK) that enriches the SemRep on the basis of the world knowledge information both carried by the individual words but also by the basic world knowledge frames they are associated to (Abelson, 1981; Minsky, 1974), frames that carry information regarding the likely events, participants and roles that are associated with a given object, action, etc and have been shown to play a key role in semantic enrichment during language comprehension (Metusalem et al., 2012). This world knowledge information relevant to the ongoing comprehension process is processed by the World Knowledge working memory (World Knowledge WM) before updating the SemRep.

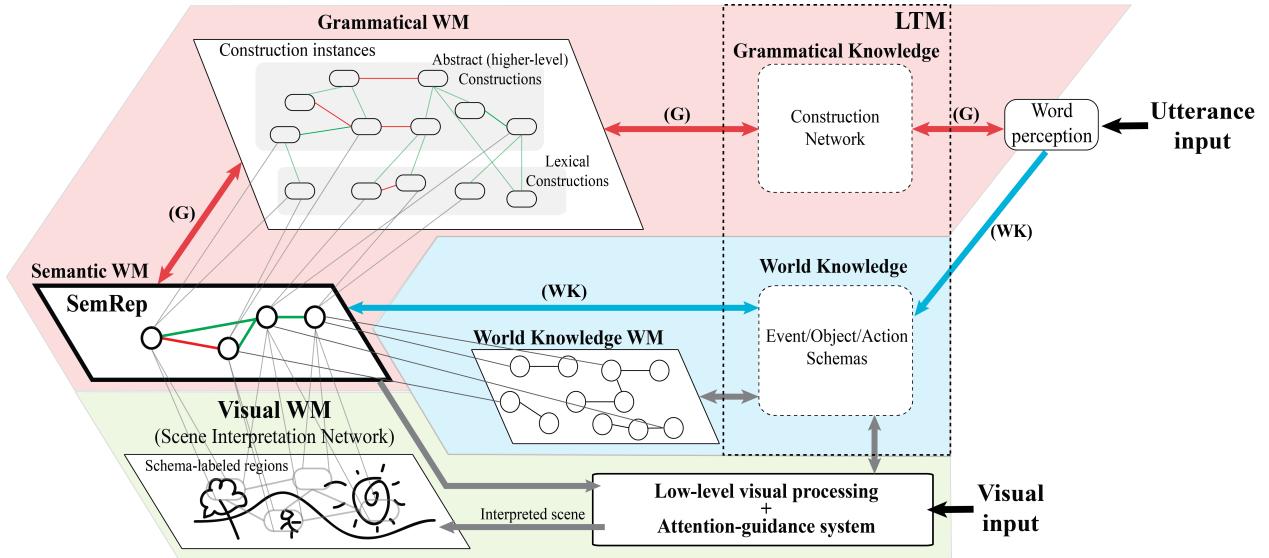


Figure 5.1: The full (informal) SALVIA comprehension architecture. The top-part of the model represent the grammatical route (G). It can be understood as reversing the processes that were at play in the production model (see ch. 2, fig. 2.2). The bottom-part of the model correspond to the visual route, similar to the one used in production. The main change is the addition of a third world knowledge driven route (WK). With those three routes, the semantic representation hold in Semantic WM (SemRep) becomes the locus of interaction between three sources of information: grammatical, pragmatic (world knowledge) and contextual (visual). Those enter in cooperative computation. The semantic interpretation of a given linguistic input is therefore the result of dynamic competition cooperation processes between multiple, opportunistically used, sources of information.

The Semantic WM that holds the SemRep becomes the locus of a cooperative competition where the construction assemblage on the one hand and the world knowledge hypotheses on the other compete and cooperate at each time to update the SemRep. The Visual working memory remains a source of input for the SemRep as in the production model.

A key feature of the SALVIA comprehension model's architecture lies in the fact that each input word will have an effect on 2 different routes triggering two parallel processes to update the SemRep.

Following the principle of cooperative computation, and in line with the work on production, the SemReps graphs active in semantic working memory are also defined both in terms of their structure and in terms of the activation levels of their components (nodes and edges), activation levels that reflect the degree of confidence associated with a conceptual relations schema instances (edges) or with non-relational conceptual schema instance (node). Any modification made by either the world knowledge or the grammatical route on a SemRep subgraph is expressed in terms of a change in activation levels (that can result in piece of graph being discarded altogether if its activation level becomes too low). In case of semantic incompatibility, competition between routes can directly trigger competitions between incompatible concept schema instances (e.g. between two different role assignment). These can therefore register the competition and cooperation of both routes.

By computationally grounding the comprehension process into the cooperative computation, SALVIA highlights the problem of determining when the computation should stop. A grammatical processing can therefore be good-enough to support a semantic interpretation of the input without necessarily exploiting or satisfying all the syntactic constraints, a position that echoes the empirical findings of (Ferreira and Patson, 2007) related to the notion of “good-enough comprehension”, a principle that had been extended to “good-enough production” in the work on SALVIA as a model of language production (see ch. 2).

### 5.3 Semantic WM: Incremental and Dynamic Semantic Representation (SemRep)

The goal of all the concurrent processes underlying the comprehension schema system (grammatical route and world knowledge route) is to dynamically and flexibly build structures (grammatical or world knowledge based) mapping an incrementally received sequence of word forms (form content) onto a semantic representation (decoded message).

The semantic representation is defined by the state of the Semantic WM and, as in the case of the production model (see ch. 4) is composed of concept schema instances where concepts can be of a few general types: EVENT, ENTITY, ACTION, PROPERTY, which not relations , and RELATIONS with subtypes for each of the general non-relational classes mentioned above: OBJECT\_FRAME relations, ACTION\_FRAME relations (thematic roles), and EVENT\_FRAME relations (which in turn include SPATIAL and TEMPORAL relations). This anchors form of the semantic representation in a type of neo-Davidsonian event representation, with the important caveat that all the relations (predicates) are binary in the model.

As a result from the binary nature of the relations, the state of the SemanticWM can always be represented as a graph (SemRep) where the nodes are the non-relational concept schema instances while the edges are associated with the relation concept schema instances. This situation is similar to the TCG production model which also relied on the graph representation of the message to anchor the processes necessary to generate meaning-to-form mappings.

Comprehension however introduce an important modification to the nature of the SemRep. In the case of production, the SemRep, at each time step, stood for the current state of the message to be communicated. The hypotheses was made then that there was no competition between concept schema instances: the decision process leading up to the incremental building of the semantic representation was thought to be handled at the level of conceptualization, with no residual uncertainty between conceptualization options (which would have led to competition) persisting at the SemanticWM level<sup>2</sup>.

Comprehension and production are inherently asymmetric with respect to the relations between message (SemanticWM) and utterance (input or output). Clearly, for production, the only way to go from a message (SemRep) to utterances is via the use of grammatical processes (GrammaticalWM). And in addition, since the output of the grammatical processes has to yield an linguistic form that will be ultimately converted into a motor pattern to generate an utterance, the nature of the output requires a single choice of linguistic form to be made (there can be only one winning assemblage of grammatical construction instances among all the competing ones) (Cisek, 2006; Cisek and Kalaska, 2010). Acting in the world constantly requires decision among many possible options.

Comprehension, on the other hand, has not as a result an action in the world, but only a cognitive state. Therefore, decision is not required (at least not right away) and multiple grammatical and semantic interpretation, **including contradictory ones**, can coexist. This fact is exacerbated by the main asymmetry between production and comprehension, v.i.z that during comprehension, any source of information that can help in the process of decoding the message carried by an utterance can be opportunistically used. Below we will discuss the roles that in particular World Knowledge and, in the case of visually situated comprehension, perceptual content, play in comprehension alongside grammatical processing<sup>3</sup>.

The literature on Good Enough Comprehension (Ferreira and Patson, 2007) has shown that comprehension is usually imperfect in that it is best described as a satisficing process (Simon, 1972) that builds the ‘best’ interpretation of an utterance given the set of constraints and goals the comprehension task is embedded in. But importantly, it has also shown that it is **not** a requirement for the final semantic interpretation to be devoid of contradictions (which in our schema theoretic framework would translate into competing instances with no termination of the competition).

---

<sup>2</sup>For production, the SemRep was hypothesized to grow monotonously: only new compatible semantic information could be added to the SemRep (hence to the message) during a pass at verbalizing some semantic content. This specification was both a simplification but also a theoretical choice to consider that at each time step the grammatical construction instances should have access to a semantic structure that would include its path-dependent history (if revision were made they should appear to respect monotonous growth) as this could impact the choice of grammatical structure used to express the message (in line from the usage-based credo).

<sup>3</sup>Everyone who has tried to learn a new language knows that it is (at least initially) much easier to understand than to speak!

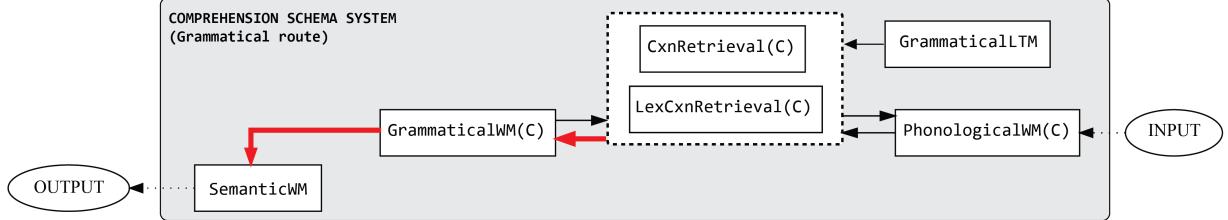


Figure 5.2: First component of the language comprehension sub-system of the Schema Architecture Language-Vision InterAction comprehension model (SALVIA\_c). The figure presents the architecture of the ‘Grammatical route’ whose goal is to dynamically process incrementally received verbal input to generate and update online a semantic representation (state of the SemanticWM) using grammatical knowledge. Each box corresponds to a system with arrows indicating message passing. The core of the system lies in the articulation and temporal coordination of the three main working memory systems: PhonologicalWM (storing the incrementally received verbal input), the GrammaticalWM (grammatical processing) and SemanticWM (where the decoded message is incrementally built).

Patson et al. (2009) carried out a study that directly demonstrate this point: They showed that semantic partial mis-interpretations can linger when garden-path sentences are processed and the correct parse is only partially recovered. Presented orally with sentences containing optionally transitive (OT) verbs (e.g. ‘bath’, ‘hunt’) such as (1) ‘While she bathed the baby played in the garden’ or (2) ‘While he hunted the deer ran through the woods’, a significant portion of the subjects who were asked to then paraphrase the sentences they had just heard would respond (1) ‘The woman is bathing the baby... the baby is playing in the garden’ or (2) ‘A man is hunting a deer who is running through the woods’. (1) and (2) have a unique interpretation but are initially garden-path as the OT verbs tend to be analyzed as transitive when they are used as intransitive verbs, leading to the mis-assignment of the following NP as their object when the rest of the sentence reveals that it should be assigned as the subject of the main clause transitive verb. Paraphrases such as (1) and (2) indicates that subjects upon encountering the main clause verb, did reanalyze the second NP as subject of the verb but the incompatible interpretation of the NP as object of the first OT verb lingered instead of being discarded. This provide evidence that the state of the semantic interpretation of a sentence need not be devoid of contradicting conceptual content.

For those reasons, in the TCG comprehension model, concept schema instances active in SemanticWM can enter in competition. **The SemanticWM becomes a locus of cooperative computation (C2).**

The fact that conceptual relations can enter in competitions means that multiple (competing) relations instances can now exists between concept instances. The SemRep has to be lifted from a labeled DiGraph to a labeled MultiDiGraph structure, allowing for multiple directed edges between similar source and target nodes<sup>4</sup>.

The following proceeds incrementally. It starts with a comprehension route relying only on the TCG\_c grammatical processes (Step 1): this is the linguistic comprehension per se. It then moves on to describe a concurrent and parallel comprehension route that bypasses grammatical knowledge to use pragmatic world knowledge to interpret a linguistic input (Step 2). Finally, the question of the dynamic integration of the processes supported by those two routes into a semantic representation is tackled (Step 3).

## 5.4 From Form to Meaning via Grammar: The Grammatical Route (GR) (Step1)

### 5.4.1 Template Construction Grammar as a Schema-Theoretic Model of Grammatical Processing for Incremental Language Comprehension

Fig. 5.3 presents an informal example of incremental interpretation. In contrast with production, the SemRep is here a multi-graph. Each set of edges between two nodes correspond to competing relational interpretations. Cooperation between instances is denoted by green links, competition by red ones. Just as in the case of production, the state of the Grammatical WM is composed of active construction schema instances and of the C2 network of cooperation and competition links they have formed. The state of the Semantic WM consists of active concept schema instances with the addition, compare to production of competition links. The concept schema instances receive activation from the construction instances whose mapping has participate in their instantiation (cross-WM cooperation links not shown). (Leftmost panel) The linguistic content “The boy” has been received. This triggered the instantiation of lexically anchored construction instances (BOY and DEF\_NP constructions) as well as bottom-up predicted instances (SVO and PAS\_SVO). This state of Grammatical WM has generated a SemRep that contains the semantic information regarding the BOY referent (grey node). Both SVO and PAS\_SVO yield partially similar predictive semantic representations involving the expectation of a transitive action (TR\_ACT) and of another ENTITY. However, they differ in the role they assign to the two entities resulting in competition. (Center panel) Upon receiving the linguistic content “eats”, the PAS\_SVO construction has been pruned since its prediction are contradicted by the input. The SemRep is updated with a EAT referring node. It is worth noting that the competition at the semantic level between role assignment that had been triggered by the competing SVO and PAS\_SVO construction instances has not yet been resolved. The dynamics of the Grammatical and Semantic WM are asynchronous. (Rightmost panel) Upon receiving the input “the cake”, the relevant construction instances are invoked, resulting in a complete assemblage that can provide the missing referent entity (CAKE) to the SemRep. At this point the losing semantic relations have been pruned out of Semantic WM leaving only a single unambiguous interpretation.

Fig. 5.4 continues the example started in Fig. 5.3.

### 5.4.2 A Look at a More Complex Example (Limited Prediction)

The (conceptual) example presented above (cf. Fig. 5.3 & 5.4), are limited to incrementally generating meaning for an utterance containing a single clause structured by a transitive verb with a single prepositional phrase.

This is by no means a limitation of the formalism. Those simple utterances and semantic representations are however often sufficient to illustrate the core theoretical and computational features of the model.

Representing the full state of the model for more complex cases is challenging but it is worth looking at one more complex processing example. Fig. 5.5 & 5.6 show a simplified version of the model’s incremental processing (in the present context of focusing only on the Grammatical Route) of the utterance “*The woman that is kiss-ed by the man kicks the dog the red ball because the boy cries.*” This example contains one main clause organized around a di-transitive verb, a center-embedded subject relative clause in passive voice, an adjectival phrase, with the ditransitive relations between verb and arguments structured by the double-object construction. This first sentence is conjoined to a second one by a causal discourse marker (subordinating conjunction “because”). The second sentence consists of a single clause organized around a intransitive verb (compare this example to the complex processing example for production provided in Ch. 3, sec. 3.4.3.)

In the example of processes presented here, the actual simulation has been simplified for legibility, while respecting the general outcome of the model. In order also to simplify the output, the model was run under limited bottom-up prediction capacities and on a smaller grammar. The bottom-up prediction is here limited to the bottom-up invocation of construction instances whose left-corner (initial prediction) correspond either to a lexical input (lexically triggered retrieval) or match onto a completed construction instance (i.e. a construction instance does not trigger predictions until it is completed).

<sup>4</sup>This change has important computational repercussions since, compared to production that was using the already NP-hard sub-graph isomorphisms algorithms, core processes now rely on sub-multigraph isomorphisms algorithms (see below)

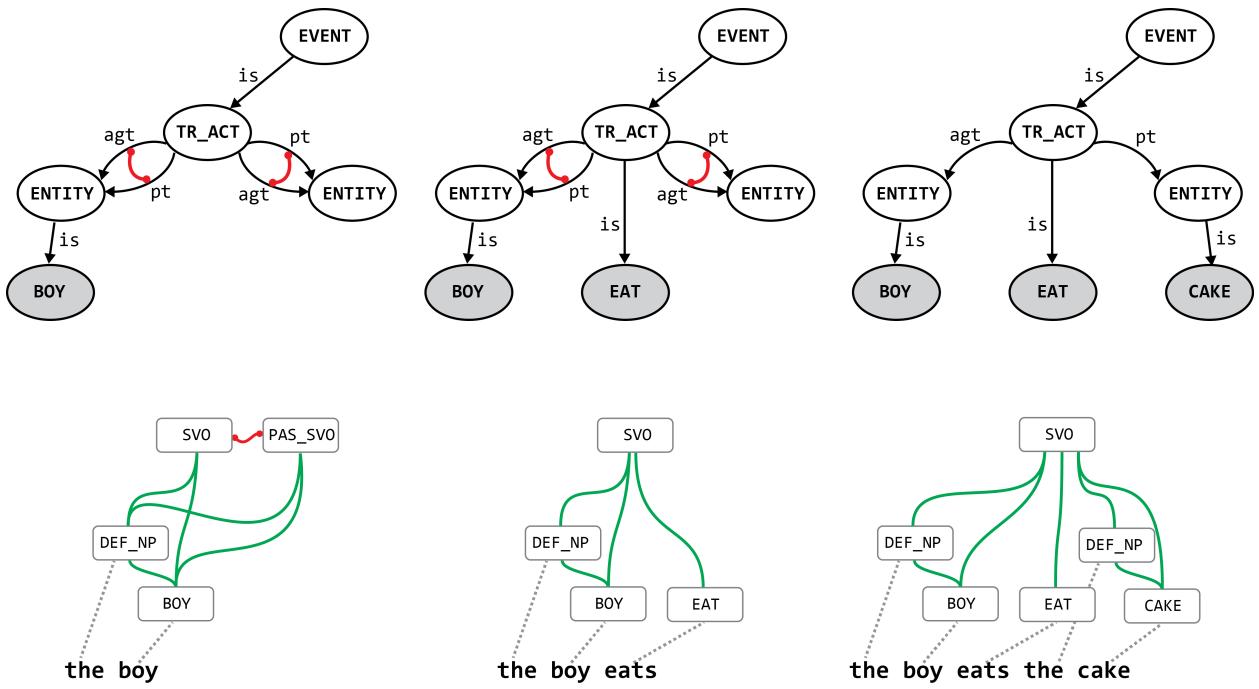


Figure 5.3: Grammatical route only: Informal example of incremental interpretation, through the grammatical route only of the utterance “the boy eats the cake”. From Left to Right, snapshots of the states of both the GrammaticalWM (Bottom) and of the SemanticWM (Top). At each time step, the content of the Semantic WM corresponds to the semantic representation (SemRep) derived from the form-meaning mappings emerging from the C2 processes taking place in Grammatical WM and that have reached a certain degree of confidence. (Details in text)

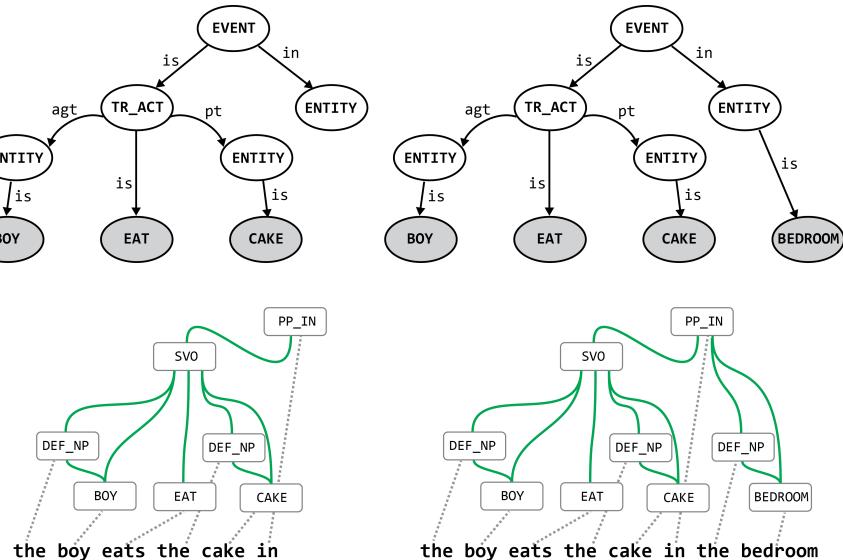


Figure 5.4: Grammatical route only (2). Here the example described in fig. 5.3 continues. The previously interpreted utterance “The boy eats the cake” continues with the prepositional phrase “in the bedroom”. The Grammatical WM state is continuously updated to dynamically adapt the emerging form-meaning mapping based on the incrementally received linguistic input.

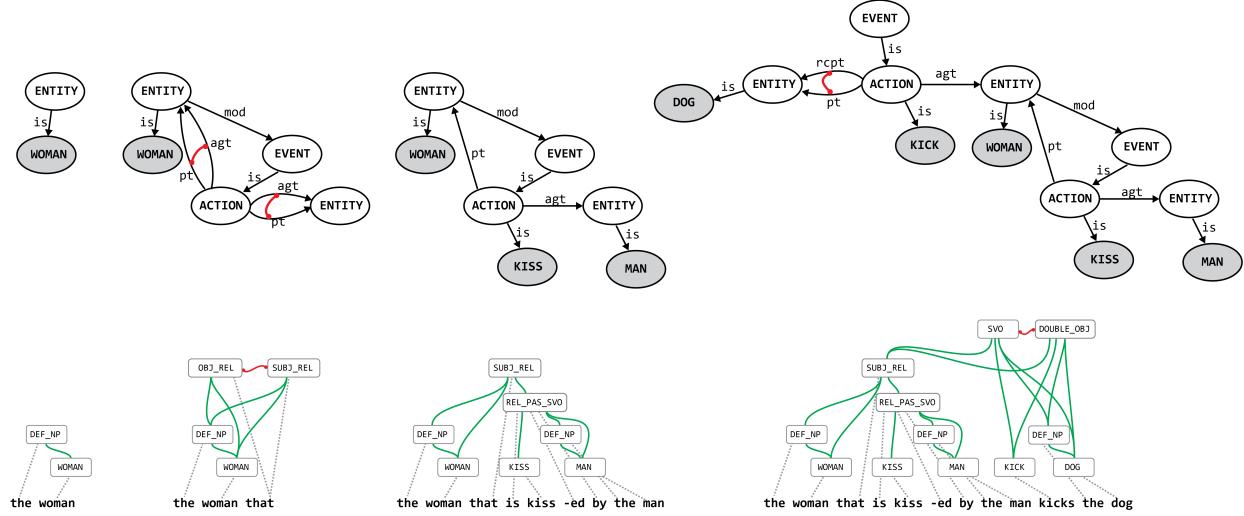


Figure 5.5: A more complex example of grammatical route only processing (1). For simplicity, the system here only carries out limited predictions: it does not apply top-down predictions and limits the depth of left-corner bottom-up predictions to one level. This example illustrates how the processes at play can handle inputs of greater grammatical and semantic complexity than single clause utterances.

This corresponds to a case in which only direct input or instances that have been fully confirmed by the incrementally received inputs can generate predictions.

The conventions used in the figures are the same as the ones used above.

The processes underlying the grammatical route will be detailed in ch. 6, sec. 6.2.

## 5.5 From Form to Meaning via World Knowledge: the World (event) Knowledge Route (WKR) (Step2)

After presenting a first processing route using grammatical processes to go from form to meaning, this section to a second route using world knowledge to, concurrently, attempt to fulfill a similar goal.

### 5.5.1 From Incremental Linguistic Input to Incrementally Built Semantic Representations: Case of World Knowledge Route Only

Fig. 5.7 presents the architecture of the ‘World knowledge route’ whose goal is to dynamically process incrementally received verbal input to generate and update online a semantic representation (state of the SemanticWM) using general pragmatic world knowledge. Each box corresponds to a system with arrows indicating message passing. The core of the system lies in the articulation and temporal coordination of three main working memory systems: PhonologicalWM (storing the incrementally received verbal input), the WorldKnowledgeWM (world knowledge processing) and SemanticWM (where the decoded message is incrementally built). The world knowledge processes hosted by WorldKnowledgeWM are triggered by the retrieval of the conceptual information carried by lexical items. The model shows that alongside the usual conceptual knowledge (ConceptLTM, used, although not shown, in Fig 5.2 and in the production model), the comprehension model enriches this world knowledge with event frame knowledge that forms the basis of the FrameLTM. Those Frame event schemas hosted by FrameLTM are the main drive behind the World Knowledge route capacity to generate Semantic Representation. They offer a world knowledge counterpart to the construction schemas used in the grammatical route.

Fig. 5.8 presents an informal example for the case in which only the world knowledge route is available. The main differences with the Grammatical route only scenario is that when the verb “eat” is received, it triggers the retrieval of the EAT\_FRAME even schema in World Knowledge WM. This in turn results in the

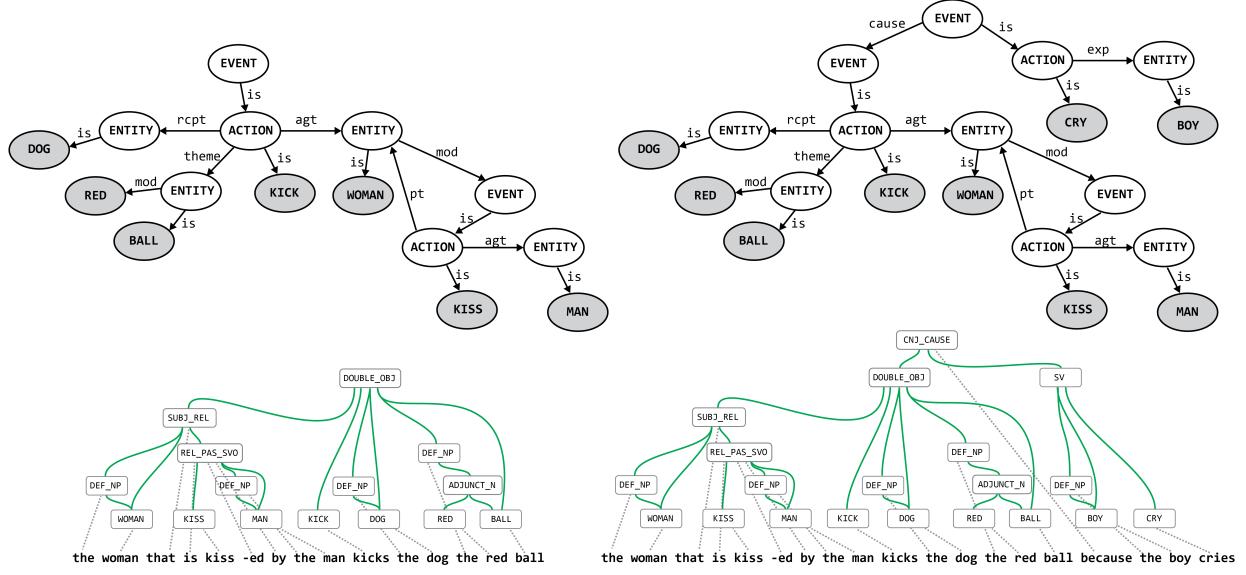


Figure 5.6: A more complex example of grammatical route only processing (2). Continue the example started in fig. 5.5.

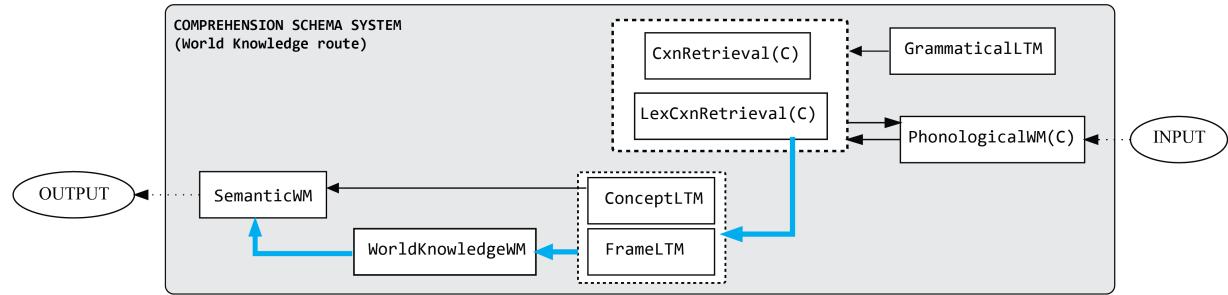


Figure 5.7: Second component of the language comprehension sub-system of the Schema Architecture Language-Vision InterAction comprehension model (SALVIA\_c). (See main text for details)(crossWM links are not shown.)

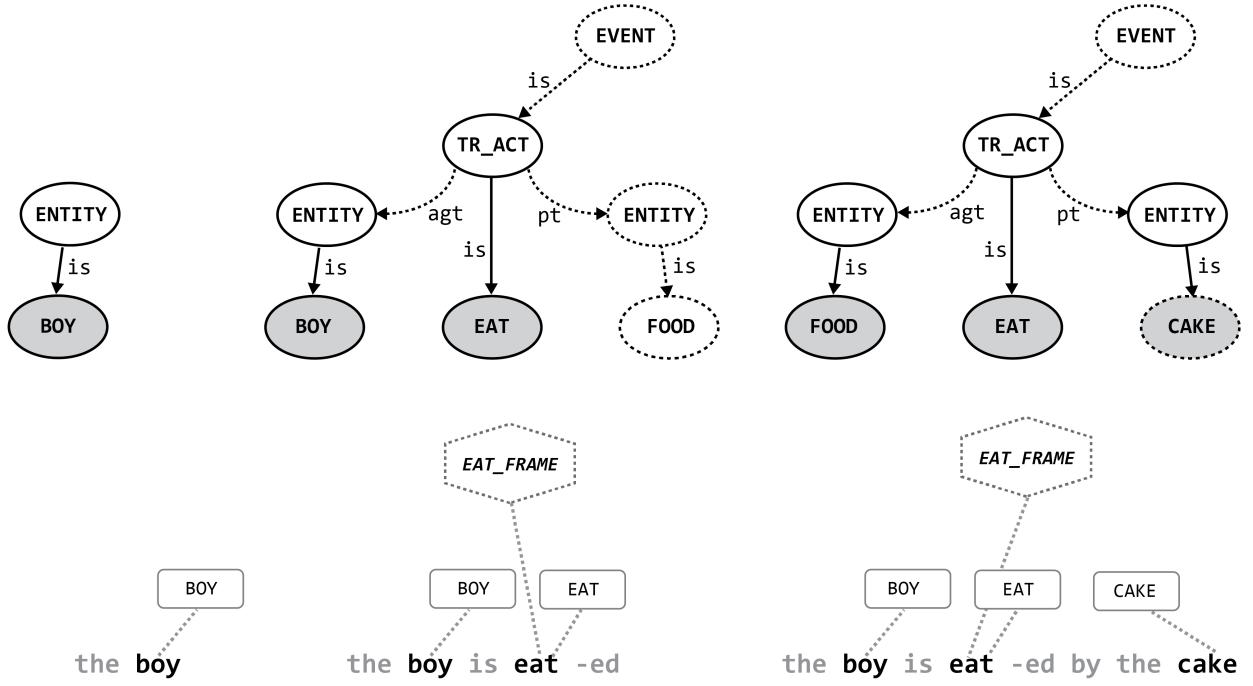


Figure 5.8: World knowledge route only (and lexical construction retrieval): Informal example of incremental interpretation, through the grammatical route only of the counter-factual utterance “the boy is eat-ed by the cake”. The figure follows the same general conventions as fig 5.3 and fig.5.4. A few key differences. Hexagons denote world knowledge event frames schemas active in World Knowledge WK. Dashed arrows and nodes in the SemRep denote the concept instances that have been instantiated based on world knowledge. The input words in bold correspond to those that trigger the retrieval of lexically anchored construction instances associated with conceptual knowledge. (see main text for details)

SemRep assigning the role of agent to BOY, on the basis of pragmatic and not grammatical considerations. It also results in a FOOD ENTITY being predicted to fill the role of patient (Central panel). This prediction is confirmed when “cake” is received, leading a full SemRep without competition that, however, assigns to the entities the semantic roles opposite to those stipulated by the utterance’s syntax but pragmatically most likely (a direct result of the counter-factual nature of the utterance)

The processes underlying the world knowledge route are detailed in ch. 6, , sec. 6.3.

## 5.6 Multi-Route Concurrent Incremental Processing of Verbal Input: C2 Between Routes (Step3)

This section turns to full system in which both grammatical and world knowledge routes are concurrently processing the input and enter in cooperative computation in their attempts to incrementally generate a semantic representation for a given input.

### 5.6.1 Semantic WM as a Locus of Cooperative Computation

Fig. 5.9 shows the full language comprehension sub-system of the Schema Architecture Language-Vision InterAction comprehension model (SALVIA\_c). Both routes described above are not integrated within a single schema system. As a result, the SemanticWM is now dynamically updated by two concurrent processes. Those therefore enter in *cooperative computation at the system architecture level* (in contrast to

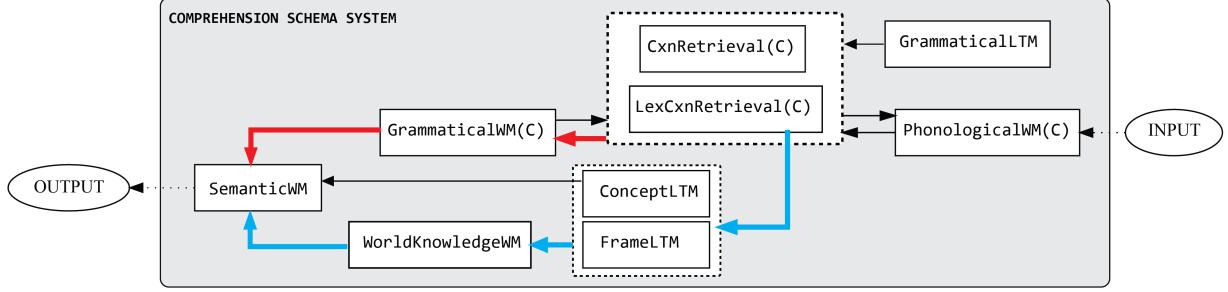


Figure 5.9: Two-route comprehension schema system. Full language comprehension sub-system of the Schema Architecture Language-Vision InterAction comprehension model (SALVIA\_c). Grammatical and the World Knowledge routes (red and blue respectively) are combined and the processes they support take place concurrently. They enter in cooperative computation at the level of the SemanticWM.

the cooperative computation taking place at the purely functional level within a WM). The SemRep becomes the locus and mediator of this between-route C2.

Fig. 5.10 & 5.11 revisit the examples provided in the previous two sections, taking into consideration the two concurrent parallel processing routes and the cooperative computation they engage in at the Semantic level.

Fig. 5.10 illustrates the case in which the two routes reinforce each-other and cooperate to generate a robust semantic representation for the utterance “The boy eats the cake”. For simplicity only three stages of the process are shown.

**Left** Only the Grammatical route has started to generate a form-meaning mapping, with the SemRep reflecting the competition between two possible semantic role assignment carried by the competing PAS\_SVO and the SVO construction instances (here again, in this informal example, the set of construction involved in the process is deliberately limited so as to keep the figures legible).

**Center** Upon receiving the verb, the world knowledge route start generating its own form-to-meaning mapping based on the invocation of the EAT-FRAME schema instance. The semantic content of the frame instance confirms the subgraph of the SemRep shown in solid lines. The PAS\_SVO construction instances has lost the competition since it is disproved by the linguistic input. Compared with the similar situation shown in fig. 5.3, this time the competition between semantic role assignment at the SemRep level is already resolve: the world knowledge route cooperate with the grammatical route in supporting the BOY as agent hypothesis. This precipitates the resolution of the semantic ambiguity: the combined impact of the two routes is not a linear composition of the impact of each route taken separately! At this stage, in contrast to fig. 5.3, the world knowledge route sets up semantic expectation about the nature of the patient (FOOD, dashed lines).

**Right** The final state is similar to that of fig 5.3 with an important caveat. In this case, the grammatically generated form-meaning mapping generated after having received the inputs “the cake” does not build new graph structure in the SemRep but simply confirm the existing structure and specify the referent for the patient. A second scenario is possible although not implemented here. If predictive feedback is defined between Semantic and Grammatical WM, the construction instances used to update the form-meaning mapping can be invoked prior to receiving the linguistic input on the basis of the SemRep structure predicted by the world knowledge route. In this case, incoming word forms themselves are predicted top-down by the state of the Grammatical WM , state that they in turn corroborate or contradict.

Fig. 5.11 presents the opposite case. The processing of a counter-factual utterance in which the two routes compete and frustrate each resulting in a situation in which the final state of the system depends can take multiple form (phase state with multiple attractors): the general analysis of those is beyond the scope of this work but the system can at least either converge toward a semantic state that is the one supported

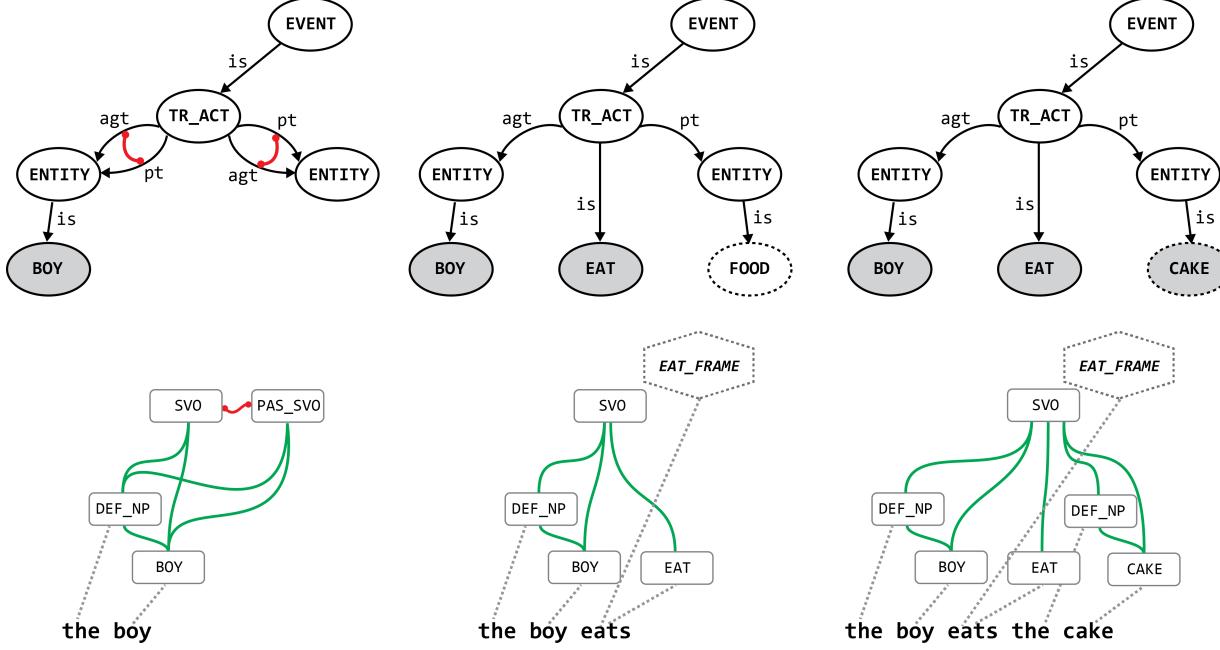


Figure 5.10: Case 1: Cooperation between routes. Here the example that was presented in fig. 5.3 is re-analyzed with now both processing routes concurrently involved in incrementally mapping the linguistic input onto a semantic representation. (Details in main text)

by the world knowledge or to a semantic state that is the one supported by the grammatically derived form-meaning mapping. But there is no guarantee that the system will converge toward a state that in which all competition have been settled (which fits with the idea that competing interpretation could remain active until a decision force a choice, see above). Finally, it is worth noting that the convergence towards an fixed point attractor is also not guaranteed. The relevance or desirability of such dynamical regimes to the comprehension process is unclear. It is the contention of the model however, that one should not a priori attempt to eliminate those in the absence of compelling empirical evidence that they do not exist. Rather, it is the role of dynamical system modeling to enter in a dialogue with modelers regarding possible behavior of a system that they might not have considered.

The processes underlying the interactions between grammatical and world knowledge routes are detailed in ch. 6, sec. 6.4

## 5.7 Conceptual Account of Agrammatic Comprehension Performances in Sentence-Picture Matching Tasks

The empirical evaluation of comprehension performances in aphasics requires the use of experimental paradigms that let the neurologist or the researcher probe the interpretation of the sentence that the patients generate during the comprehension process. As we described in the previous section on agrammatism (see sec. 5.1), sentence-picture matching tasks are commonly used in aphasia battery tests or in neurolinguistics experiments. The patient listens to an utterance while or before being presented with one or multiple visual scenes. The task consists then for him to answer questions about these scenes. In the case of single scene presented the question can be “Does the scene match the utterance?” which requires a yes or no answer. Alternatively, if multiple scenes are presented, the question can be “Which scene matches the utterance?” which requires the patient to point to the correct scene. Basically, the task tests the capability of the patient to determine whether there is a SemRep for the sentence that matches a SemRep for a given picture in order to reach a yes/no decision. Alternative approaches could

1. use vision to generate SemReps for each scene, use SALVIA in comprehension mode to generate a

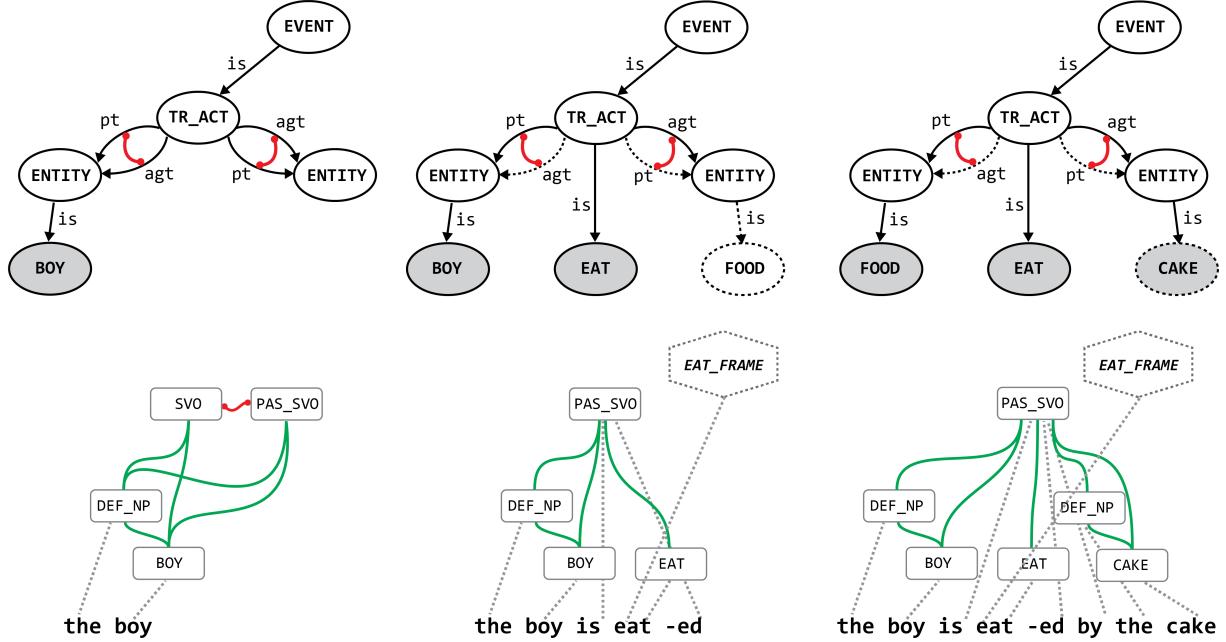


Figure 5.11: Case 2: Competition between routes. Here the example shown in fig. 5.8 is revisited, this time with both route concurrently and incrementally generating their own form-meaning mappings. The world knowledge route, as was shown in fig. 5.8, yield the pragmatically justified form-meaning mapping that assigns the role of agent to BOY, CAKE being the patient. The grammatical route generate a form-meaning mapping that results in the opposite role assignment. Once the verb is received, the SemRep becomes the locus of competition between two role assignments even though each route generate an unambiguous interpretation (dashed lines represent the assignment derived from the world knowledge route). The issue of the outcome of this competition is at the core of the question addressed by the SALVIA model: only by simulating the dynamics of the processes under a given set of parameter, among which the relative weight assigned to each route and the time characteristics of the processes in each WM play a crucial role.

SemRep from the target sentence, and test whether these match in some sense;

2. use SALVIA in production mode to generate a sentence describing the scene, and then test whether this matches the target sentence in some sense.

This begs the question of whether the processing of visual scenes used in the sentence-picture task is segregated from the workings of the “language system” or whether the two cooperate throughout to reach a decision. The design of the experiments could be such that the effects of the visual scene presentation are counterbalanced and controlled at the level of multiple trials and/or subjects, leaving the possibility to specifically target the language system. However, the goal of computationally guided models is to simulate the behavior of a single individual performing a single trial of a sentence-picture matching task and cannot eschew the problem of incorporating the role of visual processes. The SALVIA models of comprehension and production allow us to revisit the experimental results on agrammatism focusing on (conceptual) simulations of a sentence-picture matching trial, limiting ourselves to tasks involving only one visual scene for which the subject has to decide whether or not it matches the utterance they hear.

Fig. 5.12 provides four conceptual examples of sentence-picture matching trials illustrating the collaborative computation of the grammatical and the world knowledge routes but also of the visual processes in generating the semantic representation associated with an utterance. In each panel, the top section represents the perceptual schemas active in visual working memory, the bottom section represents the state the construction assemblage in grammatical working memory, the left section represents instantiated world knowledge hypotheses in world knowledge working memory (the number indicates their activation level). At the center of each panel, all three routes converge to update the content of semantic working memory. For simplicity, we present in each panel a static snapshot of the system that illustrates a key property of the SALVIA comprehension model’s dynamics in relation to modeling agrammatic comprehension. We assume that the utterance has been received to insist on the cooperative computation between routes rather than the incremental process of word by word comprehension (that has been described in the previous sections). We therefore also assume that all the nodes associated with content words have been instantiated, the three routes collaborating and competing for the assignment of relations between them (edges).

Fig. 5.12 (a) illustrates the situation in which an utterance neutral in terms of world knowledge is received, “the cheetah is chased by the tiger”, while a visual scene that accurately matches its meaning is presented. The TIGER, CHEETAH and CHASE nodes are linked to their respective perceptual schemas, hypotheses in the world knowledge working memory, and to the respective TIGER, CHEETAH and CHASE constructions in grammatical working memory. The goal of the cooperative computation between routes is therefore to judge what roles to assign to the TIGER and CHEETAH nodes in relation to the CHASE action nodes. The thin arrow mapping the grammatical working memory to the SemRep simulates a partial lesion of this route resulting in a deficit in using the form-meaning mapping generated by the grammatical route to update the SemRep (see Fig. 5.12 (d) for an alternative lesion simulation). The interpretation of the utterance input rests therefore essentially on the world knowledge route. Since, from a world knowledge perspective, it is just as likely to have either the cheetah or the tiger as the agent of the action, such situation can result in an at chance assignment of the agent and patient role. However, the fact that, in the passive construction, the CHEETAH node is created first, results in an earlier activation of the CHEETAH related world knowledge that could bias the assignment of the CHEETAH node as an agent, agent role confirmed by the then generated CHASE node. Once the TIGER node is created it would then fill in the patient role in the already quite highly activated and stable SemRep. Assuming some residual but weak capacity to assign roles based on grammatical processes, for an input in the passive voice, the weak assignment generated by SVO\_Passive construction would compete with the world knowledge “agent received first” hypothesis while, for the active voice, the weak SVO construction would cooperate with the similar world knowledge “agent received first” hypothesis. Such an interpretation would explain the classic pattern of agrammatic comprehension for which patients are above chance for active sentences and at chance for passive sentences even when no world knowledge cues are apparently available. However, in our case, the explanation of the comprehension pattern is not found in a differential treatment of SVO and SVO\_Passive constructions (one being “more lesioned than the other” or “harder to process”) but emerges from the cooperative computation between routes.

In Fig. 5.12 (b), the input utterance is “The ball is swatted by the cat”. This example illustrates the case in which the world knowledge route generates a unique highly plausible hypothesis that can collaborate

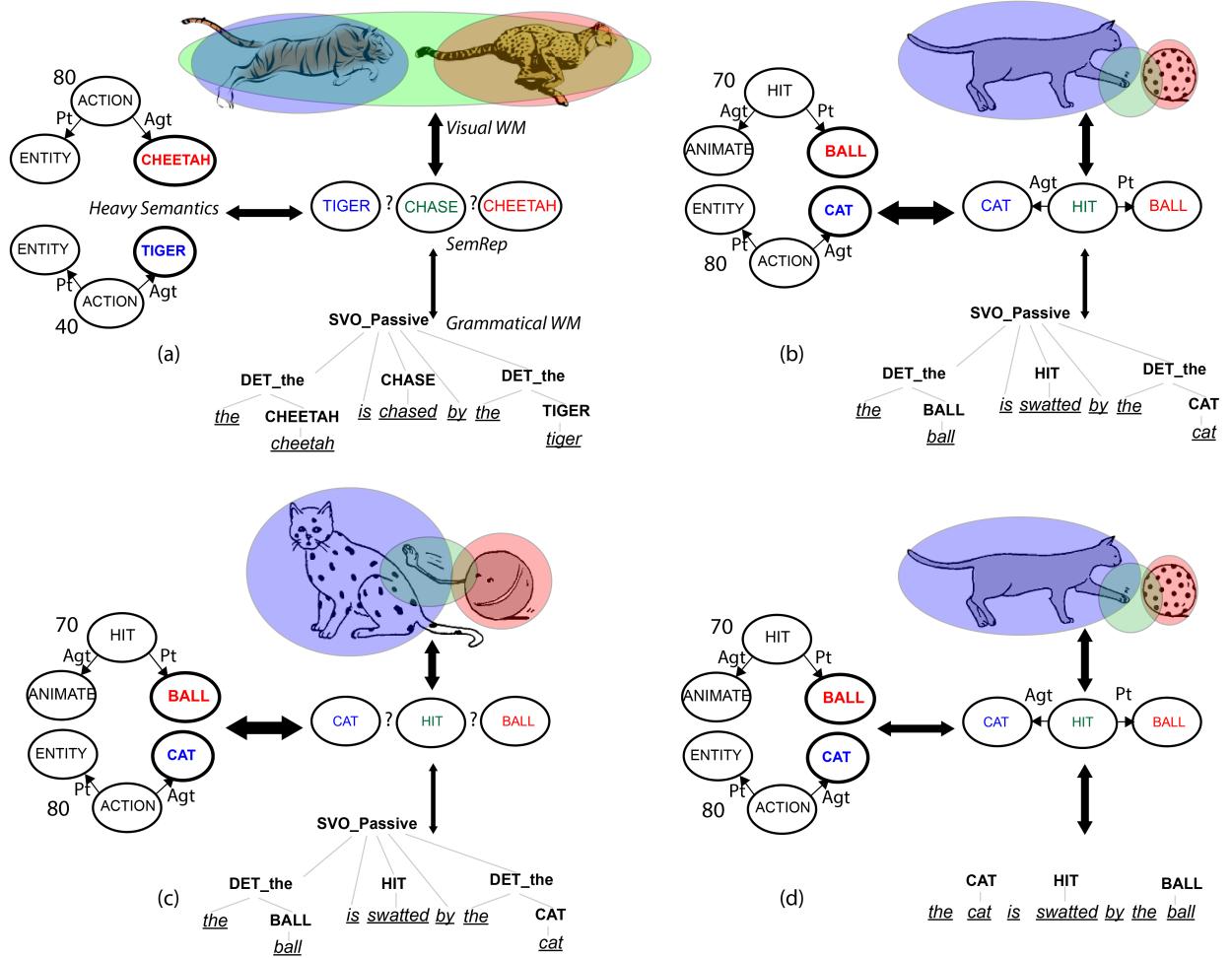


Figure 5.12: Modeling the sentence-picture matching task with SALVIA employed in comprehension mode. In each panel, the top section represents the perceptual schemas applied to the visual scene. The bottom section represents the grammatical working memory maintaining a construction assemblage generated by the grammatical route. The middle section represents the SemRep built in semantic working memory, locus of collaborative computation between the three routes (indicated by double arrows). The thickness of the double arrows represents the weight assigned to the interpretation of each route in the collaborative computation of the SemRep. Lower weights can be due to partial lesions. In panel (d), the absence of construction assemblage represents the effect of lesions that would impair the processing capacities of the grammatical working memory. Details for each panel are given in the text.

with the weakened grammatical route and allow the system to generate the proper interpretation of the utterance input. Here the visual scene matches the meaning of the input and therefore all the processes converge to the same stable interpretation, compensating the lesion to the grammatical route. Fig. 5.12 (c) presents the situation in which for a similar input utterance, the visual scene presented does not match and is counterfactual with respect to our knowledge of the world (a ball swatting a cat). The hypothesis generated by the world knowledge would cooperate with the output of the weakened grammatical route to generate a SemRep graph that this time enters in competition with the graph generated by the perceptual schemas active in visual working memory. This competition signals the mismatch between the utterance received and the visual scene.

## 5.8 Conclusion

Such examples of visual scenes in Fig. 5.12 (b) and (c) are directly drawn from sentence-picture matching tasks used to test agrammatic comprehension (Sherman and Schweickert, 1989). We see that the behavioral result of the sentence-picture matching trial rests on the complex interactions of three sources of information at the level of the SemRep. A bias towards one source of information or another can tip the cooperative computation in favor of one of the possible interpretations of the linguistic input. Discounting perceptual information while boosting the role of world knowledge to compensate for the degradation of grammatical processing simulates the role world knowledge plays in agrammatic comprehension. However, the model puts at the forefront the fact that when using a sentence picture matching task, the impact of the perceptual content of such an image on the language comprehension system cannot be fully dissociated from that of the linguistic and heavy semantic content.

For the first three examples we have focused on possible lesions affecting the link between and an intact grammatical working memory and the semantic working memory. Such lesion that would keep the grammatical process per se intact but deteriorate its capacity to impact the semantic interpretation partially echoes the conclusions of Schwartz et al. (1987) who moved away from a purely syntactic explanation of agrammatism and hypothesized that the deficit resulted from an impaired participation of the extracted syntactic information in the thematic role assignment process. Fig. 5.12 (d) illustrates the fact that the degradation of the role of grammatical information on generating the SemRep could also be simulated by a lesion limiting the computational capacity of the grammatical working memory to process grammatical constraints. This hypothesis echoes the capacity approach to agrammatism developed by (Miyake et al., 1994, 1995) who hypothesized that agrammatic comprehension is a result from a reduction, following brain lesions, of working memory resources available to compute the syntactic information contained in linguistic inputs. In this case, the connection between semantic and grammatical working memory remains intact, but the grammatical working memory is limited in the complexity of the construction assemblages it can build, a limitation that we illustrate here by allowing only word level construction to be invoked into working memory. Constructions invoked in grammatical working memory are instances of schemas and therefore represent active processes mapping their SynForm to a SemFrame but also creating links between their slots and other constructions. Since these processes involve both detecting temporal sequences of inputs or constructions that match a SynForm or assessing the match between the light semantic constraints of a slot with possible construction inputs, reduced computational capacity of grammatical memory can be generated by a variety of lesion affecting parts or whole of the process of construction matching. In particular, specific lesion to the light semantic constraints matching can be simulated, lesions that would result in the deterioration of the grammatical working memory to build stable construction assemblages while the world knowledge system remains intact.

To conclude we go back to the tripartite distinction found by (Berndt et al., 1996) in their meta-analysis of comprehension patterns for reversible sentences (agent and patient role for the entity described are equally plausible) in sentence-picture matching tasks. They found that about the same number of agrammatic aphasics were (1) at chance for both passive and active, (2) at chance only for passive and better than chance for active, or (3) better than chance for both. In our analysis of Fig. 5.12 (a) we showed how the SALVIA model of comprehension can account for the comprehension pattern (2) as emerging from the cooperative computation between a weakened grammatical route and the world knowledge route without assigning the deficit to a processing difficulty to specifically associated with the SVO\_Passive construction. SVO and

SVO\_Passive are treated equally. The lesion deficit is assigned to the capacity to use the form-meaning mapping built in grammatical working memory to update the SemRep. Such deficit results in the equal deterioration of the capacity to use the grammatical cues associated with each one of these constructions to assign relations between nodes, deterioration that can be alleviated in the case of the SVO construction only thanks to the general world knowledge hypothesis that tend to assign the role of agents to the first encountered content word (if it describes an animate entity that is usually involve in doing something). The model can explain the comprehension pattern (1) by making the hypothesis that patients showing degraded capacity to process both active and passive constructions for sentence-picture matching task could suffer from lesions affecting not only the grammatical route but also the world knowledge route. This indeed would result in a difficulty to use the “agent received first” hypothesis efficiently. As for the comprehension pattern (3) it can be accounted for by allowing for only mild lesion of the grammatical route, allowing the grammatical constraints to weigh in the final role assignment.

Finally, the explanations of the comprehension patterns (1) and (2) by the SALVIA comprehension model entail the following predictions. Patients that show good performances for active constructions even in the case where no world knowledge cues can be used, should be significantly better for sentences of the type “the tiger chases the cheetah” in which the “agent received first” hypothesis applied, than for sentence of the type “the ball hit the bat” for which the world knowledge would not instantiate hypothesis that would have BALL as an agent, removing the possibility of world knowledge to early on help building anticipations of the relations that will link the nodes generated by the content words hit and bat.

The next chapter is dedicated to turning those conceptual considerations into a computational TCG-SALVIA model of language comprehension.

# Chapter 6

## TCG-SALVIA: Formalism for Incremental and Dynamic Grammatical Processing of Utterances.

*“Quality is nothing but difference in quantity and corresponds to it each time forces enter into relation.”*

Deleuze

Nietzsche and Philosophy

### 6.1 Introduction

This chapter provides a computational counterpart to ch. 5. It also establishes computationally the validity of the SALVIA architecture as a novel model of agrammatic comprehension: the qualitatively different patterns of comprehension that have been observed in agrammatic aphasics find here an explanation in terms of dynamic cooperative computation between two concurrent parallel processing routes: a Grammatical route and a World Knowledge route. *Quantitative* differences in the relative weights that those route carry as they generate the interpretation of a linguistic input result in the emergence of distinct *qualitative* comprehension patterns.

Template Construction Grammar (TCG) as a computational construction grammar (CompCxG) framework has already been described and used as part of the Schema Architecture for Language-Vision Interaction (SALVIA) cognitive model simulating the dynamics at play during language-vision interactions in the context of online visual scene description production. This chapter presents the extension of the TCG model of language production implementation (from now on TCG\_p) to a model of language comprehension (from now on TCG\_c). Schema Theory (ST) provides guidelines to implement cognitive-level hybrid computational models that operate in style of the brain. As for production, TCG\_c extends here schema theory to language. TCG\_c share some important formal and computational characteristics with TCG\_p that won't be repeated here. Rather, this paper will focus on the element of the comprehension model that differ from production in an effort to, in the process of presenting this novel computational framework, outline some key asymmetries between production and comprehension processes, asymmetries that make TCG differ from the other implemented computational construction grammar. In contrast with the presentation of TCG\_c that was kept separate from the presentation of the SALVIA model of language production (from now on SALVIA\_p) in which it was put to use, comprehension processes are described directly within the SALVIA

architecture of language comprehension in which they are hosted (from now on SALVIA.c).

The main focus of the TCG-SALVIA for comprehension is to provide a framework to model the brain's capacity to seamlessly and dynamically coordinate the multiple sub-systems involved in situated language use: grammatical knowledge, world knowledge, and contextual (visual) information.

The empirical motivations behind the development of a concurrent multi-route comprehension system have been presented in the previous chapter (Ch. 5). They result from what is hypothesized to be the functional and anatomical organization of the language system.

## 6.2 Grammatical Route: Cooperative-Computation Driven Generation of Dynamic Construction-Based Form-Meaning Mappings

This section provides a theoretical counterpart to the informal description of the Grammatical route given in ch. 5, sec. 5.4.

### 6.2.1 Template Construction Grammar as a Schema-Theoretic Model of Grammatical Processing for Incremental Language Comprehension

The Template Construction Grammar (TCG) has been shown to provide an adequate representational support for the grammatical knowledge used in language production by SALVIA. Since the goal here consists in using the new task of comprehension not to build a new model but to expand and refine the SALVIA model, the methodology imposes to think of the new model in terms of the changes and additions that need to be introduced, the production model serving as a background on which comprehension model should be contrasted. The reader is therefore strongly encouraged to refer back to the chapters 2 and 4 as the theoretical and algorithmic choices that have emerged from the work on SALVIA production model, when necessary, will be here alluded to but not repeated.

In the rest of this section, the processes underlying the grammatical route are detailed.

### 6.2.2 Phonological WM

The Phonological WM incrementally receives word forms as input (speech word recognition is not tackled by the model). Each word form received results in the instantiation of a 'phonological' schema instance<sup>1</sup>. The state of the Phonological WM is defined as a sequence.

### 6.2.3 Grammatical Knowledge Representation

#### Template Construction Grammar

The representational formalism of Template Construction Grammar will not be repeated here as it is the same as the one used for production (see ch 4. The representation of constructional content is the same for production and comprehension, however, as the following will show, the construction schemas that are derived from them differ between the two modalities.

#### Language Schemas

Language schemas (LS) in TCG are defined as construction schemas (i.e. schema containing form-meaning mappings derived from construction representation in a given grammar).

The LS used in comprehension (LS(c)) are defined in most the same way the LS used in production were defined (LS(p)) (cf. Ch. 4).

The form pole (SynForm) of constructions in TCG representational format is converted into an active process when a language schema is derived from a TCG construction representation. The SynForm is

<sup>1</sup>'Phonological' is used here for convenience as the model's processing starts at the 'above the word' level. The phonological schema (instances) can be thought of as 'form content' schema instances.

implemented as a state sequence. This operationalization of the SynForm will enable each LS, as well as instances they spawn, to hold a ‘form’ state that is a function of the incrementally received linguistic input. It lets the LS and the LS instances generates state-dependent prediction regarding the nature of the upcoming linguistic inputs, predictions that can then be confirmed or disproved. In doing so, the construction schema instances provide a schema theoretic counterpart to the notion of active arc in a chart parser.

### Grammatical Long Term Memory

Grammatical knowledge is defined, following Schema Theory as a network of construction schemas. This stipulates the state of the Grammatical Long Term memory (GrammaticalLTM in Fig 5.2<sup>2</sup>). As in the case of production, the current work does not tackle the nature of the organization of grammatical knowledge and simply defines the grammatical knowledge as a set of construction schemas<sup>3</sup>. TCG does not make the assumption that the grammatical knowledge supporting comprehension is the same as the grammatical knowledge supporting production. Although the two do usually converge in their manifestation at a behavioral level during language acquisition - a fact that warrants regarding them as a unique entity as linguistic competence explaining objects in the fluent speaker - TCG, as a brain theory deeply rooted in question of performances, chooses to keep the two distinct with the goal to let the neurolinguistic data decide later on the proper treatment of the relations the two entertain.

#### 6.2.4 Dynamic Grammatical Processing: from Incremental Linguistic Input to Incrementally Built Semantic Representations

##### A Dynamic Incremental Parser: A Schema-Theoretic Counterpart of Left-Corner Parsing

To the author’s knowledge, Left-Corner parsing has been used by most of the existing computational construction grammar frameworks (see (Bryant, 2008) for a discussion of the use of LC parsing in Embodied Construction Grammar) and more generally by unification-based grammars (Tomuro, 1999).

Such choices have been usually driven by algorithmic considerations rather than by concerns with psychological plausibility (but for a discussion of the latter see (Resnik, 1992)).

TCG as a Schema Theory based model of grammatical processing uses LC parsing as a good algorithmic starting point to define the schema theoretic operations that fuel the incremental invocation of construction instances as well as the incremental building of the competition-cooperation network defining the dynamical trajectories followed by the activation levels of those instances. TCG does not implement an LC parser but its own parser that can however be thought of as a schema theoretic counterpart of the LC parser.

TCG representational format can be, in part, mapped onto a CFG unification grammar, and more generally onto a constraint based grammar supported by a CFG. In TCG we can start by defining,  $\forall cxn \in \mathcal{G}$  with  $cxn = (class, head, SynForm, SemFrame, SymLinks)$ :

$$\alpha = class \in V \quad (6.1)$$

$$\beta = SynForm \in (\mathcal{P}(V) \cup \Sigma)^* \quad (6.2)$$

We can associate to  $cxn$  the CFG-like production rule<sup>4</sup>:

$$class \rightarrow SynForm \quad (6.3)$$

As a form-meaning mapping, each construction also contains symbolic links  $\mathcal{SL}$  that link elements of the SemFrame to elements of the SynForm, binding the syntactic and semantic constraints.

---

<sup>2</sup>This should really be GrammaticalLTM(C) since the SALVIA model does not a priori assume symmetry between production and comprehension. Compare with the strong focus on symmetry in Fluid Construction Grammar (Van Trijp et al., 2012)

<sup>3</sup>But see (Wellens and Steels, 2011) for an implementation of grammatical knowledge as a priming network of constutions.

<sup>4</sup>Classic CFG formalisms, production rules are of the form  $\alpha \rightarrow \beta$  with  $\alpha \in V \wedge \beta \in (V \cup \Sigma)^*$  (c.f. Appendix for a formal presentation of classic left-to-right Left-Corner parsing with top-down filtering)

So all construction  $cxn \in \mathcal{G}$  can be associated with CFG-like production rule supplemented with a constraint set:

$$\begin{cases} class \rightarrow SynForm \\ class.concept \doteq Head.concept \\ \forall s \in SynForm, s.concept \doteq \mathcal{SL}^{-1}(s).concept \end{cases} \quad (6.4)$$

### Language Schema Instance: Comprehension Construction Schema Instances

Left-Corner parsing sets up an algorithmic framework. TCG provides schema theoretic parser inspired by LC. One should always keep a clear distinction between the computational, algorithmic and implementation levels (Marr, 1982). TCG shows how LC parsing can be adapted to fit with the computational and algorithmic requirements of Schema Theory, and in particular with the requirements that the parsing trajectory be in part driven by dynamic processes resting on cooperative computation. The goal is to build a brain-theory guided model, not the efficiency of computation that LC affords. However, at an algorithmic level, the notion of a bottom-up driven top-down filtered incremental process fits nicely with the core tenets of ST that insist on the constant interactions between bottom-up and top-down processes, with perception being constantly occurring on a background of goals and expectations set up by both the temporal history of the system as well as external factors.

Just as in the case of production a language schema instance results from the invocation in Grammatical WM of a construction schema, they are therefore construction schema instances<sup>5</sup>.

As schema instances, they not only contain information (the form-meaning mapping associated with the construction) but also hosts processes and states. Those differ from the ones that the construction schema instances dedicated to production host.

The state of the construction instance is simply defined as a sequence composed from top-to-bottom of the pair  $S_i = (s_i, \mathcal{SL}^{-1}(s_i).concept)$  for  $s_i$  the  $i^{th}$  SynForm element and  $\mathcal{SL}^{-1}(s_i)$  the concept of the SemFrame that is symbolically linked to  $s_i$ . The stack is therefore simply composed of form-meaning pairs, with the form being either a form content (i.e. a word form) or a form slot (i.e. a set of classes), and with the meaning being either a concept or empty<sup>6</sup>.

$S_i$  define the syntactico-semantic constraints associated with the  $i^{th}$  element of the SynForm.

The initial state of the construction instance is  $S_0$  (and corresponds to the syntactic and semantic constraints carried by the construction's left-corner).

At each time step, a construction instance can output a prediction which correspond to its state  $S_i$ . That is, the construction can post what associated form-meaning constraint it expects for the upcoming linguistic input.

The construction has a single process  $next\_state()$  that moves its state from  $S_i$  to  $S_{i+1}$  if  $i + 1 < len(SynForm)$ , to  $\emptyset$  otherwise. It also resets `has_predicted()` to False.

As in the case of construction schema instances during production, here construction instances also contain input and output ports with the input ports tied to SynForm slots and destined to receive inputs from the other children instances whose content provides an hypothesis on how the parent instances slots could be specified.

In the case of comprehension instances however, the state of the construction's state defines which unique element of the SynForm is active at each time step. For a port to be active, ie available to build cooperation links, it needs to be linked to an active SynForm slot. If the active element of the SynForm is a not a slot but a form-content element, then this specific form content defines what is expected to be received as the next input. This imposes an irreversible progression in the construction instances states. Once a transition to a next state has been triggered, the previous states of the instance are not accessible anymore (e.g. the ports associated with already used SynForm slots cannot be used anymore to build cooperation links).

Fig. 6.1 provides an example. The SYMMETRIC\_TRANS construction provides argument structure form-meaning mapping for utterances of the type "the boy and the girl shake hands". In the current formalism it assigns to both conjoined participants the role of "AGT+PT" making each both the agent and

---

<sup>5</sup>For clarity, I will when necessary, abbreviate as LSI\_c the language schema instance for comprehension to clearly distinguish them from those used in production (abbreviated LSI\_p).

<sup>6</sup> $\mathcal{SL}^{-1}$  is defined as a partial function, but, without loss of generality, when necessary, it is extended into a total function by giving it the value  $\emptyset$  where it is not defined.

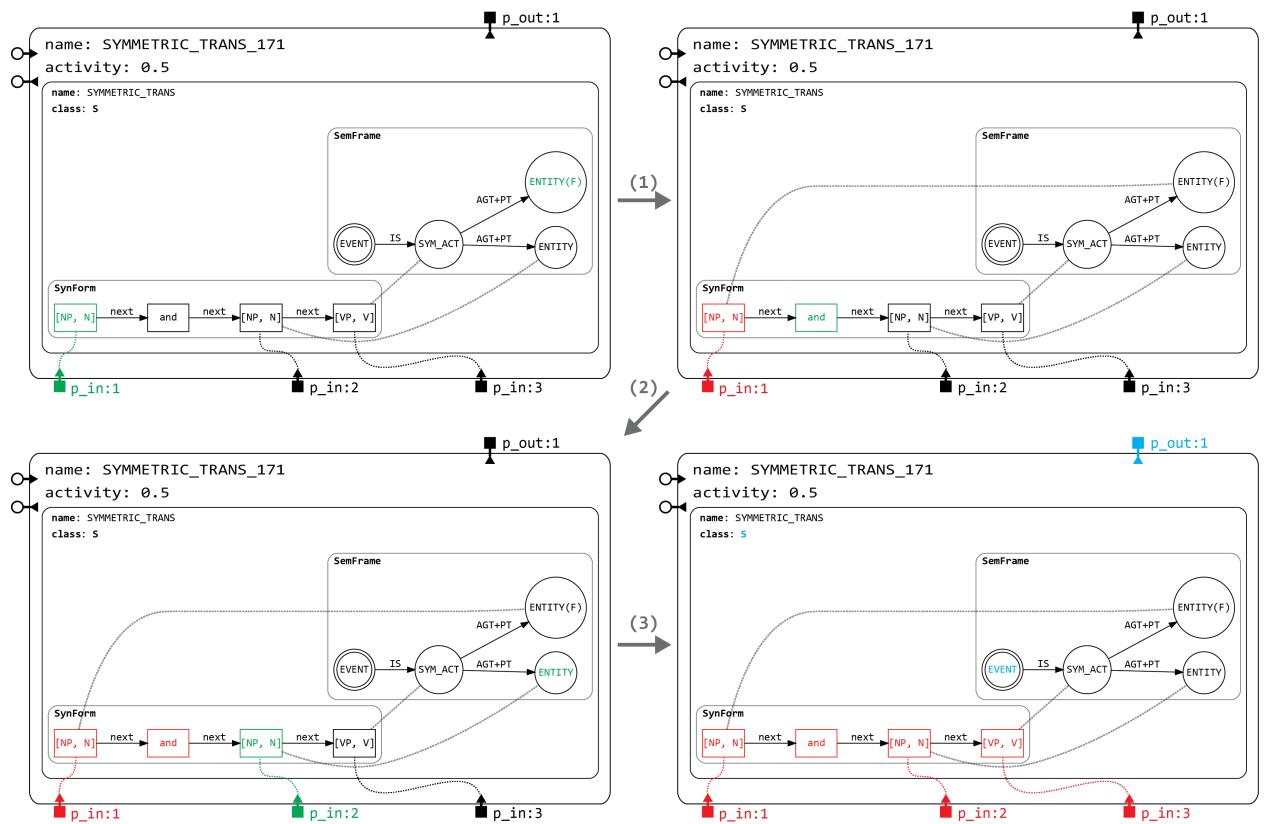


Figure 6.1: Examples of various states for a SYMMETRIC\_TRANS construction instance (Here the processes are describe for the case of pure bottom-up, without top-down filtering). See main text for details.

patient of the action. (Top-Left) Here the construction instance is in its initial state. The active SynForm element is shown in green. Since it is a slot element, also shown in green is the active input port it is associated with. And since it is associated with a symbolic link (dashed line) the symbolically associated conceptual content is also shown in green. In this initial state, the instance will attempt to link with the active output port of a child construction instance through the port p1, with the child being of class N or NP and with head of the child being compatible with the ENTITY conceptual requirement. (Top-Right) If such linking succeeds, the construction switch to its next state. In red are shown the elements that are now closed and cannot be used for the purpose of C2 processing anymore. The active SynForm element is now a form content element, setting up the expectation of “and” as the upcoming input. (Bottom-Left) If “and” is received, a new state change occur. The situation is now fairly similar to that of the (Top-Left) but for the second conjoined element. (Bottom-Right) Having skipped the VP state, here the construction instance is in its final state. All the elements of its SynForm have been linked to children construction instances are matched bottom-up with linguistic inputs. At this point the construction instance’s output port becomes active and can attempt to link to a parent construction that would have an active input slot whose associated syntactic and conceptual constraints would match the syntactic class “S” and head “EVENT” of this instance (in blue).

The key message is that the compared to the construction schema instances used in production, the behavior of the construction schema instances used in comprehension is dependent upon the state of their SynForm, a state that is therefore inherently linked to the irreversible sequential (‘left-to-right’) nature of the linguistic input. In production, the message to be communicated (SemRep) was also built incrementally, but the incrementality did not find a counterpart in irreversible state changes in the construction instances.

Linking this formalism to chart parsers (and LC parsers in particular), the change of state in a construction instance can be seen as a counterpart to moving the ‘dot’ to the right in a chart parsing edge (see ch. D). The role played by the indices associated with chart’s edges, determining the segment of the linguistic input that they cover, is, in this schema theoretic implementation, is achieved by relying on the cross WM links that exist between construction instances and the elements of the PhonologicalWM they cover (see below).

### Incremental Retrieval of Construction Schemas

The retrieval of construction schema is split in two different stages: the retrieval of construction schema based on form content i.e. the retrieval of construction on the basis of direct bottom-up lexical evidence, and the retrieval of construction schema based on the combination of bottom-up and top-down predictions<sup>7</sup>. In doing so TCG explicitly operates a distinction between lexically driven and non-lexically driven constructional processes (and hence a distinction in the treatment of lexical vs abstract constructions.)

The input received by the system is a temporal sequence of word forms noted ( $w_t$ ) where  $t$  is the time at which the word form  $w_t$  is received.

**Lexically Driven Construction Schema Retrieval** In Fig. 5.2, this corresponds to the processes supported by LexCxnRetrieval(C). Given a new phonological instance phon.inst invoked from PhonologicalWM(C), the LexCxnRetrieval(C) returns the set  $CS_{lex}$  of construction schemas whose initial predictions match the phon.inst content<sup>8</sup> as well as the  $BU_{lex}$  set of all the (class, head.concept) associated with those.

This set  $BU_{lex}$  corresponds to the set of bottom-up predicted syntactico-semantic types. It is passed on to CxnRetrieval(C) where it is used to shape the bottom-up driven instantiation of abstract constructions schemas.

**Non-Lexically Driven Construction Schema Retrieval** In Fig. 5.2, this corresponds to the processes supported by CxnRetrieval(C). CxnRetrieval(C) receives both  $CS_{lex}$  and  $BU_{lex}$  as well as phon.inst. It recursively builds the set  $CS$  of all construction schemas whose initial prediction intersect with the set  $BU$ <sup>9</sup>. CxnRetrieval(C) recursive process is initiated with  $CS_0 = CS_{lex}$  and  $BU_0 = BU_{lex}$ . For each recursive pass  $i$ , the construction that have not yet been retrieved and whose initial prediction intersect  $BU_{i-1}$  are

<sup>7</sup>See operations (4) and (5) of the LC parser in Appendix D

<sup>8</sup>This corresponds to checking for RHS of production rule matching form input, cf. op5 in ch. D

<sup>9</sup>This is a counterpart to the retrieval of all the construction schemas whose left-corner match the right hand side of a construction that has just been retrieved, see op4 in ch. D

added to  $BU_i$  and their type is added to  $CS_i$ . The process stops when  $BU_n = \emptyset$ . CxnRetrieval(C) sends to Grammatical WM the set of construction schemas to be instantiated  $\bigcup_k CS_k$  as well as phon\_inst.

### Grammatical WM: Incrementally Building the C2 Network

**State** At each time step, the state of the Grammatical WM consists in a cooperation-competition network (C2 network) of active construction schema instances, dynamically self-organizing to incrementally build a mapping from the input onto meaning, updating the current state of the semantic representation (SemRep).

The C2 network is composed of cooperation and competition links (coop\_link and comp\_link respectively), and as a whole define the dynamic system that control the temporal trajectory of the construction instances' activation values.

In addition, in order to implement the LC parser, the state of the Grammatical WM in the case of production is augmented with a *state* value that is modified each time a new phon\_input is received. It allows the GrammaticalWM to have a unique signature for each of its state following the reception of new input, a feature necessary to implement chart parsing.

**Updating C2 Network** At each time step, the state of the Grammatical WM (C2 network of construction instances) can be seen to reflect a dynamic version of a chart parser state. New cooperation and competition links are built each time a new construction instance is invoked.

Borrowing from the Earley parser terminology, the *Scanner()* method, upon receiving from CxnRetrieval(C) a new value of phon\_inst, checks the phon\_prediction() made by each of the construction instance currently active in Grammatical WM. If the instance in its current state predicts a word form input (i.e.  $phon\_prediction() \neq \emptyset$ ), then two situations ensue.

Either this prediction matches the bottom-up received content of the phon\_inst in which case the construction instances enters its next state and a cross WM cooperation link between the phon\_inst and the construction instance is built: The construction instance will receive external supporting activation from the phon\_inst.

If rather this prediction is contradicted by the content of the phon\_inst, then the construction instances is directly marked as dead (it has been disproved by direct bottom-up input) and will be pruned out of the Grammatical WM state<sup>10</sup>.

All the construction instances that have yield predictions matching the phon\_inst form content enter in competition (comp\_links are added between each pair of construction instances in this set). This reflects the fact that they all correspond to competing grammatical hypotheses regarding the mapping of this form content onto a meaning frame.

The second important process, with the terminology here again borrowed from Earley parser, is the *completer()* process.

Simply stated, the completer checks for construction instances that have reached their end state (i.e., in LC parser terms, completed arcs). For each of those construction instances  $inst_c$ , it gets their type (class, head.concept) and checks for all incomplete construction instances  $inst_i^k$  whose coverage of the phon\_inst sequence ends where the coverage of  $inst_c$  ends and whose current state generate a prediction compatible with  $inst_c$ 's type. For each construction instance  $inst_i^k$  that fits such requirements set by  $inst_c$ , a cooperation link is created linking  $inst_{complete}$  to the input port of  $inst_i^k$  mapping onto the SynFrom slot that defines the state of  $inst_i^k$ .  $inst_i^k$  is set to cover now the union of the phon\_inst sequence it used to cover and the phon\_inst sequence covered by  $inst_c$ .<sup>11</sup>

All the  $inst_i^k$  enter in mutual competition since they all represent competing hypotheses for top-down predictions matching the bottom-up data defined by  $inst_c$ 's type. In addition, they all move to their next state. This makes the *completer()* process recursive: it is applied as long as its previous application led construction instances to reach their end states.

**TD Predictions** So far I have not mentioned the role of state dependent TD predictions in filtering the BU driven construction schema instantiation.

---

<sup>10</sup>A smoother and more lax version simply imposes a competition link between the phon\_inst and the construction instance.

<sup>11</sup>This is a counterpart to finding incomplete chart arcs whose left-most daughter matches the mother of the production rule associated with the completed arc and updating the chart arc indices, cf rule op2 in ch. D, but without top-down filtering.

At each time step, the Grammatical WM sends to both the LexCxnRetrieval(C) and CxnRetrieval(C) the set of all syntactico-semantic pair of prediction constraints generated by the currently active construction instances. This set defines the state dependent top-down constraints imposed on the incremental process of construction schema retrieval. The current model can be set to use such TD filtering or not.

In the model we present, the TD predictions were not used to constraint the BU retrieval in order to maximize the flexibility of the system's capacity to parse fragmented utterance that include restarts. However, the proper integration of top-down and bottom-up information in the retrieval of grammatical knowledge is a subject of much debate in psycholinguistics and future work should focus on address this point.

### **Cooperative Computation (C2)**

**C2 Dynamics** The Cooperative Computation dynamics taking place in Grammatical WM (and in any WM of the model), is described by the same set of leaky-integrator equations that are used in described in the TCG production model (see ch. 4)

Suffices here to say that the C2 network of construction instances forms an incrementally built network of leaky integrator units (the instances), with the cooperation and competition links forming the WM-internal weighted inputs, while the cooperation (and competition links) built across WM (e.g. between the Grammatical WM instances and the Phonological WM instances) form the WM-external weighted inputs. It weight associated with those WM-external inputs is defined by the model's connectivity parameters: each connection between sub-systems is associated with a weight. This will become important below when multiple external activation inputs concurrently support sets of competing and cooperating instances.

#### **6.2.5 Generating Meaning**

##### **Construction Instances Assemblage and Assemblage Unification**

Sets of cooperating construction instances form an assemblage. Following the same principles of unification as those used for production, an assemblage  $A$  of construction instances can yield a assemblage equivalent construction instance  $eq\_inst_A$  corresponding to the unification of all the pairs of cooperating construction instances in  $A$ . The process of unification of the assemblage  $A$  returns  $eq\_inst_A$  as well as a mapping  $\mathcal{M}_A \subset eq\_inst_A.SemFrame \times \{inst.SemFrame\}_{inst \in A}$ , pairing each element of  $eq\_inst_A.SemFrame$  with the SemFrame elements from where it was constructed during unification<sup>12</sup>.  $\mathcal{M}_A$  plays an important role in building the cross working memories cooperation links between Grammatical WM and Semantic WM (see below).

##### **Updating SemRep**

The Grammatical WM, following ST, is associated with a confidence threshold  $\theta_{conf}$ . As soon as an assemblage  $A$  has a score  $> \theta_{conf}$ ,  $eq\_inst_A$  as well as  $\mathcal{M}_A$  is send to the Semantic WM where  $eq\_inst_A.SemFrame$  is used to update the state of the SemRep. All the construction instances of  $A$  are then marked as *expressed*. Only the assemblages that contain non-expressed instances are considered as a basis to generate meaning.

### **6.3 World (event) Knowledge Route (WKR): Generating Dynamic Frame-Based, Pragmatically Motivated, Form-Meaning Mapping**

This section provides a theoretical counterpart to the informal description of the World Knowledge route given in ch. 5, sec. 5.5.

---

<sup>12</sup>Unification in TCG is non-destructive.

### 6.3.1 Phonological WM: Interaction with World Knowledge

The Phonological memory incrementally invoked the phonological instances corresponding to the word inputs. Its interactions with LexCxnRetrieval() play the same role as in the previous route. Lexically anchored constructions are retrieved. However, for this sub-system, LexCxnRetrieval() in addition to triggering the invocation of the corresponding instances in GrammaticalWM, also sends the conceptual knowledge associated with those lexical constructions to the Frame LTM. This conceptual knowledge consists in the set of all the concepts associated, for a given lexicalized construction schemas, with the nodes of its SemFrame:  $SemFrame \rightarrow \{n.concept\}_{n \in SemFrame.nodes}$ . For a given lexicalized construction, the structure of the semantic content is lost (graph to set), and of the concept tokens only their types are kept.

LexCxnRetrieval() sends, alongside this set of conceptual content, the phon\_inst that triggered the retrieval of this knowledge.

### 6.3.2 WK Event Frame Schemas

World knowledge event frames extend the world knowledge content given to the system. Event frames are defined as (trigger, WK\_Frame graph) pairings.

The WK\_Frame graph stands for the meaning of the frame and follows the same general graph structure used to express the semantic representations (as in SemRep and SemFrame). The WK\_Frame, in the current implementation are all event frames and therefore stipulates world knowledge based expectation regarding the structure of an event (the entities, actions and roles that are likely to be put at the forefront of the event conceptualization.)

The trigger is defined as a set of concepts. It simplifies the complex cognitive processes that result in the retrieval and use (invocation) of a event frame to structure event comprehension. Fig. 6.2 provides examples.

### 6.3.3 WK Frame Retrieval

As a concept is received from LexCxnRetrieval(), all the WK\_frame schemas that contain this concept in their trigger set are retrieved but each one will remain in a subliminal state until all the concepts that are in its trigger set have been received. At which point, the WK\_frame schema is instantiated in WorldKnowledgeWM where it build cross WM cooperation links with all the phon\_insts that have played a role in generated the concepts that triggered the instantiation. See Fig. 6.2 for examples.

### 6.3.4 WK Processing

The dynamic processes taking place in WorldKnowledgeWM are kept very simple in the current implementation. There is no cooperative computation. This is tantamount to saying that the model, for now, does not implement pragmatic reasoning that would involve the combination of WK\_frame instances (in a way that is similar to the processes taking place in GrammaticalWM). This choice stems in part from the desire to keep the model simple, developing a C2 fueled reasoning system is beyond the scope of this paper. More importantly, this choice also reflects that decision to focus on the process of fast online world knowledge based *semantic enrichment* of semantic representations as opposed to full-blown *inference processes*. Inferencing is always part of comprehension but can be distinguished from enrichment with the latter being fast, coarse and event oriented while the former is usually slower, precise, and can tap into the entirety of the pragmatic knowledge (Altmann and Mirković, 2009).

### 6.3.5 Generating Meaning

Since the WorldKnowledgeWM does not host cooperative computation between WK\_frame instances, there is no self-organization of complex composed world-knowledge based form-meaning mappings.

Each WK\_frame instance's activation is in part driven by the activation it receives from the phon\_insts that participated in triggering its instantiation. Any WK\_frame instance whose activation value rises above the WorldKnowledgeWM confidence threshold will pass its WK\_frame information to the SemanticWM to update the SemRep in a process that is equivalent to that of a (assemblage equivalent) construction instance updating the SemRep by passing its SemFrame information to the SemanticWM. The WK\_frame

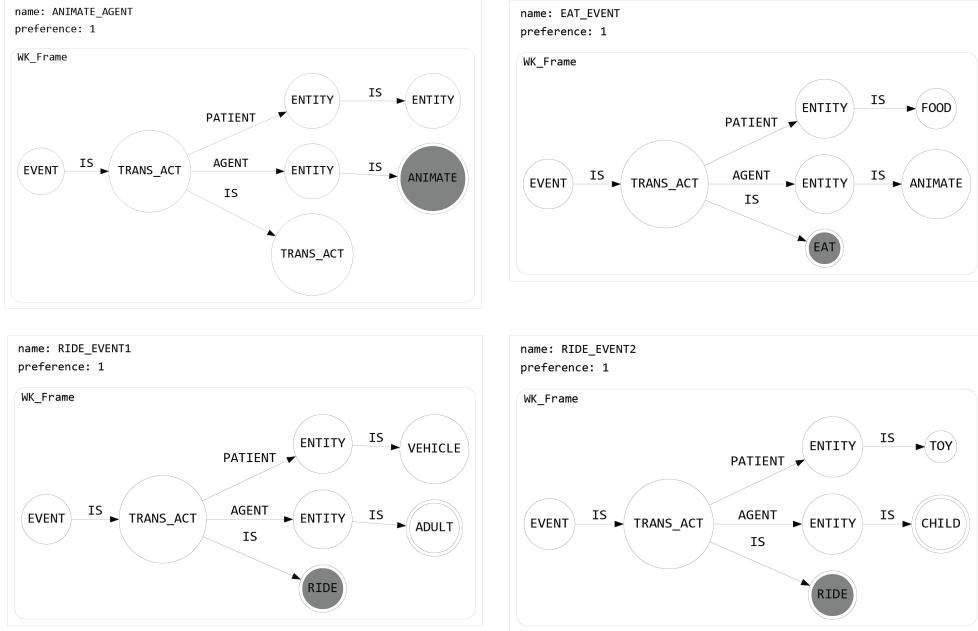


Figure 6.2: Four examples of World Knowledge Frames (WK\_Frames) schemas. A WK\_Frame schema is defined as a concept schema graph structure. It shares most of its representation convention with that of construction SemFrame (Semantic Frame). Importantly, in the case of WK\_Frames schemas, double-circle nodes indicate the concepts that are part of the WK\_Frame schema concept trigger set: a WK\_Frame schema instance is invoked in World Knowledge Working Memory (and hence participate in building a WK based form-meaning mapping) only once every concept of the trigger set has been activated as part of some of the world knowledge meaning associated with the input's lexical items. (Top-Left) Animate Agent Frame. An instance of this WK\_Frame schema only has ANIMATE as a trigger and carry the pragmatic assumption that ANIMATE referents tend to be agents (here of transitive actions) (see below for more discussion of this frame). (Top-Right) Eat Event Frame. Retrieved in association with any action concept subsumed by EAT concept, stipulates that the agents of eating events are usually animate while the patients are usually food items. (Bottom-Left and Bottom-Right) Those two bottom frames represents two variations on the Ride Event Frame, they are not triggered by a RIDE concept alone, the left one also requires a concept of the ADULT type, while the right one requires a concept of the CHILD type. They respectively stipulates that adults tend to ride types of VEHICLES while children tend to ride types of TOYS. Having two frames is a simplification of a more complex process that would first retrieve an ambiguous ride frame before specifying the type of patient as more information is received based on inference/composition processes that are not modeled here.

content can add information to the SemRep (grow the SemRep graph), corroborate already present semantic content (bolster the activation of parts of the SemRep through cross WM activation links), or results in new competitions taking place within the SemRep (adding competition between edges). The WK\_frame instances also receive activation from the SemanticWM instances, resulting in a activation loop between the WorldKnowledge WM and the SemanticWM (Flow of activation not shown in fig. 5.9)

## 6.4 Dynamic Interactions Between Grammatical And World Knowledge Route Routes as Cooperative Computation Between Concurrent Processes

This section provides a theoretical counterpart to the informal description of the multi-route cooperative computation architecture given in ch. 5, sec. 5.6. This operates an important change in the processes hosted by the Semantic WM as it becomes a locus of Cooperative Computation (cf. ch. 5, sec. 5.6.1)

### 6.4.1 Good Enough Comprehension of Utterances: Speaker and Task Relevant Parameters

The SALVIA comprehension model follows in the footstep of the production model and defines speaker and task relevant parameters. Those parameter allows to set up the model's state so that the computation reflects some aspects of a language user's idiosyncrasies and the impact that task requirements might have on the deployment and organization of the cognitive resources. The reader can refer back to the section on "good-enough production" in ch. 4, sec. 4.5. The SALVIA model of comprehension, modulo some obvious appropriate modifications, follow the same approach for the implementation of Good Enough Comprehension. That is with one major exception: that of the existence of a new set of parameters highly relevant to the good-enough process and that are absent from the production model: the weights of concurrent parallel processing routes. It is on this parameter space that the following section will focus.

### 6.4.2 Semantic WM C2 Dynamics: The Key Role of Route Weights

The C2 dynamics taking place within Semantic WM follows the computational requirements of Schema Theory which won't be restated here. The key computational question lies in the relations between external and internal C2 factors.

The full descriptions of the cooperative computation dynamics can be found in ch. A. However, in order to make the following discussion clearer, the key parts of the dynamical system equations relative to the role of C2 weights are stated here, focusing on Semantic WM.

For a concept schema instance  $i$ , active in a Semantic WM as part of C2 network, its activity  $Act_i^t$  is updated following a leaky integrator equation:

$$Act_i^{t+1} = \alpha Act_i^t + (1 - \alpha)\sigma(Input_i^t + noise^t) \quad (6.5)$$

$\alpha$  defines the characteristic time of the system,  $\sigma$  is a logistic function, and  $noise^t$  is a Gaussian noise.  $Input_i^t$  is defined as:

$$Input_i^t = w_I \left\{ \sum_{k \in comp(i,k)} w_{comp} \cdot Act_k^t + \sum_{j \in coop(i,j)} w_{coop} \cdot Act_j^t \right\} + w_{ext} \{ w_{gram} \cdot Ext_{(gram,i)}^t + w_{wk} \cdot Ext_{(wk,i)}^t \} \quad (6.6)$$

$Ext_{(e,i)}^t$  represents activation that an instance  $i$  receives from outside the working memory by subsystem  $e$ , with  $gram$  standing for Grammatical WM, and  $wk$  standing for World Knowledge WK.

## Internal C2 Factors

The internal factors governing Semantic WM C2 are the same as those governing the C2 within, for example, the Grammatical WM. Of particular importance are the weights of the cooperation and competition links shaping the internal dynamic C2 network ( $W_{comp}$  and  $W_{coop}$  in eq. 6.6). However here, the strengths of internal cooperation and competition have to be analyzed in their relation to the external C2 factors (which can be independently defined by the relation between  $W_I$  and  $W_E$ , in eq. 6.6) (see also below).

## External C2 Factors: Route Weights

When a schema theoretic model is defined, connection between sub-systems used for message passing are also associated with a weight parameter characterizing the relevance of the information sent from the output sub-system for the input sub-systems. Those weights, as in any network, are to be interpreted not in isolation but in their mutual relation within the graph of processing systems.

In the present case, the situation is at its simplest since only two parallel routes are considered. The relative weight of the routes is a key parameter determining the system's dynamic behavior (relation between  $w_{gram}$  and  $w_{wk}$  in eq. 6.6, also see below).

It also impact that role of internal C2 factors. Three regimes emerge depending on the relative order of magnitude of internal vs. external C2 factors ( $W_I$  vs.  $W_E$  in eq. 6.6 given the other weights fixed).

- External  $>>$  Internal results in a situation in which the internal competitions will not be able to resolve the possible conflict emerging from differences in the semantic interpretations generated by each route. The Semantic WM will simply be a repository of potential interpretations.
- Internal  $>>$  External results in a situation in which the internal competitions will act (quasi) autonomously to generate solve interpretation conflicts: all the hypotheses (instances) are weighed equally, regardless of the process they stem from. The only route-linked factor that can impact the dynamics is the characteristic time of the route-defined processes. (see below).
- Internal  $\sim$  External results in the situation in which the interactions between internal and external processes yields results that can only be analyzed computationally. This is the state the model will always be placed in.

## External C2 Factors: Route Characteristic Times

The relation between the characteristic times of the dynamic processes associated with each route (v.i.z. with the Grammatical and WorldKnowledge WM), impacts the C2 processes resulting from their interactions. Just as above, those needs to be analyzed in relation to each other as well as in relation to the time characteristic of the Semantic WM. The regimes emerging from those relations won't be detailed here. In the present work, the model will be placed in a situation in which all the characteristic times are the same. This choice is purely motivated by the desire to focus on the impact of route weights, leaving the analysis of the impact of characteristic times as well as the analysis of the complex interactions between time and weights for future work.

## Feedback Propagation of Activation

If both the Grammatical and the World Knowledge routes build cross-WM links through which they can support the activation of subsets of Concept schema instances. The opposite is also true.

Feedback propagation of activation from the Semantic WM to the Grammatical and World Knowledge WM establish an indirect flow of activation signals between those two concurrent processes.

### 6.4.3 Dynamic Coordination of Multiple Information Sources: Sub-Multigraph Isomorphisms

Compared to production, the existence for comprehension of multiple concurrent, asynchronous, form-meaning generating processes results in a new coordination challenge.

Just as in production, for comprehension, when the two routes are considered independently, the first coordination challenge is that of the dynamic and flexible coordination between the state of each form-meaning mapping generating WM (Grammatical and WorldKnowledge WM) and both the states of the Phonological WM and Semantic WM between which it mediates. This is solved by using dynamic C2 computation principles within the form-meaning mapping generating WMs.

In comprehension, however, a second coordination challenge emerges since the information from two concurrent processes need to be integrated: the state of the Semantic WM is updated by both the Grammatical and World Knowledge Route, each asynchronously generated hypotheses regarding the nature of the semantic representation.

To handle this issue, the SemMatch operation that was defined for the purpose of production (cf. Ch. 4), is generalized. This generalization introduces a significant increase in algorithmic complexity, but opens new possibilities, not simply for comprehension but also for production.

**SemMatch generalized (FrameMatch)** As a reminder, the SemMatch operation in the case of production consisted in using subgraph isomorphism to trigger the instantiation of a construction schema if its SemFrame (source graph  $G_s$ ) matches a subgraph of the current SemRep (target graph  $G_t$ ), indicating that the construction schema offered a hypothesis for the mapping of this matching SemRep subgraph onto a linguistic form.

The first change in the current framework is that the SemRep is no longer a digraph but a multidigraph. This does not per se require to change the nature of SemMatch but it is worth noting that since subgraph isomorphism algorithms are NP, any growth in the target graph size imposes a serious computational penalty.

The second and significant change emerges from the necessity of SemMatch to handle partial matching, allowing for both construction schemas and WK\_frame schemas to generate hypotheses regarding the nature of the semantic representation and not simply confirm or disprove it.

Since the situation is symmetric between both the Grammatical and World Knowledge WM, the former will be used to discuss this problem and the solution.

At each time step, construction schema instances active in Grammatical WM can form assemblages that reach the confidence threshold and attempt to update the SemRep.

The information contained in the SemFrame of the (assemblage equivalent) construction instance will update the SemRep by possibly adding semantic content (growing the graph). Therefore finding the places where the SemFrame information should be applied to update the SemRep state cannot be tested by looking for subgraph isomorphism from the SemFrame ( $G_s$ ) onto the SemRep ( $G_t$ ): Only a partial (and possibly null) portion of the SemFrame might match onto the SemRep, the non-matching part denoting the information to be added.

SemMatch can therefore be generalized. Instead of finding isomorphism between the whole SemFrame ( $G_s$ ) and subgraphs of the SemRep ( $G_t$ ) (subgraph isomorphism), it needs to find *subgraphs of SemFrame* that match subgraphs of the SemRep. More precisely: SemMatch needs to find the **set of the maximum subgraphs of the SemFrame that match subgraphs of the SemRep** (maximum in the sense of inclusion).

The computational complexity of this operation is of course NP and now in addition depends not just on the size of  $G_t$  but also on the size  $G_s$ .

However, generalizing SemMatch in this way allows to instantiate schemas that partially match the Semantic WM states and therefore introduce predictions in the system:

**In production**, construction schemas can be instantiated as their SemFrame is considered to be a good prediction of what the future state of the SemRep will be, or even can actively shape the updating of the message.

**In comprehension**, the same principle can be applied to instantiate construction instances on the basis of the SemRep state rather than on the basis of the incoming linguistic input. Since the SemRep state reflects also the hypotheses resulting from applying world-knowledge based processes, this results in having semantically instantiated construction instances that can predict upcoming linguistic forms on the basis of world-knowledge expectation of meaning.

**Maximum Partial Sub-Multigraphs Isomorphisms Update** The generalized SemMatch needs to allow for the possibility to not only find an initial set of mappings between a information carrying frame and the SemRep, but also to update these mappings as the states of the frame or of the SemRep changes. This is handled by constraining the search for maximum partial sub-multigraphs isomorphisms to the sub-space of isomorphisms that intersect with an initial set of isomorphisms.

The solution considered are, for each already existing mapping, the set of maximum partial mappings that include this pre-existing mapping.

This can be seen as an extension of the utterance continuity principle that was used in production to a “semantic continuity principle”: in production the system attempted to generate form outputs that were smooth continuation of what had already been uttered, here the system attempts to generate continuations of SemRep that provide a smooth continuation of its previous state.

**Co-Reference Constraints** Another core set of constraints on the way a given frame (WK\_frame or SemFrame) can update the SemRep are carried by co-reference requirements. Co-reference resolution is beyond the scope of the present work and is not treated in any depth by the current model (c.f. (Bryant, 2008) for an in depth analysis, in the context of ECG of the issue of co-reference resolution). The co-reference will be required to be unequivocal. This is a very strong simplification that will limit the type of utterances that can be parsed by the SALVIA system under the multi-route architecture.

**Concept Updating** A final important change introduced by the need to organize the cooperative computation between Grammatical and World Knowledge route is that the SemRep growth is not restricted to be monotonous anymore: concept instances can be updated (the concept they carry can be changed) as a result of new information carried by one of the routes.

**Summary** By tackling updating of the SemRep by multiple concurrent processes, the Semantic WM and its information carrying structure (SemRep) faces all the challenges that most of the “blackboard-like” systems face. Although there are solutions to optimize blackboards architecture, the fact that the present model limits itself to work within the framework of Schema Theory makes that such optimizations are either unavailable or their introduction would muddy the schema theoretic nature of the model. As it should be clear by now, the purpose of this model is not to offer an optimized architecture but a functional one that highlights the benefits and challenges of implementing language models according to computational brain operating principles.

## 6.5 Input-Output

**Input** An input is simply defined by a *string sequence* (each string representing a word form), *(rate, std)* values that define the uniform distribution from which are drawn the time intervals that separate to word form being fed to the comprehension model.

**Output** The output of the model is the incrementally built SemRep (and in particular the final one, representing the final semantic interpretation on which the system has settled). The full output contains both the time stamped series of SemRep graph states, but also the activation trajectory for each concept schema instance. A lighter version of the output only records the first half of the full output. The ISRF writer can be used to store the output in a text file that can then be used as an ISRF input to the production model (the ISRF writer translates the state of the Semantic WM into a ISRF formula, see Appendix 4, sec. C.1.5).

## 6.6 From Conceptual To Computational Examples

### 6.6.1 Sanity Check

In order to ensure that the model was performing correctly in an idealized “competence-like” set-up, it was tested with inputs ranging from single NP to single clause, embedded relative clauses (center and final

embedded, subject and object relative), each with transitive, intransitive, or ditransitive verbs, up to a level of complexity involving one of those main sentences conjoined with a second subordinate sentence through a discourse causal marker (see sec. 5.4.2 for an example.)

A ground truth was associated with each input in the form of a target SemRep (unique since the sentences were unambiguous). The model was able to generate the correct SemRep in the case of the Grammatical Route only situation.

To test the World Knowledge route, using simpler single transitive sentences, a ground truth was defined regarding the expected thematic role assignment (TRA) and the sentences were marked as counterfactual, pragmatically expected, or neutral with respect to world knowledge. No sentences were ambiguous with respect to the world knowledge. In absence of the grammatical route, the world knowledge route systematically correctly assigned the TRA for pragmatically expected sentences, incorrectly assigned the TRA for counterfactual sentences, and was unable to provide TRA for the neutral sentences.

These results should only be taken as a sanity check indicating that the model is able, from a pure competence perspective, to deliver what it was designed to deliver.

### 6.6.2 Simulation 1: Grammatical Route Only

SALVIA was run to generate a computational simulation for the grammatical route only situation. This simulation corresponds to a slightly different example than the conceptual example presented in ch. 5, sec. 5.4.1. The conceptual example focused on generating a form-meaning mapping for a pragmatically expected sentence. Here, the model is given the counterfactual input sentence, “*the boy is eat -ed by the cake*”. Using the grammatical cues only, SALVIA recovers the proper counterfactual semantic representation. This is done in absence of other constraints and with all constructions considered equal in preference (in particular, the active voice is not given a higher preference than the passive voice construction), and there are no time constraints or WM capacity limitations.

Although the impact of those factors would be important to understand even in the context of normal comprehension, future work will analyze their relations to good-enough comprehension results in normal subjects (see for example Dick et al., 2001)

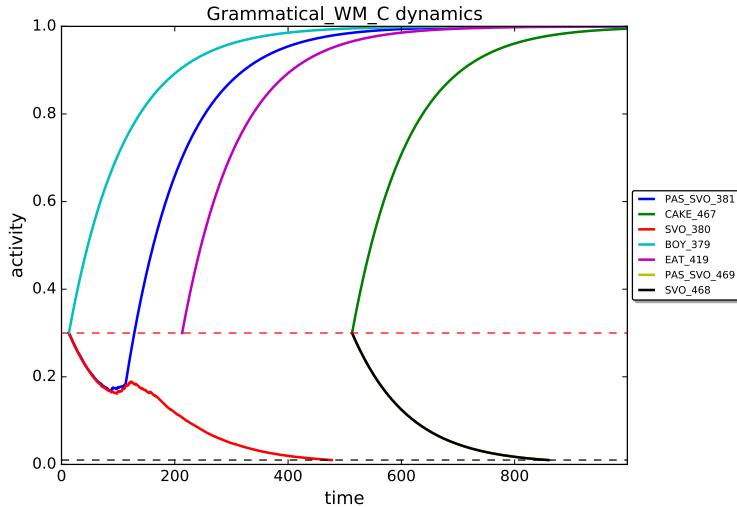


Figure 6.3: Activation levels of the construction instances active in Grammatical WM as a function of time. Top dashed line corresponds to the confidence threshold. Bottom dashed line marks to the pruning threshold. See Appendix B, sec. E.2 for representations of the Grammatical WM states over time. See main text for details.

Fig. 6.3 shows the activation levels of the construction instances invoked in Grammatical WM as a function of time for the system processing the input “(the) boy is eat -ed by (the) cake”.

At  $t=0$  the first word input is received (“boy”). This triggers the invocation in Grammatical WM of the lexical construction instance BOY\_379, but also of the two argument structure construction instances SVO\_380 and PAS\_SVO\_381 corresponding to predictions of an upcoming transitive action event described using the active or passive voice respectively. Those two argument structure instances cooperate with BOY\_379 but compete with one another. For this reason, BOY\_379’s activation level quickly rise while at first both SVO\_380 and PAS\_SVO\_381 see their activation level decrease due to the effect of mutual inhibition.

As the next input “is” is received at  $t=100$ , this boosts the PAS\_SVO\_381 hypothesis. The symmetry between the two argument structure construction instances is broken and PAS\_SVO\_381 emerges as the winner. At  $t=200$  “eat” is received triggering the invocation of the EAT\_419 construction instance whose activation quickly rise. Later on “-ed” and then “by” are received, boosting again the passive voice instance. The SVO instance loses the competition and gets pruned out of the Grammatical WM. Finally “cake” is received, triggering the retrieval of the CAKE\_467 lexical instance. Two new SVO and PAS\_SVO instances are also invoked at this point, setting up the state for the possibility that a new utterance has began, with the former input being left unfinished (partial utterance). Since this is not the case here, those will simply decay and be pruned out.

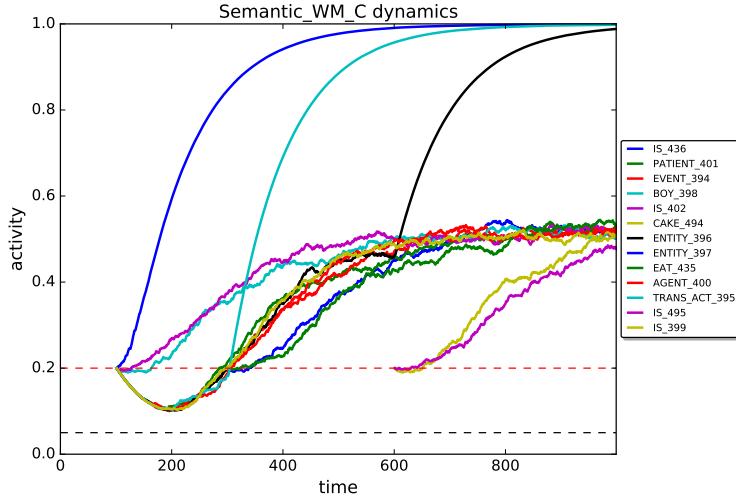


Figure 6.4: Activation levels of the concept instances active in Semantic WM as a function of time. Dashed line marks to the pruning threshold. See Appendix B, sec. E.2 for representations of the Semantic WM states over time. As the activations traces show, the SemRep is updated at various time points that do not correspond directly with the time at which new linguistic inputs are received. The updating of the SemRep is triggered when a construction instance assemblage has reached the confidence level.

(Refer to Appendix B, sec. E.2 for a more detailed view of the simulation run)

### 6.6.3 Simulation 2: World Knowledge Route Only (+ Lexical Constructions)

SALVIA was run to generate a computational simulation in the case of the world knowledge route only situation. It provides a simulated counterpart to the conceptual example presented in 5, sec. 5.5.1. The input is the same as the one used for Simulation 1 above: “*the boy is eat -ed by the cake*”.

Using the world knowledge cues only, SALVIA does not recover the proper counterfactual semantic representation. Rather it assigns the thematic roles that are most likely given the system’s knowledge of the world, in particular that usually animate agent eat food items (and not the opposite).

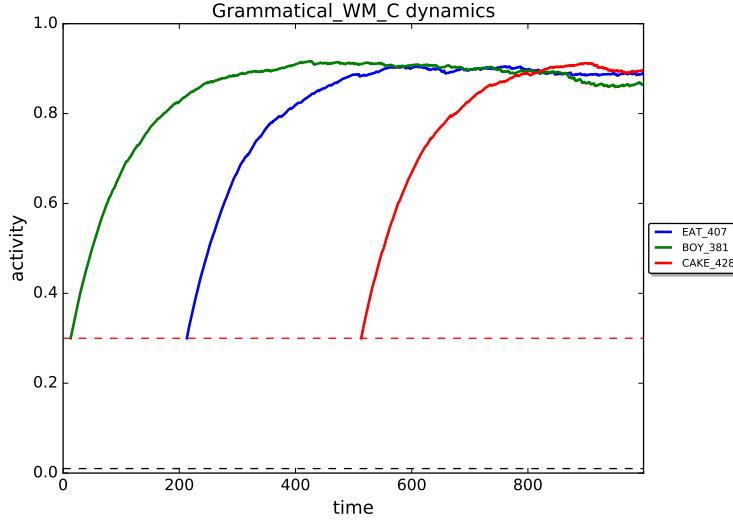


Figure 6.5: Activation levels of the construction instances active in Grammatical WM as a function of time. Top dashed line corresponds to the confidence threshold. Bottom dashed line marks to the pruning threshold. See Appendix B, sec. E.3 for representations of the Grammatical WM states over time. See main text for details.

Fig. 6.5 shows the activation levels of the construction instances invoked in Grammatical WM as a function of time for the system processing the input “(the) boy is eat -ed by (the) cake). Compared to fig. 6.3, here only the lexical construction instances are retrieved. The processes of the Grammatical route is limited to lexical construction retrieval.

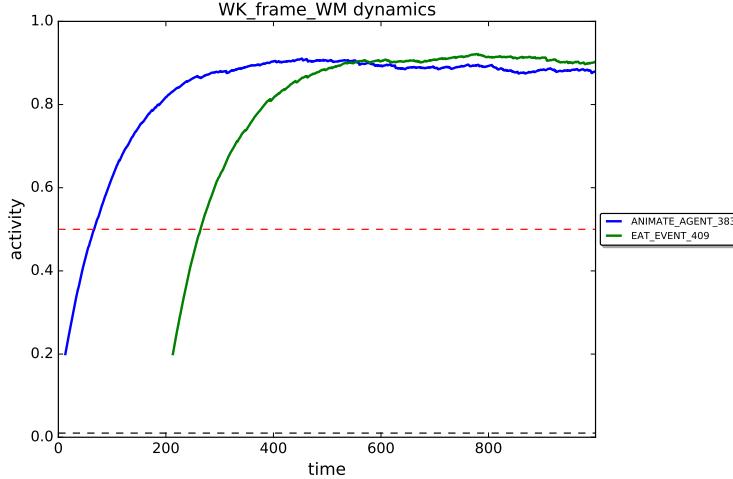


Figure 6.6: Activation levels of the world knowledge (event) frame instances active in World Knowledge WM as a function of time. Top dashed line corresponds to the confidence threshold. Bottom dashed line marks to the pruning threshold. (See main text for details)

Fig. 6.6 shows the activation levels of the world knowledge (event)frame instances invoked in World Knowledge WM as a function of time for the system processing the input “(the) boy is eat -ed by (the) cake). Upon processing “boy”, a world knowledge frame schema predicting that an animate referent is likely to be the agent of an action (here we assume for simplicity that this will be a transitive action), is instantiated in World Knowledge WM. The ‘eat’ input triggers the invocation of the EAT frame schema instance in World Knowledge WM that stipulates that an “eating event” usually involves an animate agent

and a FOOD patient.

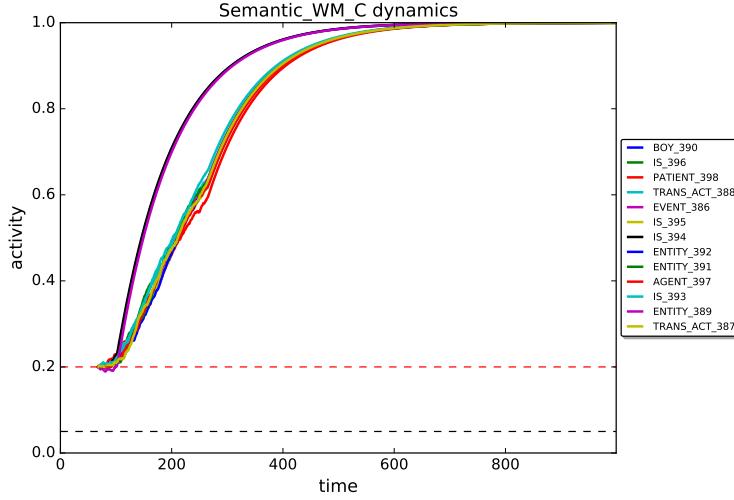


Figure 6.7: Activation levels of the concept instances active in Semantic WM as a function of time. Dashed line marks to the pruning threshold. See Appendix B, sec. E.3 for representations of the Semantic WM states over time. Here, compared to the Grammatical route only case, the whole SemRep graph receives its general shape as soon as “boy” is received since right away the world knowledge predicts that the animate BOY will be the AGENT of a transitive action. As more inputs are received, the SemRep is simply updated. The final interpretation is incorrect as it places BOY in the AGENT position and CAKE in the PATIENT position, as expected by world knowledge, but contrary to what is suggested by the grammatical cues.

(Refer to Appendix B, sec. E.3 for a more detailed view of the simulation run)

#### 6.6.4 Simulation 3: Cooperation Between Routes

SALVIA was run to generate a computational simulation in the case of the cooperation between grammatical route and world knowledge route. It provides a simulated counterpart to the conceptual example presented in 5, sec. 5.6.1, fig. 5.10. The input sentence describe a pragmatically expected event “*the boy eats the cake*”

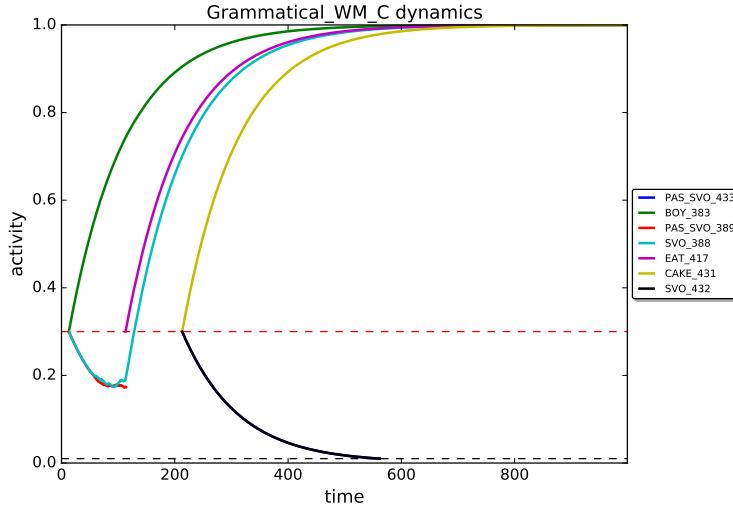


Figure 6.8: Activation levels of the construction instances active in Grammatical WM as a function of time. Top dashed line corresponds to the confidence threshold. Bottom dashed line marks to the pruning threshold. See Appendix B, sec. E.4 for representations of the Grammatical WM states over time. See main text for details.

Fig. 6.8 shows the activation levels of the construction instances invoked in Grammatical WM as a function of time for the system processing the input “(the) boy eats (the) cake”. The processes taking place here are similar to those described in the Grammatical route only case shown in fig. 6.3, the main difference being that the active voice is now winning over the passive voice.

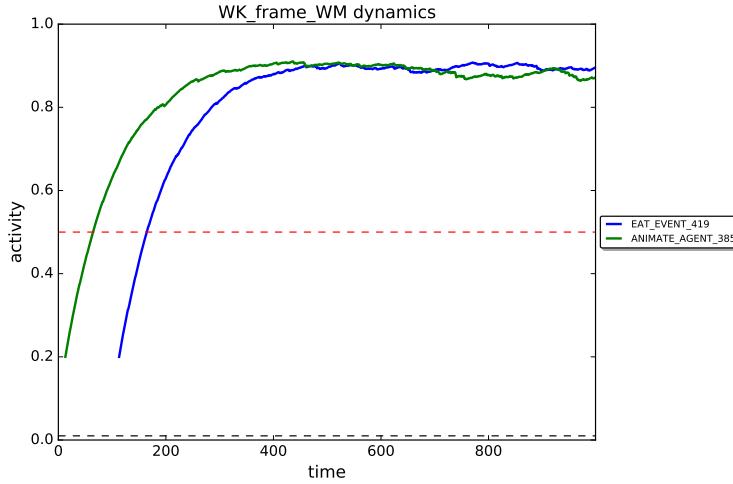


Figure 6.9: Activation levels of the world knowledge (event) frame instances active in World Knowledge WM as a function of time. Top dashed line corresponds to the confidence threshold. Bottom dashed line marks to the pruning threshold. (See main text for details)

Fig. 6.9 shows the activation levels of the world knowledge (event)frame instances invoked in World Knowledge WM as a function of time for the system processing the input “(the) boy eats (the) cake”. The processes taking place here are similar to those described in the World Knowledge route only case shown in fig. 6.6. Indeed, although the input is different “(the) boy is eat-ed by (the) cake” previously vs. “(the) boy eats (the) cake” here, the inputs processed by this route are the same and appear in the same order [boy, eat, cake] (albeit at different time points).

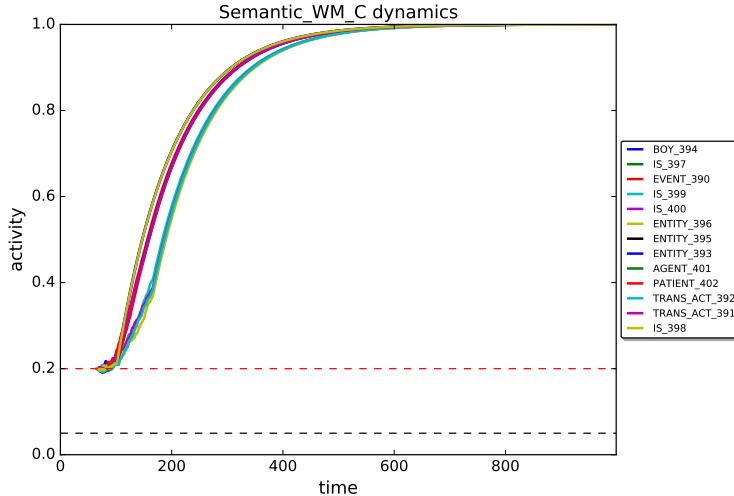


Figure 6.10: Activation levels of the concept instances active in Semantic WM as a function of time. Dashed line marks to the pruning threshold. See Appendix B, sec. E.4 for representations of the Semantic WM states over time. Here the temporal profile of the state of the Semantic WM combines the characteristics of the ones described for the Grammatical route only and World Knowledge route only case (fig. 6.4 and fig. 6.7 respectively.) Since the two routes cooperate, i.e. stipulate the same thematic role assignment (the active voice here places the BOY as AGENT and CAKE as PATIENT which is also the hypothesis put forward by the World Knowledge route), the main difference with the previous cases is that here a consensus on the final semantic interpretation is reached more quickly. The concept schema instances benefit from the combined activation received by both the Grammatical and the World Knowledge WMs. The routes reinforce each other.

(Refer to Appendix B, sec. E.4 for a more detailed view of the simulation run)

### 6.6.5 Simulation 4: Competition Between Routes

Finally, SALVIA was run to generate a computational simulation in the case of the cooperation between grammatical route and world knowledge route. It provides a simulated counterpart to the conceptual example presented in 5, sec. 5.6.1, fig. 5.11. Here the input is a pragmatically counterfactual counterpart of the input used in the previous simulation “*the boy is eat -ed by the cake*”

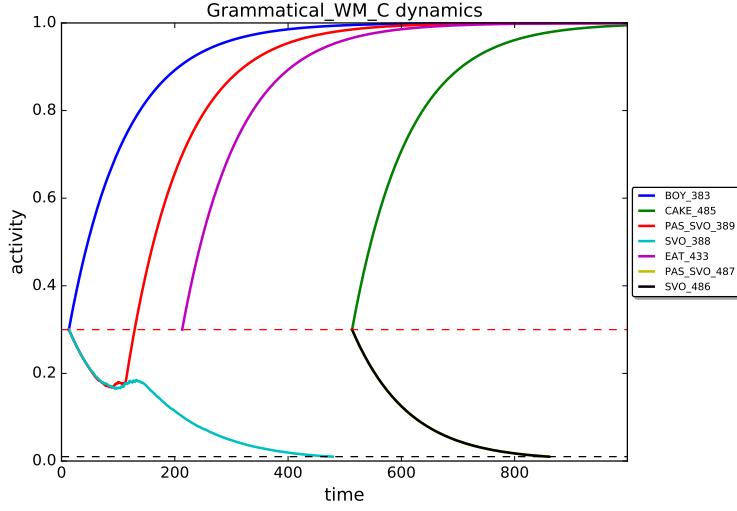


Figure 6.11: Activation levels of the construction instances active in Grammatical WM as a function of time. Top dashed line corresponds to the confidence threshold. Bottom dashed line marks to the pruning threshold. See Appendix B, sec. E.5 for representations of the Grammatical WM states over time. See main text for details.

Fig. 6.11 shows the activation levels of the construction instances invoked in Grammatical WM as a function of time for the system processing the input “(the) boy is eat -ed by (the) cake”. Given the independence of computation between the two routees, the processes taking place here are the same as those described in the Grammatical route only case shown in fig. 6.3. The Grammatical route attempt to generate an interpretation of the input in which the BOY is the PATIENT of the EAT action of which CAKE is the AGENT.

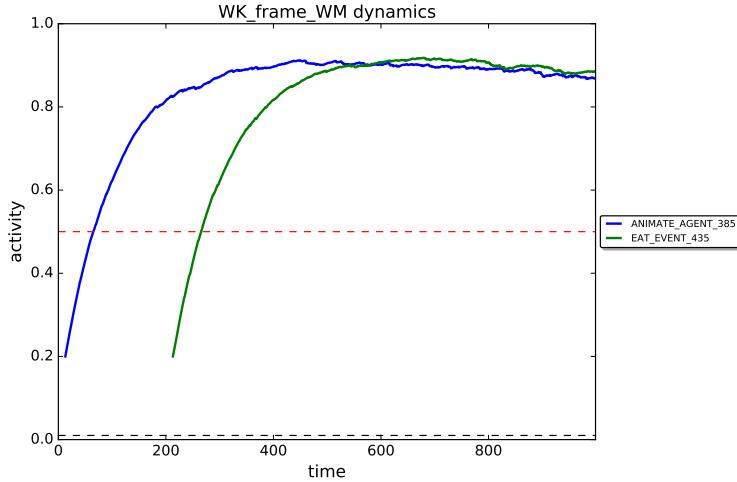


Figure 6.12: Activation levels of the world knowledge (event) frame instances active in World Knowledge WM as a function of time. Top dashed line corresponds to the confidence threshold. Bottom dashed line marks to the pruning threshold. (See main text for details)

Fig. 6.9 shows the activation levels of the world knowledge (event) frame instances invoked in World Knowledge WM as a function of time for the system processing the input “(the) boy is eat -ed by (the) cake”. For the same reason as the one stated above, the processes taking place here are the same as those described in the World Knowledge route only case shown in fig. 6.6.

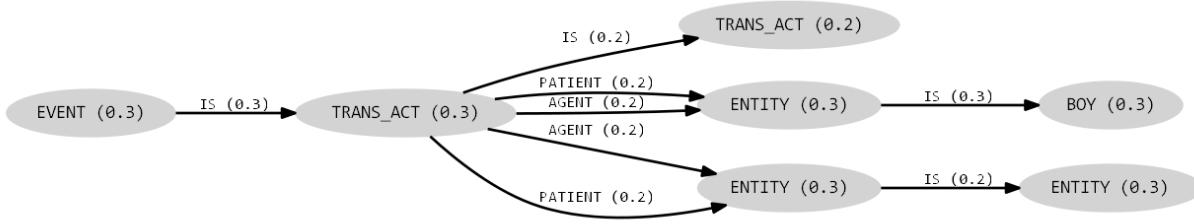


Figure 6.13: State of the semantic WM at time  $t=112$ . The SemRep has been updated by both the World Knowledge and the Grammatical route. The two routes are in **competition** since they each propose opposite thematic role assignment! This is reflected in the SemRep that is now a multigraph. Where the World Knowledge route had generated a AGENT conceptual relation (edge) for the BOY entity, the Grammatical route adds a parallel PATIENT edge (and conversely the Grammatical route adds an AGENT relation where the World Knowledge had placed a PATIENT one). Those multi-edges in the SemRep reflects **competing concept relation schemas** (competition links are not shown). it is worth noting that there is no guarantee that if the AGENT wins one of the competition, then the PATIENT necessarily needs to win the other one. This reflects the fact that the interpretation need not be coherent (“good enough comprehension” paradigm.) (See Appendix B, sec. E.5 for representations of the Semantic WM states over time.)

(Refer to Appendix B, sec. E.5 for a more detailed view of the simulation run)

## 6.7 SALVIA: Simulating the Agrammatic Aphasics Comprehension Performances

Following those initial simulation examples, SALVIA was then used to test the theory proposed in ch. 5 regarding the computational underpinning of the tripartite repartition of agrammatic performances on passive and active voice. the key empirical result is re-summarized here.

In a meta-analysis of studies that reported agrammatic aphasics' comprehension performances and that included contrasts between active and passive constructions, Berndt et al. (1996) found that the data sets could be clustered into three groups of approximately equal size, each reflecting a distinct comprehension pattern:

1. Only active constructions are comprehended better than chance
2. Both active and passive constructions are comprehended better than chance;
3. Both structures are comprehended no better than chance.

So far none of the theories linking agrammatism to a specific deficit in syntax processing has been able to account for this variety in performances.

Informally, the previous chapter suggested that this pattern of performance could be the result of

### 6.7.1 Agrammatic Comprehension: A Novel Interpretation

#### Animate agent hypothesis and agent-first heuristics

One of the world knowledge abstract event frame that is used in the model carries the general world knowledge hypothesis that agent tend to be animate. Animacy as an “agent-like” characteristic has been shown to be used a heuristic across many different language to perform thematic role assignment.

In particular Bornkessel and Schlesewsky (2006) have shown, in their extended Argument Dependency Model that animacy was directly linked to the heuristic used to determine the prominence of referents. The prominence scale is in their model then tied to TRA with one route attempting to assign the agent role to the most prominent referent.

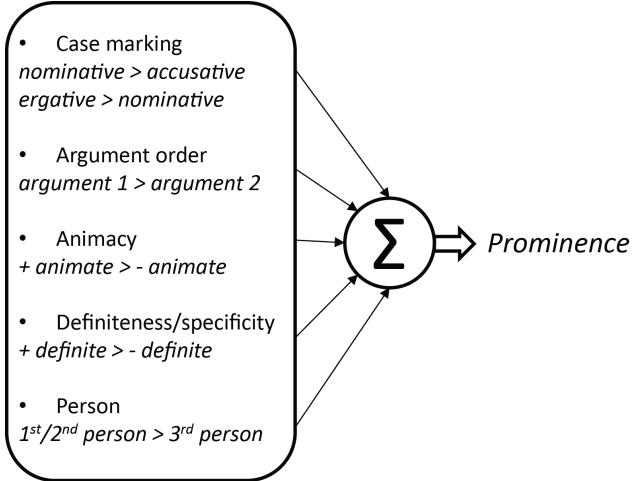


Figure 6.14: Illustration of the factors entering in the process of determining prominence according to the eADM model. SALVIA does not subscribe to the whole analysis carried out by eADM, but hypothesized that, in particular, the impact of animacy on the attribution of agent role is one of the processes at play in the world knowledge route.

### Simulation Results

The simulation results displayed in fig. 6.15 clearly show three regimes at which the model can function that define three performance patterns (see top)

P1 corresponds to a situation in which the grammatical route largely outweighs the world knowledge route. Unsurprisingly, in this case the system determines always the correct TRA since the grammatical cues dominate the semantic interpretation.

P2 is the most interesting situation and correspond to the classic agrammatic profile: The system is able to correctly assign the thematic role for the active sentence but fails for the passive one. This is due to the non-linearity of the C2 interactions between grammatical and world knowledge: the active voice construction enters in **cooperative interactions** with the world knowledge TRA hypothesis while the passive voice construction enters in **competitive interactions** with the world knowledge hypothesis. This difference between constructive and destructive dynamic interactions allows, generate this regime in which the interpretation of active voice sentences is salvaged by the fact that the active voice construction can still recover its capacity to pilot the interpretation due the cooperative boost given by the world knowledge route, while on the other hand, although here the passive voice is considered equally likely or equally preferred, the competition with the world knowledge route is sufficient to prevent it to survive as a grammatical hypothesis governing the TRA interpretation. It is key here that both active voice and passive voice constructions are considered grammatically equivalent in terms of preference and likelihood. It is the nature of the C2 interactions between routes that breaks the symmetry between the two. This strongly differs from the theories that hypothesize a clear difference of nature between the two constructions and use this difference to explain the patterns of agrammatic comprehension performances. (The situation arising in P2 can be tied to the informal example presented in fig. 5.11)

Finally P3 corresponds to the situation in which the world knowledge route largely outweighs the grammatical route. The fact that the system falls at 0% accuracy instead of being at chance is a result of a particular collapse of the animate agent world knowledge heuristics. In the case of the active voice, due to the fact that the sentence is rather short, the heuristic will assign the last animate referent as agent. While, in the case of the passive, due to the longer time it takes for the sentence to be received, the hypothesis of the animate agent assigned to the first referent will have time to significantly build up its activity before the world knowledge hypothesis that the second animate referent is the agent as a chance to be invoked. The latter will therefore always lose the competition. In both cases, this leads to the incorrect TRA. This pattern is interesting but is not characteristics of all the dynamical regime of the system (many will yield a 50% assignment for this pattern).

The patterns P4 and P5 are worth noting as they can vary in width depending on the parametrization of the model. P4 corresponds to regime in which the active voice is consistently correctly interpreted while the proportion of correct TRA for the passive voice decrease with the weight of G. Taken as a whole, it corresponds to a pattern in which the passive voice is overall interpreted at chance. P5 displays as similar patterns but for the active voice.

P1 and P4 taken together can be thought to represent the “high-functioning agrammatics” while P3 and P4 taken together would represent the “low-functioning” agrammatics.

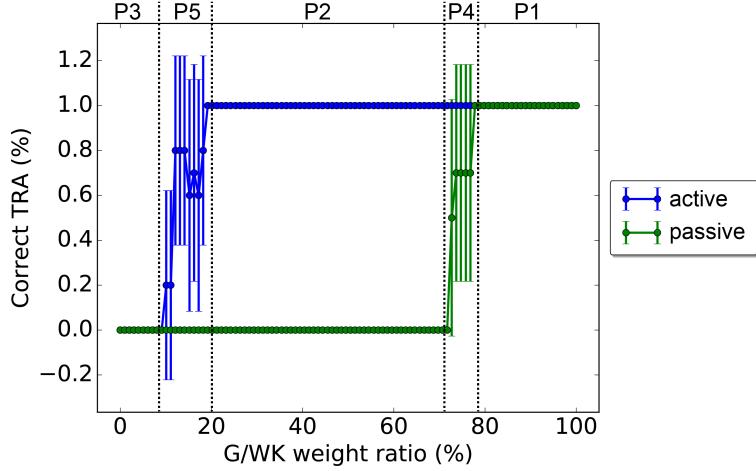


Figure 6.15: Proportion of correct thematic role assignment (TRA) as a function of the relative weights of the Grammatical and the World Knowledge route for the active sentence “**the lion chases the tiger**” vs the passive sentence “**the tiger is chased by the lion**”. (details in text)

## Conclusion

TCG-SALVIA shows how a tripartite distribution of agrammatic aphasic comprehension scores (Berndt et al., 1996) could derive, not from the loss of a specific process, but from variations in the impacts of routes’ relative weights on the cooperative computation dynamics **between** processing routes.

Dynamic cooperation and competition within and between routes in a system-of-systems generates complex behaviors that can remain unsuspected in ‘time-less’ or ‘unstructured’ models.

## 6.8 Discussion

### 6.8.1 Towards a Full SALVIA Model: Linking Vision, Production, and Comprehension

This chapter presented the extension of the SALVIA model into a computational model of language comprehension. The link to visual attention was not directly addressed and neither was the relation to the SALVIA model of language production. Fig. 6.16 gathers together the various sub-systems that have been presented in order to present a picture of what an integrated SALVIA model including Vision, Production, and Comprehension could be. In this presentation, the choice was made to be maximally cautious regarding the processes shared between systems. Only the Semantic WM serves as a interfaces between the three of them.

In particular, it leaves open for future investigation the question of the relations that between the grammatical processes of each modality.

As was mentioned in the informal analysis of the SALVIA comprehension model (see ch. 5), both psycholinguistics (through the use of the visual world paradigm) and neuropsychology (through the use of sentence picture matching tasks) have a common ground in their use of visually situated language compre-

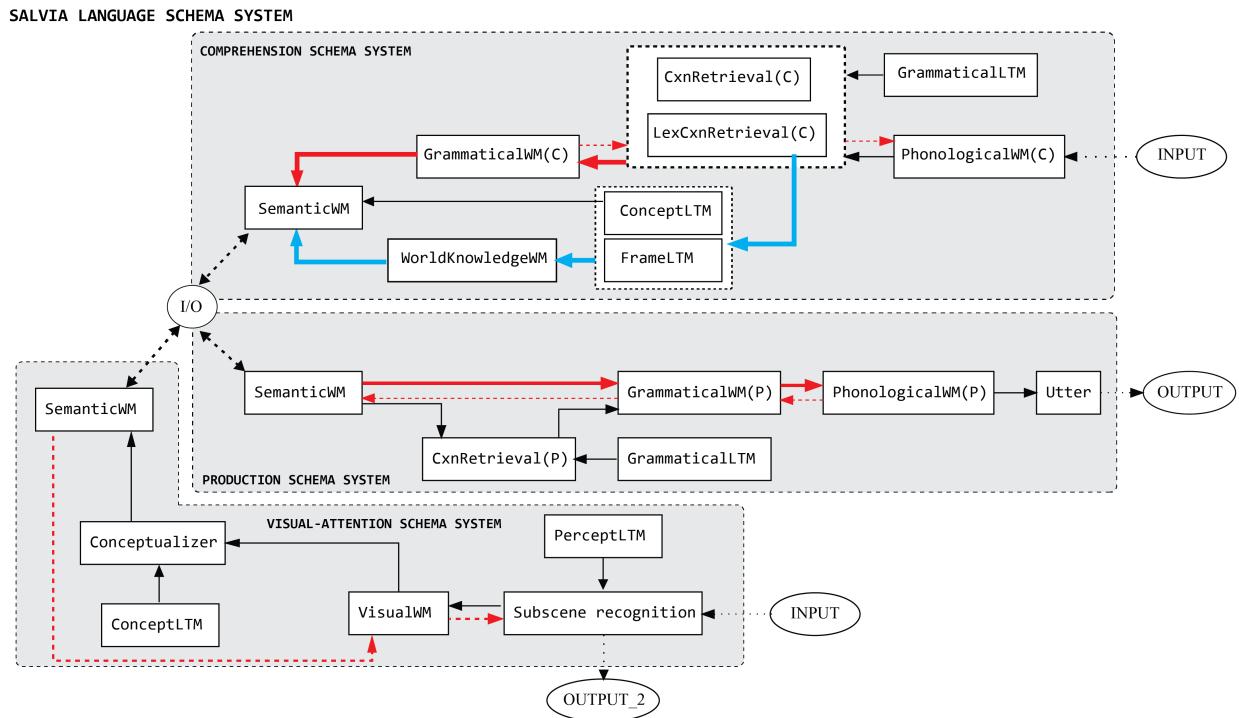


Figure 6.16: SALVIA Production and Comprehension models unified. The full model encompasses a Production Schema System that corresponds to the model that was described in ch. 4, a Comprehension Schema System that was described in this chapter, as well as a Visual-Attention Schema System that was described in ch. 2 in the context of production. The Semantic WM is represented in each schema system for convenience of representation as well as to insist on the question of how the semantic representation are shared between systems, and in particular between comprehension and production systems.

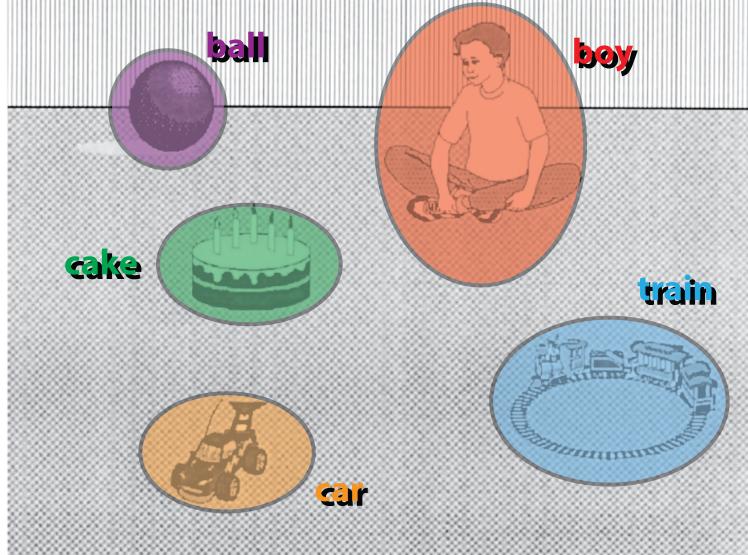


Figure 6.17: Altmann & Kamide 99. Compare anticipatory saccades at the verb in the case of ‘The boy will eat the cake’ vs ‘the boy will move the cake’. Minimal scene (only 1 possible agent and 4 possible patients). In this context, it is assumed that the identity of the set of referents present in the scene is apprehended very quickly. The perceptual schema instances associated and therefore their availability as source of semantic information for the SemRep is assumed to precede the beginning of the sentence comprehension process.

hension to respectively understand the mechanisms underlying contextual language use and testing aphasic patients to determine their symptoms.

The present chapter has not discussed the simulation of the linkages between vision and language in the case of comprehension. The computational support of this task is not yet completed. However, in the case of the “minimal scenes” of the types used in psycholinguistics and in neuropsychology, the model can already partially handle the sentence picture matching task.

Fig 6.17 shows an example of visual scene used by Altmann and Kamide (1999) in psycholinguistic experiments. The key idea here is that such scenes are very minimal in their content (only 1 possible (animate) agent and 4 possible (object) patients), and are usually presented prior to the beginning of the comprehension task (in psycholinguistic studies, but this is of course also true of similar scene used in neuropsychology experiments that are always carried out offline). It is therefore reasonable to assume that, prior to the beginning of the comprehension process, the Visual WM and possibly the Semantic WM are already seeded with the perceptual and conceptual information regarding those referents: the elements labeled in fig. 6.17 can be assumed to correspond to concept schema instances already instantiated in Semantic WM prior to the beginning of the comprehension process.

The comprehension process can therefore be simulated in the same way it was simulated for the two-route model, but as a three-route model, with an initial Semantic WM state. Fig. 6.18 and 6.19 present a conceptual example of this process. Dark grey nodes stand for the concept schema instances that are assumed to be already readily available (either already in Semantic WM or at least already in Visual WM). The model processes the sentence “the boy will eat the cake”.

As soon as the input “the boy” is received, it is already matched with the BOY referent. If we assume that such matching result in saccades towards the referent, the first saccade will be directed toward the boy. Based on both grammatical and world knowledge processes, the semantic representation already grows to predict some transitive action involving another entity, with the role of BOY as patient or agent still unclear but with agent favored due to the animate agent preference world knowledge event frame (event frames are not shown).

When “will eat” is received, the role of BOY as AGENT is confirmed, EAT is added to the semantic representation, and, crucially, the world knowledge already predicts that the patient should be a FOOD item.

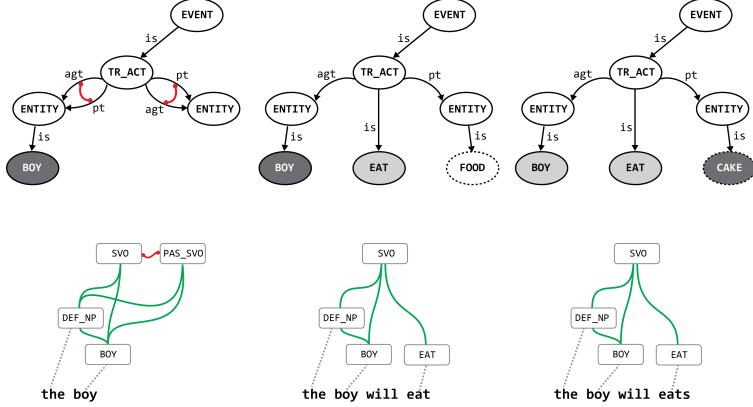


Figure 6.18: Dynamic coordination of three incremental sources of information (1) (see main text for details)

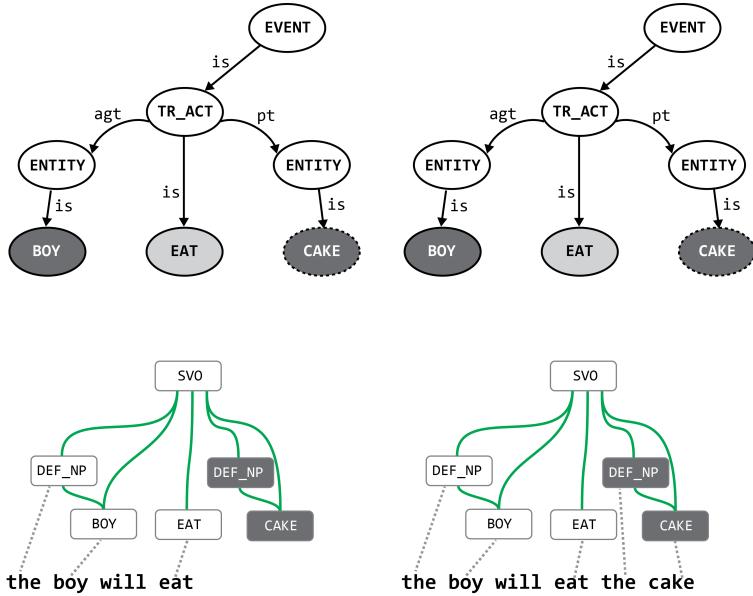


Figure 6.19: Dynamic coordination of three incremental sources of information (2) (see main text for details)

This results in the SemRep to right away match FOOD with the referent CAKE, resulting in a saccade towards the CAKE predicting the upcoming linguistic input.

On the basis of this SemRep updated by both the world knowledge and the visual input, new construction instances are already invoked top-down (CAKE and DET\_NP constructions instances) those predict the grammatical structure and the linguistic form of the input to come.

When finally “the cake” is received, it only confirms the state of the Grammatical and Semantic WM.

## 6.8.2 Model Comparisons

### SALVIA and Theories of Agrammatic Comprehension

SALVIA is built to account for neuropsychological data on agrammatic comprehension and simulate the deterioration of linguistic performances following brain lesions. However, its design was also constrained both by schema-theory and the desire to extend our production model. The use of schema theory ensures that the computational principles are compatible with those generally outlined by brain theory i.e. are not specific to the language system but common to all perceptuo-motor coordination systems. The extension of our production model (instead of the design of a brand new model of comprehension) enables the development

of a model that integrates the joint requirements of action and perception, and therefore can grasp behaviors as part of action-perception cycles (with the physical environment, and in the case of language with other members of the language community).

In this chapter we simulated one of the two possible scenarios that was put forward in ch. 5 depicting how computational impairments of the language system could lead to agrammatic comprehension performances with world knowledge compensating for grammatical processing deficits. SALVIA showed how degradations of the links between the grammatical working memory and the semantic working memory undermine the strength the former as it enters into cooperative computation with heavy semantics information to update the SemRep. This leads to a bias in favor of world-knowledge constraints in the assignment of edges relations between nodes in the SemRep. A second option that was not implemented here considered the possibility that limitations in grammatical working memory capacity could have similar impacts on comprehension, but this time the cause does not lie in a biased competition in favor of world knowledge at the SemRep level but in the difficulty to build a stable construction assemblage. These two accounts, that are not mutually exclusive, are compatible with the SALVIA framework used in production and each relies on a key aspect of schema theory: cooperative computation and the definition of working memory as a schema assemblage.

Such accounts go against the Trace-Deletion Hypothesis (TDH). Grodzinsky (2000) hypothesized that agrammatism results from a deficit in processing the relations between moved phrasal constituents and their traces. Using a construction grammar framework, Template Construction Grammar offers a non-transformational account of grammatical decoding that does not make use of movements or traces. In this framework, active and passive voices are considered as separated constructions each conveying its own form-meaning mapping. Therefore, contrary to the TDH, the SALVIA account of the comprehension of active and passive constructions does not involve any essential differences other than the fact that from a usage perspective active constructions might be more easily activated due to a more frequent use. This accommodates better the comprehension patterns in agrammatics that can either be at chance for both passive and active, only for passive, or better than chance for both (Berndt et al., 1996) (see also Caramazza et al., 2005), with the TDH predicting that only comprehension of passive sentence should be impaired.

These limitations of the Trace-Deletion Hypothesis have been pointed out by other theories of agrammatic aphasia that provided alternative explanations. Building upon both the above mentioned comprehension difficulty for some simple active sentences and the fact that agrammatic aphasics can in some cases be showed to retain syntactic capacities such as discriminating between syntactically correct and incorrect sentences for a wide range of syntactic errors (Linebarger et al., 1983), Schwartz et al. (1987) moved away from a purely syntactic explanation of agrammatism and hypothesized that the deficit lied in the participation of the extracted syntactic information in the thematic role assignment process. In line with the cue competition approach of (Bates and MacWhinney, 1989), they hypothesized that thematic role assignment results from the competition between general world knowledge and syntactic cues, competition that becomes biased in favor of world knowledge in the case of agrammatic aphasics (Saffran et al., 1998). The SemRep, modeling the language-vision interface, is not designed to model thematic roles (as defined by a linguistic theory) but more generally links between perceptually recognized roles in an event scene and a semantic structure. However, our accounts of agrammatic aphasia comprehension in terms of reduced role of grammatical working memory in the cooperative computing of the SemRep echoes Schwartz et al. hypothesis.

The capacity approach to agrammatism developed by Miyake et al. (1994, 1995) is yet another alternative theory. They hypothesized that syntactic comprehension is a result from a reduction, following brain lesions, of working memory resources available to compute the syntactic information contained in linguistic inputs. Our second account of agrammatic comprehension is directly compatible with the general lines of this hypothesis, although in details, the type of resource and computation limitation are strongly dependent on the type of grammatical representations and computational procedures used.

SALVIA provided an account of agrammatic comprehension that is compatible with the theories that have tried to bypass the difficulties faced by the Trace-Deletion Hypothesis. By incorporating within a single model the possibility to test for both working memory and thematic role assignment deficits, SALVIA offers a unified framework to discuss these theories and their implications both in terms of grammatical processing but also in relations to the interaction of grammatical information with world knowledge and visual information. Finally, by highlighting the theoretical and computational distinction grammatical semantics and world knowledge, SALVIA points towards new challenges for the theories of agrammatic aphasia.

## **SALVIA and Other Computational Models**

Cottrell (1985) proposed a hand wired connectionist model of sentence parsing that simulated the joint role of semantic (world knowledge) and syntactic constraints in thematic role assignment. The model uses binding units that dynamically assign a word sense (associated with a content word) to a case (verb based relations) on the basis of both syntactic and semantic features. The model simulates the effect of segregated lesion of connections linking the syntactic analyzer to the binding units, leaving the semantic cues alone in charge of assigning the correct case to the content words. By choosing to impair the connection between the syntactic analyzer to the binding units and not the syntactic analyzer itself, the model provides an implementation of the hypothesis described above that agrammatism stems from deficit in thematic role assignment based on syntactic cues and not from a deficit in processing these cues. SALVIA shares with this model the cooperative computation approach between “semantic” and syntactic cues, but goes further by distinguishing between world knowledge and grammatical semantics, and by offering a framework to analyze the interactions between language and vision in sentence-picture matching task used to test agrammatic aphasics.

The Lichtheim 2 model (Ueno et al., 2011) is a more recent attempt at modeling aphasia using multi-layer neural nets. Compared to SALVIA, the model directly replicates in its architecture the known neural organization of the language system as a network of neural layers, each layer representing a brain region. In addition, the model used plasticity rules to simulate learning processes. Lichtheim 2 uses a dual-pathway architecture of the left hemisphere. A dorsal pathway connects the primary auditory cortex (PAC) to the inferior supramarginal gyrus that then connects to the motor cortex while the ventral pathway starts at the PAC and runs along the superior temporal gyrus, before connecting to the Broca’s area that then connects to the motor cortex. The ventral pathway includes the ventral anterior temporal lobe as an interface between the language system and meaning representations, activated by other modalities. By simulating lesions at various points of its architecture, this neuroanatomically constrained model can simulate aphasics performances on tasks involving word production (naming), recognition, and repetition. However, while Lichtheim 2 is implemented in a neural net and incorporates neuroanatomical data into its architecture, it cannot be scaled up to account for more complex linguistic task involving full sentence comprehension, production, or sentence-picture matching and therefore cannot simulate the patterns of comprehension of agrammatic aphasics that SALVIA tackles at the schema level. It is the role of schema models to allow for neuro-computational simulations of behaviors that are for now beyond the grasp of biologically constrained neural network models, in a way that can then be refined as our knowledge of the system improves.

### **6.8.3 Future challenges**

In SALVIA both production and comprehension involve building construction assemblages in grammatical working memory through cooperative computation. However, so far the model remains agnostic as to whether the construction instances stored in long term memory and invoked during production and comprehension are the same. Linguistic work on idioms has distinguished between encoding and decoding idioms (Makkai, 1972). Indeed a hearer could figure out the meaning of an encoding idiom when she first encounters it although as a speaker she would not have guessed that these expressions are semantically correct (e.g. “answer the door”) while one needs to learn the conventional meaning of a decoding idiom to be able to understand and use it (e.g. “he kicked the bucket” or “he pulled a fast one”). From a usage base perspective, such differences between encoding and decoding can be extended to all constructions with speakers having their own idiosyncratic encoding preferences at the word, idiom, and up to argument structure level, while decoding expectations are shaped by the landscape of input that the speaker receives. If the question of the relation of between the grammatical knowledge stored in long term memory for production and comprehension remains to be better analyzed in SALVIA, the fact that a single brain system supports both the encoding and decoding grammatical working memory finds support in recent behavioral (Kempen et al., 2012) and fMRI adaptation studies (Menenti et al., 2011; Segert et al., 2012).

The empirical results gathered in the last 10 years by the various groups focusing on good-enough comprehension and for which, to our knowledge, no computational framework has been developed are yet another challenge that will need to be addressed by TCG (for a review see Ferreira and Patson, 2007). These studies have used rephrasing empirical paradigm to more precisely study the semantic representations that subject derives from garden path sentences (Christianson et al., 2001; Christianson and Luke, 2011) revealing

until then ignored semantic effects such as the fact that the semantic representations derived from the initial incorrect parse of a garden-path sentence lingers and can be maintained alongside the correct final semantic interpretation. The SALVIA framework could offer a way to directly simulate this rephrasing paradigm by coupling the comprehension and the production system through a SemRep which, since it emerges from cooperative processes, is not endowed with any requirement to optimally represent the semantic content carried by the linguistic input.

As a framework that tackles (so far separately) both production and comprehension, the next step in the development of SALVIA will be to integrate grammatical encoding and decoding from a computational perspective as well as in relation to possible shared neural substrates. In doing so, we will also need to expand our initial focus on agrammatic comprehension to account for the relation between the deficits in receptive and expressive aphasia. The integration of production and comprehension, linked to a deeper analysis of the interface between the language and visual system through the computational exploration the neural processes underlying sentence-picture matching tasks, would make a step towards a computational neurolinguistic model of a brain that can perceive its environment, produce, and understand utterances about what it perceives and therefore interact with others (Steels, 1999). Such a step is crucial if we ever want to be able to build a brain theory of language processing that accounts for the essentially social and interactive aspect of language.

If the SALVIA model of comprehension has not so far been tested against these real-time processing empirical results, it offers a computational framework that coarsely fits generally with these multi-stream approaches in terms of the general differentiation between a world knowledge and grammatical route. Moreover SALVIA adds a quantitative perspective on the challenges that emerge from any attempt to understand brain systems in which computation is distributed while schema theory puts time at the core of the modeling effort. Indeed, we showed how the recency bias, capturing the fact that word serves as context for what occurs in their temporal vicinity, plays a crucial role in modeling comprehension. However the question of the relation between the real-time processing, i.e. time as measured using neuroimaging techniques and especially EEG/MEG, and computational time remains a major challenge. So far this type of timing is out of reach for our model but to our knowledge this is general shortcoming of models that tackle higher level vision or language processes. Bridging the gap between neurocomputational models and real time EEG/MEG recording by allowing models (at the neural or schema level) to make clear causal contact with the measured data would allow computational neurolinguistic models to generate predictions that could be directly tested against neural timing data. As a first step in this direction, we proposed elsewhere to expand synthetic brain imaging methods to the modeling of ERP components (Barrès et al., 2013). Conversely, EEG/MEG data can be used to constrain a model's parameter space and Dynamic Causal Modeling (David et al., 2006) offer a partial answer to this problem. However, linking computational neurolinguistic model to real-time data remains in a great part an open question, even though this issue is one of the main stumbling blocks hindering the establishment of clear linkages between the work of experimentalists and modelers. This issue will be the topic of the next chapter (ch. 7).

# Chapter 7

## Synthetic Event-Related Potentials: A Computational Bridge Between Neurolinguistic Models and Experiments

### 7.1 Linking Computational Models to Brain Data

This chapter changes gears and presents a parallel and yet complementary line of work that focused in offering computational methods to tie neurocomputational models to EEG neuroimaging data. Unlike the other chapter, it is a self contained contribution.

As the next chapter will detail (ch. 8, anchoring a model in brain systems is a very delicate task. Quantitative tools are therefore required to facilitate this task.

Previous work developed Synthetic Brain Imaging to link neural and schema network models of cognition and behavior to PET and fMRI studies of brain function.

We here extend this approach to Synthetic Event-Related Potentials (Synthetic ERP). This work was originally published in (Barrès et al., 2013).

Although the method is of general applicability, we focus on ERP correlates of language processing in the human brain. The method has two components: Phase 1: To generate cortical electro-magnetic source activity from neural or schema network models; and Phase 2: To generate known neurolinguistic ERP data (ERP scalp voltage topographies and waveforms) from the putative cortical source distributions and activities within a realistic anatomical model of the human brain and head.

To illustrate the challenges of Phase 2 of the methodology, spatiotemporal information from Friederici's 2002 model of auditory language comprehension was used to define cortical regions and time courses of activation for implementation within a forward model of ERP data. The cortical regions from the 2002 model were modeled using atlas-based masks overlaid on the MNI high definition single subject cortical mesh. The electromagnetic contribution of each region was modeled using current dipoles whose position and orientation were constrained by the cortical geometry.

In linking neural network computation via EEG forward modeling to empirical results in neurolinguistics, we emphasize the need for neural network models to link their architecture to geometrically sound models of the cortical surface, and the need for conceptual models to refine and adopt brain-atlas based approaches to allow precise brain anchoring of their modules. The detailed analysis of Phase 2 sets the stage for a brief introduction to Phase 1 of the program, including the case for a schema-theoretic approach to language production and perception presented in detail elsewhere.

Unlike Dynamic Causal Modeling (DCM) and Bojak's mean field model, Synthetic ERP builds on models of networks that mediate the relation between the brain's inputs, outputs and internal states in executing a specific task. The neural networks used for Synthetic ERP must include neuroanatomically realistic

placement and orientation of the cortical pyramidal neurons. These constraints pose exciting challenges for future work in neural network modeling that is applicable to systems and cognitive neuroscience.

## 7.2 Background

In the present section, we briefly look at the advantages and drawbacks of fMRI and synthetic brain imaging and then briefly review the use of Event-Related Potential (ERP) data in neurolinguistics. To complete the background, we briefly describe how inverse and forward models relate ERPs to electrical signals within the brain.

### 7.2.1 fMRI and Synthetic Brain Imaging

We have previously explored how simulations of neural networks constrained by data from animal neurophysiology can emulate adaptive and visuomotor behavior (e.g., Fagg & Arbib 1992, Schweighofer et al 1996). With the advent of tomographic brain imaging, the lab expanded this simulation method to develop Synthetic Brain Imaging. Synthetic PET (Arbib et al 1995) facilitated modeling and comparison of saccade generation in primates and humans, and a similar approach was used to associate a synthetic BOLD signal with a model of primate imitation (Arbib et al 2000, see Husain et al 2004, Tagamets & Horwitz 1998 for related studies). The key idea is to start with a biologically grounded neural network for execution of a task set that matches a range of neurophysiological and behavioral data. A spatial and temporal average over the simulated absolute value of all synaptic activations across a region provides a viable prediction of the activation of that region for brain imaging thus enabling the use of simulations of biologically grounded neural networks to yield predictions to be tested against brain imaging studies. Here we initiate a comparable methodology for Synthetic ERP to allow us to provide a bridge between computational models of fine-grained processes in the brain and ERP data. We focus on computational models that simulate the information processing required to perform a given task. As we show in Section 1.3, this emphasis distinguishes the new method from Dynamic Causal Modeling (DCM) and Bojak's mean field model although certain techniques are common to the three methods. We concentrate our work on neural network models but we also discuss schema level computational models. Although the method should have broad applicability the focus of this chapter is on ERP data related to language processing.

To set the stage we briefly compare three sources of data for neurolinguistics: lesions, fMRI (or PET), and ERPs. There has historically been a disconnect between processing models of neurolinguistics which seek to derive linguistic phenomena from empirical data and biological constraints (Kempen & Hoenkamp 1987), and representational approaches which employ a top-down approach seeking to assign theories of language structure to biological systems (Lecours & Lhermitte 1969). Computational modeling was advanced as a means to reconcile these methods (Arbib & Caplan 1979), but because lesion data were the primary data linking brain and language at that time, the attempt to assign linguistic processing to specific cortical regions remained problematic. New techniques for lesion studies combined with subsequent hemodynamic imaging techniques allowed the researchers to support neurolinguistic models at the level of interacting brain regions (e.g., Hagoort 2005b, Hickok & Poeppel 2007), but such models often give an all-or-none match of region to function inconsistent with the dynamics of neural interactions across the brain. Computational approaches like Synthetic Brain Imaging are needed to relate such interactions to hypotheses about detailed neural circuitry or schema interactions. To extend Synthetic Brain Imaging to simulate higher cognitive functions such as language for which there are no animal data, one may explore homologies between the macaque and human brain as a basis for evolutionary hypotheses that suggest possible circuitry within areas of the human brain associated with uniquely human functions (Arbib & Bota 2003)

### 7.2.2 Event-Related Potentials: A privileged window into how the brain processes language

Hemodynamic imaging techniques provide an indirect measurement of neuronal activity mediated by the mechanisms of perfusion and metabolism, resulting in a physiological limit in their temporal resolution on the order of seconds (Heeger & Ress 2002). Unlike PET or fMRI, recordings of neuroelectromagnetic

activity taken by electroencephalography (EEG) and Magnetoencephalography (MEG) are reputed to allow one to follow the time course of neural activity on a time scale of milliseconds rather than seconds but at the high price of a drastic loss of spatial resolution. Neuroelectromagnetic fields are thought to originate within the apical dendrites of cortical pyramidal neuron populations synchronously engaged in excitatory depolarizations and inhibitory hyperpolarizations in postsynaptic potentials (EPSPs and IPSPs respectively) (Niedermeyer & Silva 2010). Unlike the radial symmetry exhibited by their basal dendrites, the linear orientation of the pyramidal cell's apical dendrites provides the basis for generation of a net dipole moment during EPSPs and IPSPs (Ritter et al 1983). The polarized geometry of the pyramidal cell's apical dendrites is repeated at the population level through the organization of pyramidal cells into a palisade formation where the axes of the cells' dendritic trees parallel one another perpendicular to the cortical surface (Nunez and Silberstein (2000), and see also Appendix F (B)). The rationale for viewing EPSPs and IPSPs rather than action potentials as the electrophysiological phenomenon underpinning EEG-recorded scalp potentials is based on the recognition that only synchronous (de)polarizations of neural populations would be able to produce measurable signals at the scalp. While action potentials have a significantly larger amplitude (100mV vs. 10mV), the PSP's time course is longer in duration (10ms vs. 1 ms) and thereby allows more opportunity for synchronization. Such electrical summation plus relative proximity to the scalp satisfies the theoretical conditions for the activity of cortical pyramidal neurons to generate the electromagnetic fields perceived on the scalp surface (Lopes da Silva & Van Rotterdam 1987).

An encephalogram is obtained by recording an array of potentials from leads attached at various places on the scalp (i.e., external to the head) over a certain time period. ERPs are extracted from the encephalogram by averaging recordings of signals at each lead over a number of trials using the same task and with recordings time-locked to stimulus onset. The challenge for interpretation of the ERP is to correctly divide the average waveform into various components that may be viewed as measurements of cognitive processes relevant to the task at hand. When researchers cite the presence of a given ERP component in response to an experimental stimulus (dotted line in Figure 1), they are generally referring to a statistically significant increase in the component's intensity compared to its baseline intensity observed with a control stimulus (solid line in Figure 1).

From 1980 onwards, ERPs have supplied an important method of investigation for neurolinguistic research (see fig. 7.1 for examples in German). The first significant contribution came with the discovery of an increased negativity (shown as an upward deflection in ERP displays!), the N400 event-related potential, which occurs approximately 400ms after the start of presentation of a semantically anomalous word within a sentence (Kutas & Hillyard 1980), suggesting that the N400 may signal "reprocessing" of semantically anomalous information. (Osterhout & Holcomb 1992) saw an increased positive brain potential 600ms (P600) after the onset of words which were syntactically inconsistent with the preceding words of the sentence. Furthermore, final words in sentences typically judged to be unacceptable elicited an N400-like effect, relative to final words in sentences typically judged to be acceptable. In this way, ERPs offer precise timing data, but the distribution of potentials across the scalp does not support a confident inference of where in the brain these changes are elicited (more on this when we discuss Figure 4). A number of neurolinguistic models thus seek to integrate the functional localization from tomographic imaging with the functional time course from ERP recordings and thereby arrive at a spatiotemporal account of various functional components of language processing. However, such an account does not address the issues of what information is represented in diverse brain regions and how they interact with each other in language processing. As we shall later spell out in detail, this gap motivates our initial work on Synthetic ERPs presented here, seeking to relate ERPs to detailed processing models.

### 7.2.3 Synthetic ERP in comparison to Dynamic Causal Modeling and other ERP modeling approaches

Perhaps the ERP modeling approach with most in common with Synthetic ERP is Dynamic Causal Modeling (DCM). The focus of this section, then, is to not only chart those commonalities but also make explicit the ways in which our approach diverges from DCM. (In Table 1 we will offer an explicit comparison of Synthetic ERP not only with DCM but also with Bojak's mean field model). To preview what follows, the difference turns on the word "causal". For DCM, the issue is "What aggregated measures of underlying neural activity could cause the observed ERP recordings?" whereas Synthetic ERP is doubly causal: "(1) What patterns

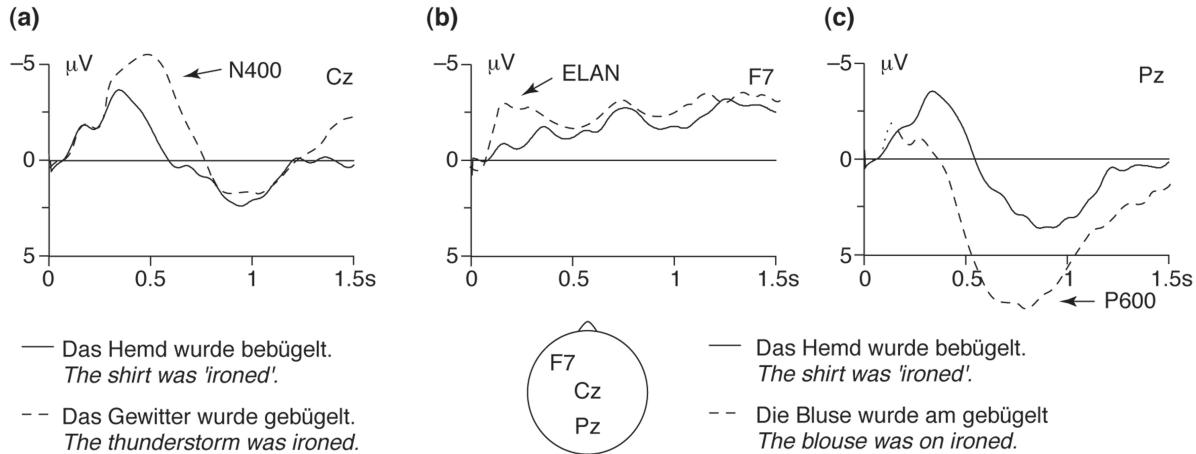


Figure 7.1: Three language-related components in the ERP: (a) the semantic N400, (b) the syntactic early left-anterior negativity (ELAN), and (c) P600. Solid lines represent the condition for a correct word, and dotted lines the condition for an anomalous word (adapted from Friederici 2002)

of interaction in neural circuitry could cause the observed behavior (and, where available, explain single-cell recordings)?; and (2) Could the aggregate activity of neurons in the circuitry so modeled cause the observed ERP recordings?"

Current models of neurolinguistic processing seek to integrate the temporal and spatial information obtained from electromagnetic recording and hemodynamic imaging respectively. To accomplish this integration, researchers have sought to model different ERP component's "neural generators" through the use of equivalent current dipoles whose location, orientation and amplitude are considered to approximate the synaptic activity of pyramidal cells in some region or patch of cortex (see Appendix F (B) for a more thorough presentation of the current dipole model). The attempt to infer generators from the scalp distribution of ERPs, commonly called the inverse problem, has long been recognized as an ill-posed problem (Helmholtz 1853) since many different patterns of neural activity can in theory result in similar EEG recordings (Nunez & Srinivasan 2005). Accordingly, the use of a priori constraints (which may be neurologically unrealistic) is required to generate a unique solution. Numerous approaches to the inverse problem have been developed including parametric approaches (Miltner et al 1994), non-parametric approaches (Pascual-Marqui et al 1994), and constrained solution space approaches using PET or fMRI imaging (Bohland et al 2009, Phillips et al 2002), and much energy has been spent comparing their various merits and the situation-dependent applicability of their respective a priori constraints (Grech et al 2008, Michel et al 2004). However, the forward problem of going from a set of dipoles via a representation of brain/skull/scalp geometries and conductivities to yield scalp potentials is well-posed (though, as we shall see later, representing brain/skull/scalp geometries is a major challenge).

Our goal for Synthetic ERP consists in developing computational tools to link biological neural network models of brain functions to quantitative predictions of the EEG recordings generated by the neural activity in such models. The "biological" neural network approach differs from the "artificial" neural networks approach by its insistence that models should be constrained by biological data including: accurate characterization of discrete cell types, anchoring of sub-networks within specific brain regions, and connectivity structures reflecting the known connectivity of the brain. In addition, such models should be causally complete information processing models. By this we mean that the purpose of the models should be to simulate the behaviors of the organism performing a given motor, sensory, or cognitive task. This distinguishes Synthetic ERP from approaches which simulate neural masses solely in terms of fitting ERP data rather than modeling networks that mediate the relation between the brain's inputs, outputs and internal states in executing a specific task. For example, the behavior of tens of thousands of neighboring neurons (neural masses) might be modeled by a few time varying mesoscopic parameters. So called neural mass or mean field models have been especially important in the field of EEG signal modeling with a focus on replicating

epileptic seizures as well as the richness of known electrocortical rhythms. A comprehensive review of such models can be found in (Deco et al 2008).

Characteristics		Models	DCM	Bojak et al. (2010)	Synthetic ERP
Computational approach	General goal	Biologically sound forward model of ERP useable in a Bayesian inverse model framework.	Biologically sound whole brain forward model of EEG patterns	Biologically sound synthetic read-out of ERP signals from brain anchored network models	
	Modeling focus	Large scale neural dynamics	Large scale neural dynamics	Causally complete information processing model	
	Implementation level	Neural mass	Mean field	Biological neural (or schemas) networks	
Anatomical constraints	Current dipole modeling	Few dipoles	Dipole distributions for the whole cortex	Dipole distributions for relevant brain regions	
	Source waveform	Read-out from pyramidal cell neural mass activity	Read-out from pyramidal cell neural mass activity	Read-out from pyramidal cell synaptic activity	
	Anatomical constraints on the dipole sources	Position only	Position and orientation based on anatomically sound cortical geometry	Position and orientation based on anatomically sound cortical geometry	
	Link to brain atlases	Not discussed	Not discussed	Emphasizes the issue of variation in cortical surface parcellation ontologies	
	Structural connectivity within a brain region	Brain regions are modeled by a single dipole	Realistic connectivity between distributed dipoles	Based on the connectivity of the neural net layer representing the region	
	Structural connectivity between brain regions	Based on existing literature	Based by homology on CoCoMac database for the macaque connectome	Based on existing literature	
	Anatomical constraints on the conduction volumes	Spherical head model	Realistic head model	Realistic head model	
Hypothesis testing	Free parameters estimation	Probability distributions through Bayesian inference	Single values	Single values	
	Model comparison	Bayesian model comparison (accounts for model complexity)	Based on capacity to simulate empirical results	Based on capacity to simulate empirical results	
	Incorporate monkey neurophysiology data	No	Connectivity only	Connectivity and single-unit recording	
Neuroimaging use	Suitable for inverse modeling	Yes	No	No	

Figure 7.2: Comparison of three computational approaches for EEG/ERP forward modeling. From left to right: Dynamic Causal Modeling (DCM) (David et al 2005, David et al 2006); the mean field model described by Bojak et al (2010); and finally our own Synthetic ERP approach. The models are compared according to characteristics related to their computational approach, their use of anatomical constraints, their mode of hypothesis testing including whether monkey neurophysiology data can be used to build hypotheses, and finally their emphasis (or lack thereof) on inverse modeling.

Table 1 presents a comparison of two EEG signal forward models with our Synthetic ERP approach. Dynamic Causal Modeling (DCM) and Bojak's mean field model were chosen because they represent two different modeling approaches which share a common goal with synthetic ERP: to develop biologically realistic forward models. They differ in that DCM focuses on the inverse problem while Bojak et al. focus on the anatomical details of the forward model. We note that Sotero et al (2007) developed an approach very similar to that of Bojak. The models are compared based on 4 types of characteristics: their computational approach, their use of anatomical constraints, the way the hypotheses represented by the models are linked to empirical evidence, and finally their suitability as inverse models for EEG neuroimaging source localization.

In terms of *computational approaches*, both DCM and Bojak's model use a network of mesoscopic neural masses (David & Friston 2003, Jansen & Rit 1995) or mean field models (Liley et al 2002) whose purpose is to account for large scale neural dynamics. The connections between neural masses reflect large-scale white matter connectivity but also, only in case of Bojak et al, local connectivity between neighboring cortical patches. Synthetic ERP on the other hand proposes to link causally complete neural networks models designed with the purpose not to simulate large-scale activation patterns but instead to simulate the information processing required to perform a cognitive task. Both long and short scale neural connectivity constraints can be incorporated in such networks.

The three models vary in their use of *anatomical constraints*. DCM models tend to put less emphasis on cortical topology and head geometry. However this is less an intrinsic limitation of the approach than a consequence of their focus on the Bayesian framework for hypothesis testing (see below). In contrast, Bojak's model puts a heavy emphasis on anatomical constraints for sources and volume conductor modeling. Sources are defined as distribution of current dipoles constrained in position and orientation by the geometry of the cortex. In addition, it uses the CoCoMac database for the macaque connectome (Stephan et al 2001) to generate, by homology, the patterns of connectivity between human brain regions. Finally, Synthetic ERP follows Bojak and insists on the role of cortical geometry in current source modeling. In contrast to Bojak's whole brain modeling approach, Synthetic ERP only models those cortical regions implemented in the underlying neural network model, and in doing so tackles often ignored quantitative issues regarding the role that cortical surface parcellation ontologies, brain atlases, idiosyncratic cortical variations, and neurohomologies (in the case of the comparison of human and non-human primate brain structures) should play in modeling EEG signals.

All three approaches use computational models to express *hypotheses*. Synthetic ERP and Bojak et al. simply focus on generating the simulated EEG data associated with a given model. Such simulations can then be compared to empirical ERP results that can validate or not the hypotheses made by the modeler. Embedded within a Bayesian inference framework, DCM has been specifically developed to allow for an optimal use of empirical evidence, e.g. ERPs, to infer an inverse solution defined as a neural mass network model. Our current aim is to link forward modeling to computational neural networks than can causally explained an observed behavior as opposed to computational models of neural dynamics. If this for now forces us to put aside discussions related to optimal hypothesis testing, it underscores the issue of incorporating within neurolinguistic computational models non-human primate neurophysiology results. The focus of Synthetic ERP on such models will facilitate framing neurolinguistic modeling into an evolutionary perspective that makes full use of both human and non-human primate data. Finally, only DCM has been designed specifically to be used as an inverse model for neuroimaging use. Working outside the inverse modeling framework relaxes many assumptions commonly made to constrain the ill-posed inverse problem such as the limitation in the number of current sources.

Both the Dynamic Causal Modeling approach and the mean field modeling by Bojak et al. provide important tools and insights concerning EEG signal modeling. DCM enables solving the inverse problem in a mathematically sound Bayesian inference framework that allows both parameter fitting and model comparison. David et al (2011) and Yvert et al (2012) applied DCM to ERPs resulting from various linguistic tasks (prosodic and syntactic violations, phoneme detection in pseudo-words, and semantic categorization). They were able to optimally assign the evidence provided by ERP measurements to highlight the role that the subcortical thalamic relay could play in language comprehension as well as the potential structure and change of effective connectivities in various language-related neural pathways. Bojak et al. offer a whole brain account of how to simulate the EEG signal generated by the activity of large brain regions constrained in shape by the 3 dimensional geometry of the cortex. They provide an invaluable means to analyze the structure of brain networks supporting different cognitive tasks which can then be used to

constrain information processing in biological neural networks. However, they leave questions related to simulating behaviors untouched. Our Synthetic ERP approach follows in Bojak et al.'s footsteps by insisting on the importance of cortical geometry. It focuses on the issue of going beyond assigning neural network computational layers to a brain region's name to linking such layers to a 3D model of such region suited for forward EEG modeling. Departing from the use of neural masses and looking into the possibility to link neural network to ERP data, the Synthetic ERP approach insists on the relevance, in an evolutionary perspective, of such models that have been exhaustively used to model processing in the monkey brain. The goal of Synthetic ERP is to build a framework for the causal simulation of both linguistic behaviors and their associated EEG neurolinguistic data (ERP scalp voltage topographies and waveforms). In this chapter our methodology sidesteps the ill-posedness of the inverse model by showing how computational models of linguistic processes may be linked to dipole distributions modeling the neural electric activity resulting from the simulation of such processes, and how these in turn may via forward modeling yield scalp distributions of ERPs that can be tested against neurolinguistic data. Unfortunately, our analysis will reveal that current neurolinguistic models are inadequate for such detailed analysis, thus posing new challenges for both modeling and empirical studies.

## 7.3 “The 2002 Model”: Friederici’s 2002 Model of Auditory Sentence Processing

### 7.3.1 A basic feedforward model of the timing and localization of processes involved in integrating the auditory form of a word into the comprehension of a sentence

Our aim in this chapter is to introduce a new methodology, Synthetic ERP, and show its relevance to providing fine scale models (at the level of neural or schema networks) of brain processes underlying human use of language. To frame the issues involved and the challenges that our approach makes explicit for future research in neurolinguistics, we will focus on a conceptual model (as distinct from a processing model developed for computer simulation) of the temporal activation of brain areas during auditory sentence comprehension developed by Friederici (2002). Subsequent work has refined some aspects of this view (see Friederici 2011 for her own update) and there are now many data from other laboratories that complicate the picture, but the 2002 model remains a valuable benchmark for the approach offered here. The key idea in forming the 2002 model was this: (a) Assess ERP data to time-stamp processes against the beginning of a word with certain characteristics relative to the preceding words of the sentence; and (b) seek other data (whether from PET, fMRI, TMS or lesion studies) that suggests the localization of the posited process. While some of the (a)-(b) matches are well-established, others remain debatable because we stress that (i) ERP data provides precise timing but offer little or no clue as to the source of an observed peak of potential; (ii) the non-ERP data may localize to the level of a general brain region but not to the level of specific circuits or subregions, and in any case cannot specify events with much less temporal precision than 1 second; (iii) different studies use different stimuli, making it unclear whether they elicit the same pattern of neural activity, and (iv) most of the time, multiple brain regions will be actively involved, though to varying degrees.

Figure 2 shows the localization and timing of left hemisphere processes. The model posits four phases:

**Phase 0 (0100ms):** Acoustic analysis and phoneme identification correlate with the well-characterized auditory N100 ERP component.

**Phase 1 (100300ms):** The initial syntactic structure is formed on the basis of information about the word category. An Early Left-Anterior Negativity (ELAN) correlates with rapidly detectable word category errors.

**Phase 2 (300500ms):** Lexical-semantic and morphosyntactic processes support thematic role assignment (e.g., determining which noun phrase denotes the agent of the action described by the verb). A Left-Anterior Negativity (LAN) correlates with morphosyntactic errors. In this model, the N400 occurs in response to words that cannot be integrated semantically into the preceding context.

**Phase 3 (5001000ms):** The different types of information are integrated. A late centro-parietal positivity, the P600, occurs between 6001000ms and correlates with outright syntactic violations (following the

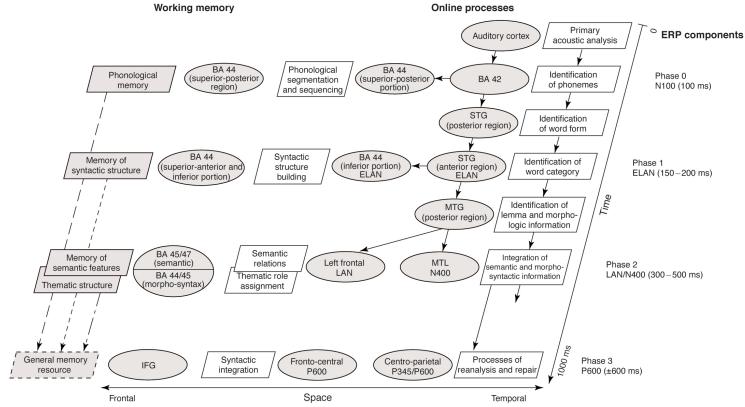


Figure 7.3: A conceptual model of the time course and localization of activation of left hemisphere activity in response to hearing a word during the comprehension of a sentence (adapted from Friederici 2002). The right hand half will be referred to as “the 2002 model.” The boxes represent the functional processes, the ellipses the underlying neural correlates. Abbreviations: BA, Brodmann’s area; ELAN, early left-anterior negativity; IFG, inferior frontal gyrus; MTG, middle temporal gyrus; MTL, middle temporal lobe; PET, positron imaging tomography; STG, superior temporal gyrus.

ELAN) and with garden-path’ sentences that require syntactic revision, and with processing of syntactically complex sentences.

On this view (see Friederici’s Phase 1), linking a new word into a syntactic phrase structure is autonomous and precedes semantic analysis in the early-time windows; these processes interact only at later times. However, word-forms can be ambiguous as to syntactic category e.g., glass can function as adjective or noun. Thus, semantic priming may dominate over syntactic category in analyzing the word form; whereas in other cases multiple interpretations may be needed (at least transiently) to continue the parse, as in deciding how glass is being used in the glass is half full versus the glass pendant is beautiful. All this suggests possible refinements of the 2002 model, but these will not concern us explicitly in the present chapter since our aim here is to establish a new methodology. The model is conceptual rather than computational in that it does not describe the computations within each region, and does not assess what data must be “in play” from earlier words of the sentence (and the broader context) to affect how the current word is processed. Although the left hand sequence highlights memory processes and the processing of the current word at right is posited to update memory structures, the model shows only a forward flow from auditory input without showing how the working memory induced by earlier words of a sentence feeds back to affect the processing of the current word. In particular, there is no way for semantic features (Phase 2) to contribute to the initial updating of syntactic structure (Phase 1) rather, it is only in Phase 3 that problems in syntactic analysis are claimed to trigger processes of reanalysis and repair that may invoke semantic features. Note how different aspects of each word are evoked in different regions during different phases suggesting a distributed representation of the lexicon but one whose components are accessed in serial fashion. As Friederici comments (personal communication) “The 2002 paper model is based on the empirical data then available, and the fact that there are no arrows linking on-line processes to working memory represents the state of the art at that time. There were separate data on working memory activation and data on on-line processes, but not on their interplay.” Given this, a future challenge is to understand how detailed modeling can offer hypotheses that build on more recent empirical data to achieve some measure of causal completeness at the neural or schema network level and then offer ideas for new experiments.

Turning to the cortical regions flagged in Figure 2 (see Friederici, 2002, for the primary references), Friederici notes the classical view that Broca’s area is the locus of syntax but argues that increased fMRI activation of BA 44 is triggered by syntactic memory but not by complexity, whereas local phrase-structure building seems to recruit the inferior tip of BA 44. Such conclusions, however, are based on the structure of the sentences to which subjects are exposed, not on an explicit computational model of parsing. Studies investigating semantic processes at the sentence level report a variety of activation loci, including the left

inferior frontal gyrus (IFG, comprising BA 45/47), the right superior temporal gyrus (STG, which includes BA 22) and the left middle temporal gyrus (MTG, which includes 21 and 37) as well as the left posterior temporal region. However, activation of BA 45/47 appears to depend on the amount of strategic and/or memory processes required.

Friederici (2011) updates this analysis in light of recent data. She shows how the right hemisphere processes prosody as a complement to the syntactico-semantic processes of the left hemisphere, citing data showing that pitch discrimination in speech syllables correlates with increased activation in the right pre-frontal cortex, violations of pitch for lexical elements in a tonal language modulate activity in the left frontal operculum adjacent to Broca's area, and processing of suprasegmental prosody involves the right superior temporal region and fronto-opercular cortex. Other data suggest that right hemisphere prosodic processes can influence left hemisphere syntactic processes. However, such extensions are outside the scope of this chapter, and we will focus on the left hemisphere processes shown in Figure 2. In what follows, the term "the 2002 model" will refer to the model shown on the right side of Figure 2 in which the contributions of working memory are not made explicit, and the flow of information is purely feedforward (the downward arrows of Figure 2).

### 7.3.2 Data on functional anatomy

In recent work, Friederici (2011, 2012) describe a predominantly left-lateralized temporo-frontal network of cortical regions that support various stages of syntactic phrase structuring and semantic integration. Regions are defined using a combination of PET, fMRI, inverse MEG source modeling, and lesion studies. In addition, functional parcellation schemes are devised for some of these regions based upon DTI tractography (Raettig et al 2007) and Granger causality mapping of DTI and fMRI data (Upadhyay et al 2008).

In what follows, we use the term "module" for any group of brain regions or subregions postulated to work together in some subfunction of (language) processing. The model shown in Figure 3 provides the neuroanatomical for the 2002 model by defining five functional language modules composed of cortical regions in the perisylvian language areas and connected via four major white matter fiber tracts (Friederici 2009). We add a number of post-2002 references, and note that some of their data suggest modifications in the 2002 module but reiterate that such refinements are extraneous to our current goal, the grounding of the method of Synthetic ERP.

The first module (Phase 0, N100, around 100 ms) is subserved bilaterally by the primary auditory cortex (PAC) and the planum temporale (PT). These areas are thought to support the analysis of phonemes (Binder et al 2000, Pantev et al 1988). Bridging the Sylvian fissure, the next module (Phase 1, ELAN, peaking between 100-300ms post stimulus onset) is considered responsible for phonemic concatenation (DeWitt & Rauschecker 2012) and incorporates the anterior STG and the frontal operculum connected by the ventral uncinate fasciculus. The posterior STG (pSTG) along with BA 44 comprise a module (Phase 2, LAN occurring in the 300-500ms window) through connections mediated by the dorsal pathways through the arcuate fasciculus and the superior longitudinal fasciculus. This module is thought to be involved in syntactic processing (Kuperberg et al 2000, Newman et al 2009). The most distributed module (Phase 2, N400, around 400ms) is claimed to process semantic information and incorporates middle and posterior portions of the STG, middle and posterior portions of the MTG, BA 47, and BA 45 (Vigneau et al 2006) connected by the ventral extreme capsule fiber systems (Weiller et al 2011). Lastly (Phase 3, P600, around 600ms post stimulus), the integration of syntactic and semantic information is carried out by a module including the pSTG and posterior superior temporal sulcus (pSTS), as well as the basal ganglia; however these regions are only implicated by inverse MEG studies (Service et al 2007). Thalamic contributions to these stages of language comprehension have also been explored (David et al 2011, Wahl et al 2008), but are not considered in the current approach.

The 2002 model and its later iterations explore a larger body of cortical and subcortical language-related regions, but only those directly associated with the generation of linguistic ERP components were selected for Synthetic ERP implementation. However, future "doubly causal modeling" (i.e., using neural and schema networks to explain single cell data and behavior as well as to drive a forward model to generate ERPs) will have to take, e.g., thalamic and basal ganglia activity into account, as indeed we have done in our classic model of control of saccadic eye movements (Dominey & Arbib 1992, Dominey et al 1995).

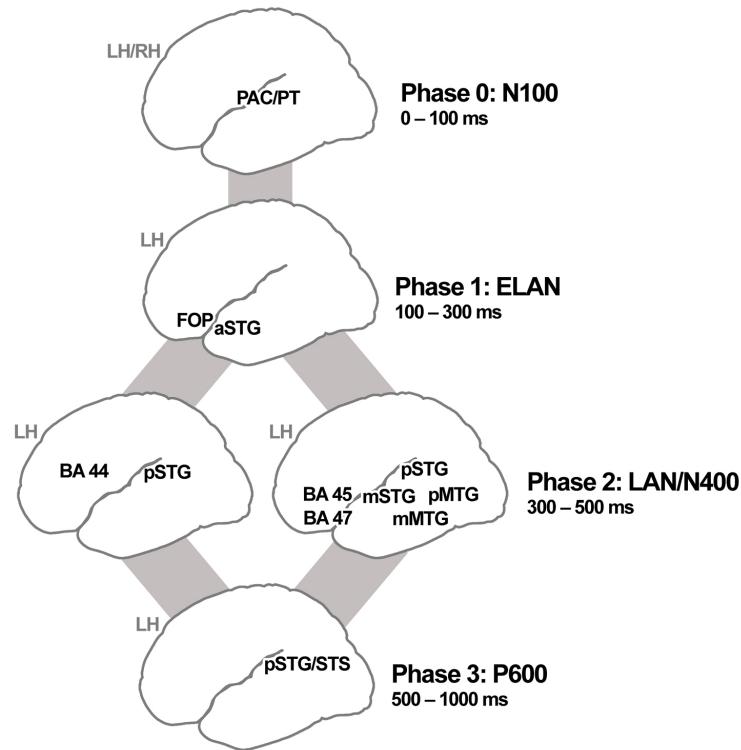


Figure 7.4: Five modules and four processing phases extracted from the 2002 model and implemented within Synthetic ERP. The cortical regions depicted within each of the five modules are left lateralized, except for those generating the N100. Two language modules supporting linguistic processing and ERP generation are activated during Phase 2 (LAN shown at left; those for N400 shown at right). Abbreviations: LH, left hemisphere; RH, right hemisphere; BA, Brodmann's area; PAC, primary auditory cortex; PT, planum temporale; FOP, frontal operculum; aSTG, anterior portion of the superior temporal gyrus; pSTG/STS, posterior portion of the superior temporal gyrus and superior temporal sulcus; mSTG, middle portion of the superior temporal gyrus; pMTG, posterior portion of the middle temporal gyrus; mMTG, middle portion of the middle temporal gyrus.

## 7.4 The Two Phases of Synthetic ERP: A Preliminary Computational Framework

Synthetic ERP is a means to use computational models of neural or schema networks to predict the scalp potentials associated with ERPs. The method has two phases:

**Phase 1:** To generate amplitudes for dipole distributions from processing models for neurolinguistics based on neural networks or schema networks;

**Phase 2:** To apply forward modeling, based on a realistic anatomical model of the human brain and head, to compute Synthetic ERP activity which can be tested against available ERP data.

Later in the chapter, we offer a preliminary perspective on Phase 1, but it is Phase 2 that receives detailed attention, and computer simulation, in the remainder of the chapter. The present section offers a preliminary theoretical framework that makes explicit some key computational issues to be tackled in order to express ERP and localization data in a form which makes it a target suited for the Synthetic ERP methodology.

### 7.4.1 A preliminary framework

Let us use  $[T_1(A), T_2(A)]$ , etc., for the period during which region  $A$  is posited to make its contribution to the current epoch of processing. The nature of such contribution in the general information processing scheme depends on the hypotheses made by a given model. This is roughly the period during which the region's activity contributes to the ERP. In formalizing the data, we seek to assign a dipole amplitude time course to the region  $A$  for that interval,

$$d_A : [T_1(A), T_2(A)] \rightarrow d_A(t)$$

which assigns to each time  $t$  the amplitude  $d_A(t)$  of the dipole oriented in the direction orthogonal to the cortical surface of  $A$ , and related to the synaptic activity of pyramidal cells (More generally, we may need to use multiple dipoles, rather than one. More on this in later sections.) The forward model then computes the electric field as a function of time resulting from such current dipole activity. This method can be applied to all the relevant regions and the electric fields generated by their respective activities can be added to yield the complete ERP for that epoch (on this additive nature of the field, see section 4.3).

The precise definition of the function  $d_A$  will be the focus of Phase 1. However, the current general specifications of  $d_A$  already imply that neural network (and schema network) models must incorporate the following constraints:

- (1) They must be tied to the 3 dimensional geometry of the cortex in order to derive the location and orientation of the dipole,
- (2) They should subdivide their neuron models into different cell-types that allow the read out of synaptic activity from pyramidal cells only.

In many ERP studies of language processing, the epoch starts with the onset of presentation of a word. In a feedforward model such as the 2002 model, it is assumed that the state of the brain at the start of the epoch may be ignored with one exception that knowing whether the word is semantically or syntactically anomalous yields different values for  $d_A$  in some brain regions.

- (a) In cases of divergent feedforward processing (i.e., a region receives input from at most one other region, and there are no loops), the timing in a chain  $A \rightarrow B \rightarrow C$  goes something like this:
  - $t(A)$ : time required after receiving coherent input for  $A$  to reach a degree of confidence for its processing of the current data.
  - $t(A \rightarrow B)$ : time required for a coherent output from  $A$  to affect activity in  $B$
  - $t(B)$ : time required after receiving coherent input for  $B$  to reach a degree of confidence for its current processing.
  - $t(B \rightarrow C)$ : time required for a coherent output from  $B$  to affect activity in  $C$ . Here  $T_1(A) = 0$ ,  $T_2(A) = t(A)$ ;  $T + 1(B) = T_2(A) + t(A \rightarrow B)$ ;  $T_2(B) = T_1(B) + t(B)$ , etc.

- (b) In general, feedforward processing might involve confluent inputs, so that  $C$  receives input from  $A$  and  $B$ . In this case, a descriptive model might allow one to conclude that  $C$  can carry out its computation with the input from either  $A$  or  $B$  alone, or  $C$  might require both inputs. In the first case

$$T_1(C) = \min[T_2(A) + t(A \rightarrow C), T_2(B) + t(B \rightarrow C)]$$

and in the second case

$$T_1(C) = \max[T_2(A) + t(A \rightarrow C), T_2(B) + t(B \rightarrow C)]$$

and other cases may obtain.

- (c) However, even more generally there will be loops in the computation and it may well be that in simple cases a region  $C$  can complete its computation in feedforward mode, whereas in other cases it may need to get further input both bottom up (this is like a switch from the first case to the second in (b)) and top down and this may involve input from regions encoding states achieved in processing earlier words of the input, and interaction between states "at multiple levels" initiated by receipt of the current word.

The 2002 model is a case of divergent feedforward processing. Later models by Friederici et al. have slightly relaxed this serial requirement, adopting a cascade type model which allows for some parallel activity and temporal overlap<sup>1</sup> (Friederici 2012, Friederici & Kotz 2003). It is our conviction going back to the "neuralization" (Arbib & Caplan 1979) of the classic HEARSAY model of speech understanding (Erman et al 1980, Lesser et al 1975), and reinforced by current computational modeling that the general models of case (c) are the rule rather than the exception, and it is this which motivates the need for Phase 1 of Synthetic ERP using detailed processing models to infer dipole activity  $d_A(t)$  for diverse cortical areas  $A$  in cases where complex interactions underlie the response to different task conditions. Such a model would yield individual trial-by-trial ERP values, but would then be averaged appropriately across an ensemble to yield predictions to be tested against the empirical data.

#### 7.4.2 The challenge of timing data for the 2002 model

To illustrate the distinction between the descriptive time course of the 2002 model and the generative time course needed to implement a computational framework for Synthetic ERP, consider the N100 component which signals the acoustic processing of Phase 0. At the conceptual level it is enough to point out that the waveform peaks around 100ms and thereby establishes the time window for phonemic processing ascribed to the bilateral primary auditory cortices by MEG (Pantev et al 1988) and fMRI (Binder et al 2000). However, the N100 component is not monolithic, so that the early phonemic processing stage could be better characterized by a combination of temporally and functionally distinct subcomponents with putative sources in *and around* primary auditory cortex (Naatanen & Picton 1987, Scherg et al 1989). McCallum and Curry (1980) described three such subcomponents; N1a generated bilaterally in the auditory cortex on the dorsal surface of the temporal lobes with a fronto-central peak at around 75ms, N1b a component of indeterminate neural origin peaking at the vertex electrodes around 100ms, and N1c generated in the STG with a laterally distributed peak around 150ms. A similarly complex cascade of cortical activation is implicit in other components such as the N400 (Kutas & Federmeier 2011). While in this cautionary mode, consider the N400 of Figure 1(a). It has been described as a small negativity around 400ms that relates to a semantic anomaly. (There are other linguistic correlates as well, but they need not detain us here.) However, what we really see is a small N400 when there is no anomaly, and an N400 which lasts longer and rises higher when there is a semantic anomaly. Thus a computational model adequate for Phase 1 would not simply turn a dipole on if there is a semantic anomaly, but would rather show in some detail how semantic processes which are employed normally will be placed under stress when a semantic anomaly occurs, and then show how that translates into a time course for dipoles that forward modeling (Phase 2) can then use to explain the variation in N400 timing and magnitude seen in Figure 1(a).

---

<sup>1</sup>Such overlap can be found in the processing of semantic and verb-argument information (N400) and morphosyntactic information and thematic role assignment (LAN) which is portrayed occurring partly in parallel during Phase 2 of the 2002 model (See Phase 2 of Figure 3).

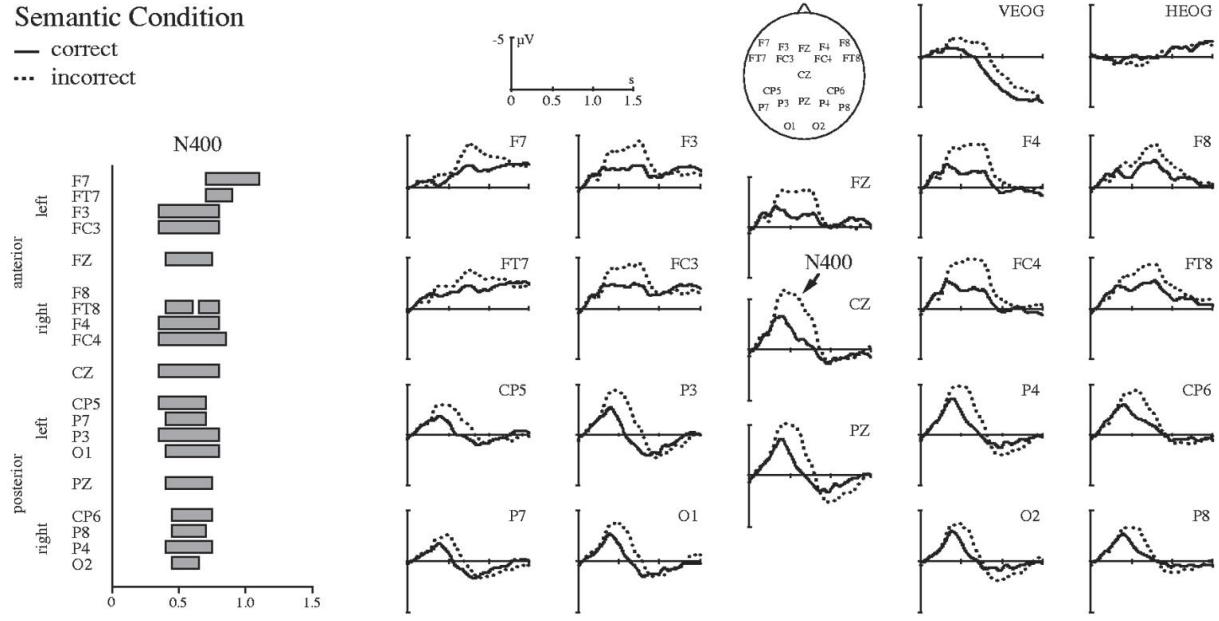


Figure 7.5: Grand average ERPs when participants judged sentences for overall correctness, with averages calculated relative to a 100ms post-stimulus onset baseline. The display shows ERPs for the semantic violation condition (e.g. *Der Vulkan wurde gegessen*’ The volcano was eaten’) as compared to the correct condition (e.g. *Das Brot wurde gegessen*’ The bread was eaten’), where in the examples the origin of the x-axis corresponds to the onset of *gegessen* and negative voltage is plotted upwards. The rectangles at left display the results of MANOVAs comparing the incorrect condition to the correct condition for each electrode, starting at the onset of the participle. Shaded bars indicate significant effects ( $p < 0.05$ ) whenever two or more successive 50ms windows revealed a reliable effect. (Adapted from Hahne & Friederici 2002)

To further highlight the challenges of Synthetic ERP simulation, consider Figure 4 which shows one portion of the ERP results from a study (Hahne & Friederici 2002) which provides an experimental basis for Friederici’s syntax-first serial approach to neurolinguistics and the resulting time course of the 2002 model. The challenge for Synthetic ERP is not simply to produce a single waveform as shown in Figure 1 (the N400 arrow in Figure 4) but also to obtain, to some acceptable degree of approximation, the distribution of activity seen across the various scalp electrodes. Note particularly the rectangles at left part of each row which show for each electrode the time interval during which there is a significant difference between the incorrect and the correct condition. Our point is that, interesting though such an analysis of significance may be, our concern is to relate ERPs to a model of the underlying processing, and for this it is the actual waveforms in the two conditions that must be predicted (to some level of fidelity), not just the time course in which the stated difference is significant.

Timescales of activation must be much more clearly delineated than was required in synthetic PET and fMRI. Whereas the low temporal resolution of fMRI tends to superimpose all stages of activation and communication within an active module of cortical regions, the instantaneous temporal resolution of ERP signals necessitates that signal resolution and propagation effects be taken into account. Signal propagation effects can be attributed to white fiber conductivity delays which can range from 1050ms (Aboitiz et al 1992, Matsumoto et al 2004). We are confronted with a situation where signals are clearly propagating through the different elements of the 2002 model’s temporo-frontal language network well before those regions contribute the ERPs with which their functions are associated. In this case, the conduction delays of feedforward signal propagation are inadequate to explain what is going on. For this reason we earlier introduced equations like  $T_1(B) = T_2(A) + t(A \rightarrow B)$  to emphasize that propagation delays sum with processing time in a region before it achieves a coherent output that affects other regions. Such processing time should incorporate constraints on the temporal characteristics of post-synaptic potentials buildup and their impact on the activity profile of a brain region. (Matsumoto et al 2004) measured conduction delays between Wernicke’s and Broca’s areas

of 2236ms, but conjectured that early end of that response window, mediated by the thickest fibers, may have been lost due to an artifact in the recording data, thereby bringing the average latency closer to the 20ms window previously mentioned. This 20ms conduction delay is also in accordance with evaluations of white matter fiber conductivity (Waxman & Swadlow 1977) and anatomical properties (Bishop & Smith 1964). Thus the present implementation of Synthetic ERP will assume a 20ms propagation delay prior to the onset of each processing phase. Later iterations will explore a more nuanced conduction delay scheme.

## 7.5 Phase 2 in Detail: From Areas of Cortical Activity to ERPs

We now turn to a detailed presentation of Phase 2 of the Synthetic ERP model, showing what is required to specify the cortical regions involved in a task with sufficient accuracy to allow first the computation of dipoles, and ultimate those of observed ERPs. We will use the 2002 model to make these challenges explicit when detailed computation of the forward model is invoked.

### 7.5.1 Forward modeling in the Synthetic ERP framework

Precise forward modeling requires both a head model and a processing model. The head model requires head meshes to define the various compartments of the head and a realistic brain mesh associated with a brain atlas to anatomically localize the brain regions. The head model also provides the conductivities of the various head volumes in which the electric field propagates. The realistic brain mesh enables one to anatomically constrain the direction and orientation of the dipoles associated with each cortical region or patch while the head meshes allow one to specify the various conduction volumes as wells as the locations of the sensors on the scalp that simulate the EEG electrodes.

In what follows, we use dipole source waveform activations partially derived from the ERP empirical results, leaving open the future use of detailed neural or schema network models (as specified in Phase 1 of the full Synthetic ERP approach).

### 7.5.2 Head model

#### Conduction volumes

The implementation of Phase 2 of Synthetic ERP reported here uses a 4-compartment head model based on the MNI Collins MRI scans (Evans et al 1993, Mazziotta et al 1995) which provides meshes representing the surfaces defined by the grey matter, the inner skull, the outer skull, and the outer surface of the scalp respectively. The reader unfamiliar with such models can refer to Appendix F (C) which offers additional details. In addition, Appendix F (A) provides the general physical formulation of the forward problem. We want to note that we use a realistic head model representing the anatomy of a specific individual. The impact of the model's specificity on conduction volumes is often neglected (many forward solutions go as far as simplifying the conduction volumes by concentric spheres). However, this specificity becomes an issue when considering individual-specific variations in cortical geometry (Mangin et al 2010) as well as contrasts between, geometrically, functionally, or cytoarchitectonically informed approaches to defining brain regions (Fischl et al 2008). Idiosyncratic anatomical variations of language-related brain regions only begins to be better quantified (Amunts & Zilles 2012, Keller et al 2009, Keller et al 2007). Such issues have long been recognized as important to neurolinguistics in general (Whitaker & Selnes 1976), and take a central position here as the accuracy of ERP simulation critically depends on accurate representations of cortical folds. For this reason, in adjunction to the head model, we next discuss cortical parcellation ontologies and atlases. We want to insist on the need to quantitatively link EEG sources to standard brain nomenclatures in order to bring issues related to cortical geometry into computational modeling.

#### Brain areas

A brain area is a patch of cortex defined as a set of faces taken from the realistic brain mesh. Such an area can be defined as anything from a single mesh face to a whole brain region. Currently, Synthetic ERP lets one define brain areas by selecting an individual face or a set of contiguous faces using the Destrieux

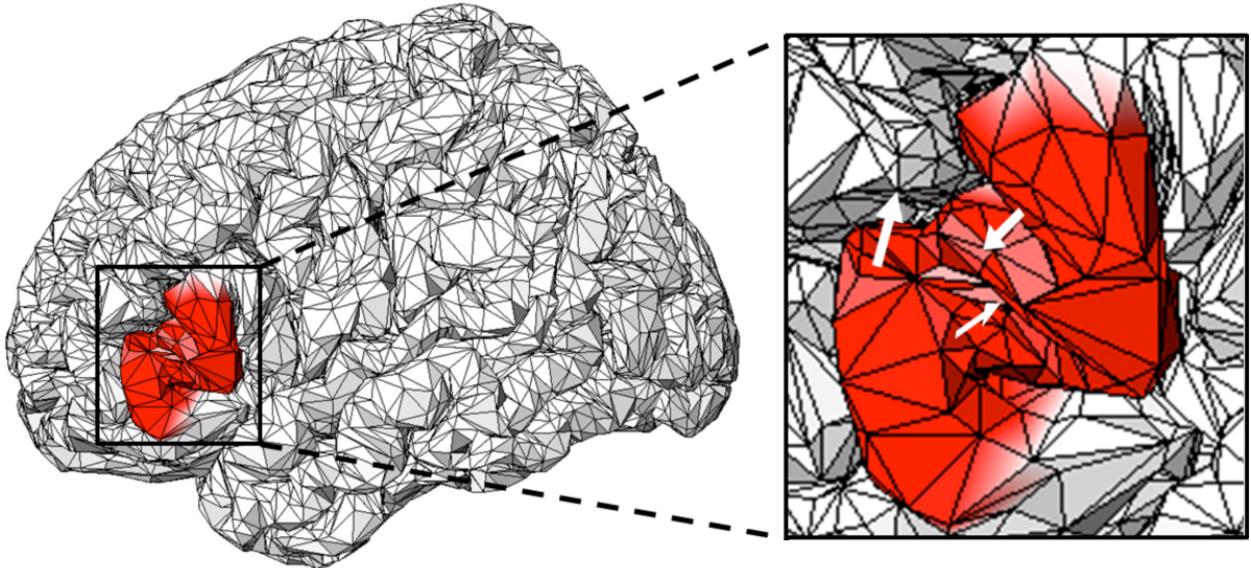


Figure 7.6: (Left) Left pars triangularis (Desikan-Killiany atlas) brain area defined on the realistic brain mesh. Brain areas can be defined as individual faces or as brain regions. The Destrieux or Desikan-Killiany anatomical atlases can be used to select a collection of faces associated with a brain region or other cortical area. The current model uses the Brainstorm default version of these atlases. Each brain region is directly defined as a list of associated vertices on the MNI Collins brain mesh. (Right) Close up view of the brain area. The orientation of source dipoles is defined by the unit vector normal to the face and pointing inward. Here we show the normals to three arbitrary faces of the left pars triangularis (though for ease of representation, the normals are here oriented outward). The orientation of a source dipole within a given brain is highly dependent on its position.

(Destrieux et al 2010, Fischl et al 2004) anatomical atlases based on its BrainStorm versions. Although other atlases exist (we also implemented the Desikan-Killiany atlas (Desikan et al 2006)), we mainly use the Destrieux nomenclature which defines regions based on gyral and sulcal borders (while the Desikan-Killiany cortical atlas defines regions based on gyri alone), giving us more flexibility. Selecting a region triggers the selection of all the vertices and faces of the realistic brain mesh that belong to this region. We incorporated the possibility of selecting only the anterior, middle, or posterior portions of an atlas based brain region. These informal subdivisions heavily used in conceptual models are defined as covering a third of the length of the area along its main axis as defined by PCA.

We apply our cortical brain atlas on the brain model of a specific individual. This approach reflects the limitation that most ERP data reported in the literature omit related data concerning the cortical anatomy of those individuals participating in the study. In addition most ERP components reported are averaged over a given population, begging the question of what is a good definition of the cortex’s “average” geometry. In this chapter we make explicit our choice of using a standard individual brain in order to insist that any discussion on ERP forward modeling must address issues related to cortical surface templates, variation in cortical geometry, surface averaging, and parcellation ontologies. For a review of the state of the art and challenges related to these questions, we refer the reader to (Evans et al 2012).

## Dipoles

Current equivalent dipoles are used to model the electromagnetic field sources associated with a given brain area. In the simulations reported below we consider mainly excitatory synaptic inputs (positive amplitudes for dipoles oriented inward). Appendix F (B) offers more details concerning the modeling of neural activity using current equivalent dipoles.

A key assumption of Synthetic ERP is that dipoles are constrained both in position and in orientation

by the cortical geometry of the active source, thereby ensuring that all dipoles have a physical meaning (see Table 1). Appendix F (B) details some of the hypotheses intrinsic to the dipole model regarding the curvature and size of the patch of cortex modeled. Synthetic ERP offers the option of summarizing a brain area by either a single dipole or as a distribution of dipoles associated with the faces of the mesh linked to the brain area. When a single dipole is used, its orientation is defined as the mean of the normal vectors associated with each of the faces contained in the brain area. The magnitude of the dipole at time  $t$  is given both by the surface  $SA$  of the associated brain area and its time course function  $d_A(t)$ .  $SA$  simply plays a role of multiplicative factor on  $d_A(t)$  to ensure that the magnitude of a dipole is proportional to surface of brain tissue it represents. It is given by the sum of magnitudes of the normal to the faces contained in the brain area. The function  $d_A(t)$  for a given dipole will link, in future work, to the activity levels in computational modules associated with the brain area this dipole covers (Synthetic ERP phase 1). In the present work, a more limited processing model is used to generate the time course activation function for each dipole.

### 7.5.3 Forward model and lead field computation

Synthetic ERP uses a standard method to compute the forward model using the FieldTrip (Oostenveld et al 2011) MatLab implementation of OpenMEEG (Gramfort et al 2011). We refer the reader to Appendix F (E) in which we offer an overview of the Boundary Element Method (BEM), and the numerical method used here to solve the equations defining the electric field (detailed in Appendix F (A)). Our key interest is to compute the lead field – the field observed at each sensor position. A computationally important characteristic of this field is that it can be expressed as the product of a gain matrix that only depends on the dipoles' positions and orientations with the vector representing the dipoles' amplitude (see Appendix F (D) for an overview of this algebraic formulation). This implies that for a given set of anatomically constrained dipoles, expansive computation of the gain matrix needs only to be performed once. The EEG signal generated by a given time course of brain activity, modeled as variations in dipoles amplitudes, can be then simply computed through one matrix multiplication.

### 7.5.4 Processing model

#### Defining dipole amplitude from a conceptual model

We now turn to the processing model whose role is to associate to each dipole a source waveform based on a conceptual model. As noted earlier the long term aim of Synthetic ERP Phase 1 is to infer these waveforms from an underlying neural or schema network model, but here we focus on an alternative problem: Given a conceptual model, formalize a set of quantitative hypotheses on the patterns of temporal activation within the brain and ultimately on the patterns of temporal activation of a set of dipoles. This then provides the input to forward modeling whose results may clarify places where the conceptual model is underspecified and thus specify more precise targets for Phase 1 building of computational network models.

We now focus on the five modules and four processing phases extracted from the 2002 model and shown in Figure 3. It may be best to think of the graph of Figure 3 as anatomical showing pathways feeding information from one module to another but in the case of a feedforward model like the 2002 model one could also interpret this as a graph of temporal precedence. However, this latter interpretation would not be applicable in models that include loops and top-down processing. The crucial point is to break down each module into brain areas or even finer subdivisions of cortex and to specify the time course  $d_A$  of dipole activity for each area so defined. At this level, it becomes irrelevant whether an area  $A$  so specified occurs in one module or many, and whether it is active once or many times during the overall task.

#### Activity modeling

Most conceptual models based on ERP results directly associate the activation time of a brain module by appealing (perhaps implicitly) to the boxcar representation critiqued in our discussion of Figure 4. In order to show the strengths and limitations of such an assumption, we choose in the present model to directly link the  $d_A$  functions (dipole amplitude waveforms), for the brain area  $A$  associated with a given brain module, to the shape of its component within the associated ERP. However, the shape of an ERP component is not always readily accessible from the EEG literature which tends to emphasize the time of occurrence, duration,

and size of the component and these values can be reported in a variety of ways, but there is no consensus (Luck 2005). Moreover, our concern here is to lay the foundations for linking ERP data to detailed models of underlying processing. As we have stressed elsewhere (Arbib et al 2000), the fact that area  $A$  is “significantly more active” in task X than area  $B$  does not mean that there is no activity in area  $B$  during task X. A processing model will provide an account of the time course of detailed neural or schema activity in both areas  $A$  and  $B$  during the task, however ERP results are usually presented without assessing the presence of possible overlaps between components. This makes the quantitative extraction of the shape of a component difficult.

To clarify the issues here, we here suggest a means to extract the N100 component generated during an auditory oddball detection task from the empirical measurement reported by (Scherg et al 1989) from an overall ERP waveform. The challenge is to hypothesize what part of the empirical data given by the solid curve in Figure 6 is actually the N100 component, and what parts of the waveform are extraneous to this module. We took the original waveform and defined the duration time  $D$  of the N100 component as the full width at half the extremum of the peak at 100 ms, spanning the interval  $[T1, T1 + D]$ . We then modeled the shape of the peak by finding the gamma function  $\gamma$  defined on  $[0, 1]$  and normalized in amplitude such that  $\gamma((t - T1)/D)$  results in the best fit for that duration (parameters for  $\gamma$  are  $k=4$  and  $\theta=0.1$ ). From there,  $d_A$  associated with the N100 brain module is defined as  $\gamma((t - T1)/D)$  for  $t$  in  $[T1, T1 + D]$ . Once we have defined the N100 component in this way, the difference between the posited N100 and the actual ERP activity in Figure 8 is then hypothesized to correspond to neural activity in other brain regions. We do not claim that a gamma function is the only option for the fitting of an ERP. It has the property of being defined on the positive real line with  $\gamma(0) = 0$  for a certain range of parameters. This fits with the assumption that the activity of a brain area has an onset time before which its activity is null. However a polynomial fit would of course also be possible but would leave the number of parameters unconstrained.

To generalize this example, recall that a brain module  $M$  is based on the hypothesis that certain brain areas are implicated in a specific function  $F$ . In a feedforward model, we may assume that a given area  $A$  is active only once and only in one module. In general, though, an area  $A$  may be involved in multiple modules. However, we here consider how to extract an estimate for  $d_A(t)$  for the time period associated with the peak (P or N) of a strong deviation in the ERP associated with an anomaly in the execution of function  $F$ . Noting that most modules in Figure 3 are each associated with several brain areas, we make (for now) the simplifying assumption that the same amplitude waveform is associated with each area linked to the given module. Given this, we start with an ERP waveform that has a significant peak posited to correspond to a particular ERP component. Then, just as before, we extract a duration from the peak, and then find the gamma function that best fits the peak during that duration. The resultant curve is our formally defined  $d_A(t)$  for that interval for all areas  $A$  associated with the module for that component. Unfortunately, such curve fitting is inapplicable when the necessary details of the ERP waveform are not reported in the experimental literature.

Given our discussion of the N400 in Figure 1, note that we are looking (at least) for the  $d_A(t)$  for a given area of a given module in two different conditions, with and without the anomaly.

As noted at the beginning of this section, we do not claim that the modeling choices we made in order to extract  $d_A$  were the only ones possible. Our goal was to give a quantitative form to the assumptions underlying the interpretation of ERP components, often implicit in the neurolinguistic literature. The two main assumptions can be summarized in the following statements:

- (1) An ERP component has a given qualitative shape. This shape is roughly the same for classes of components such as N100, N400, and P600.
- (2) The shape of the ERP component is isomorphic to the activity of its underlying brain regions.

We formalize (1) quantitatively by choosing to model each ERP components with a gamma function (best fitted to each component). The direct association between such gamma functions and  $d_A$  formalizes (2). We insist on the fact that such a phenomenological fit of the ERP data is not an intrinsic feature of Synthetic ERP but stems from a desire to formalize what we consider to be the way most neurolinguistic models interpret such empirical results. In doing so we seek a better understanding of how to link computational models to conceptual ones as well as identify the stumbling blocks that qualitative data formats might create for modelers (for a more general discussion of these issues and of the challenges posed to neuroinformatics by

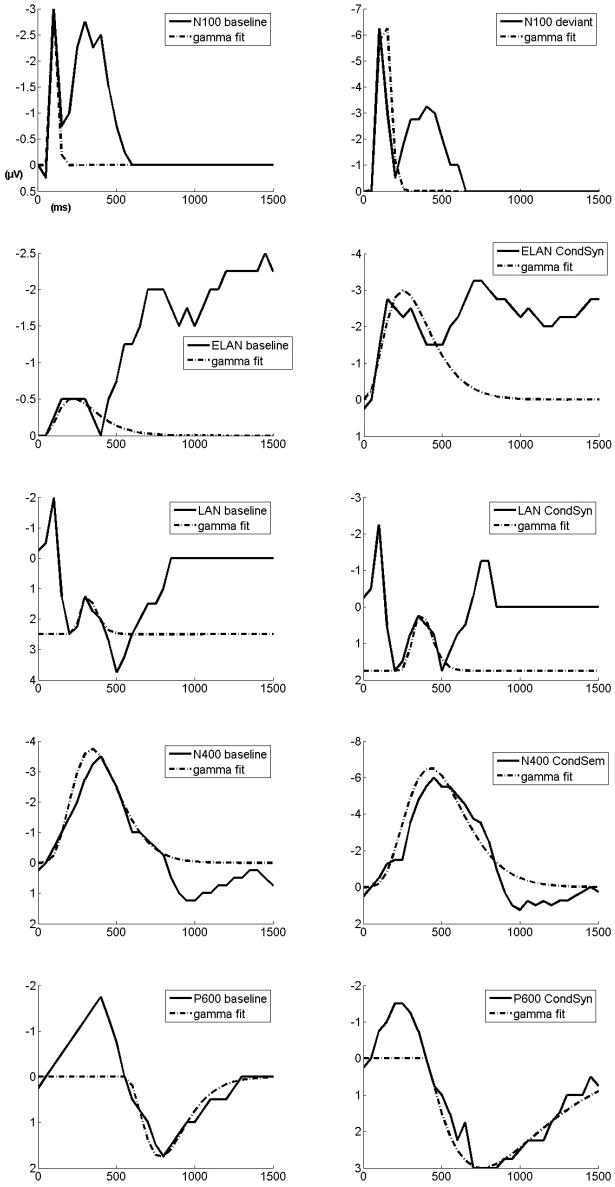


Figure 7.7:  $d_A$  shape definition from the different ERPs. In each case, the solid black line represents the experimental ERP waveform while the semi-dotted line represents the best-fit result to the stated ERP component. Data for ELAN, N400, and P600 are taken from (Hahne & Friederici 2002). Since N100 and LAN are not analyzed in their study, N100 is defined based on (Scherg et al 1989) and LAN based on (Penke et al 1997). Note that LAN presents a specific case where the component is a negative going deflection on a positive baseline value. CondSem refers to the semantic violation condition while CondSyn refers to the syntactic violation conditions. For the brain module source of any given ERP, the shape of  $d_A$  is defined as the best fitting gamma function associated with the peak for the stated component. For the actual  $d_A$  see Figure 9.

Destrieux atlas		Friederici's model	
Hemisphere	Name	Cortical Regions	Modules
left & right	Anterior transverse temporal gyrus (of Heschl)	Primary Auditory Cortex	N100
left & right	Planum temporale or temporal plane of the superior temporal gyrus	Planum temporale	N100
left	Lateral aspect of the superior temporal gyrus [Anterior segment]	Left Anterior STG	ELAN
left	Opercular part of the inferior frontal gyrus	Left Frontal operculum	ELAN
left	Triangular part of the inferior frontal gyrus	Left Frontal operculum	ELAN
left	Opercular part of the inferior frontal gyrus	BA 44	LAN
left	Triangular part of the inferior frontal gyrus	BA 45	N400
left	Orbital part of the inferior frontal gyrus	BA 47	N400
left	Lateral aspect of the superior temporal gyrus [Middle segment]	Left Middle STG	N400
left	Lateral aspect of the superior temporal gyrus [Posterior segment]	Left Posterior STG	N400, LAN, P600
left	Middle temporal gyrus (T2) [Middle segment]	Left Middle MTG	N400
left	Middle temporal gyrus (T2) [Posterior segment]	Left Posterior MTG	N400
left	Superior temporal sulcus (parallel sulcus) [Posterior segment]	Left Posterior STS	P600

Figure 7.8: Conversion of the mixed brain ontology of the 2002 model into a single anatomical atlas, the Destrieux brain surface atlas.

neurolinguistics see (Lee & Barres in preparation)) The goal of future work on Synthetic ERP Phase 1 is to simulated ERPs by applying the forward model to a realistic processing model of the interactions between neural components adequate to produce the observed language behavior.

## 7.6 Simulation Results

We now present simulation results for a forward model computation of ERPs based on the 2002 model of language comprehension. We first detail the information extracted from the 2002 model (using the methods of the previous section) that served as the basis for the forward model. We then qualitatively compare the simulated scalp potential topographies generated by the forward model against empirical data (Simulation results 1: Scalp potential topographic maps). This enables us to highlight some initial issues related to source dipole localization. We finally move on to simulating ERPs based on the 2002 model (Simulation results 2: Synthetic ERPs) and compare our results with the ERP experimental data reported by (Hahne & Friederici 2002) in Figure 4.

### 7.6.1 Mapping the 2002 model onto cortical geometry

As in Figure 3, the 2002 model defines five brain modules associated with the different phases of sentence comprehension with each considered to be the source of a specific ERP component. The brain regions associated with these brain modules are referred to using a mixed ontology of gyrus/sulcus neuroanatomical landmarks along with cytoarchitectonically defined Brodmann areas and functionally defined sensory areas. We mapped these regions onto the Destrieux atlas that offers a cortical surface parcellation based on cortical geometry. We based the conversion on the description of the Destrieux atlas given by (Fischl et al 2004). The result of this conversion is presented in Table 2. Brodmann ontology is frequently used in neurolinguistics. However, given the sensitivity of the EEG sources to the grey matter gyral geometry, idiosyncratic variations in the gyral localization of Brodmann areas are a limitation to their usefulness in Synthetic ERP (see Amunts et al 2010 for a discussion of these issues, focusing on Broca's area).

Figure 7 presents the faces of the realistic brain mesh associated with the different brain modules in the head model. Our definition of the LAN module is more extensive than in Figure 3, because the 2002 model does not make precise what parts of the frontal operculum need to be included. We were able to include the

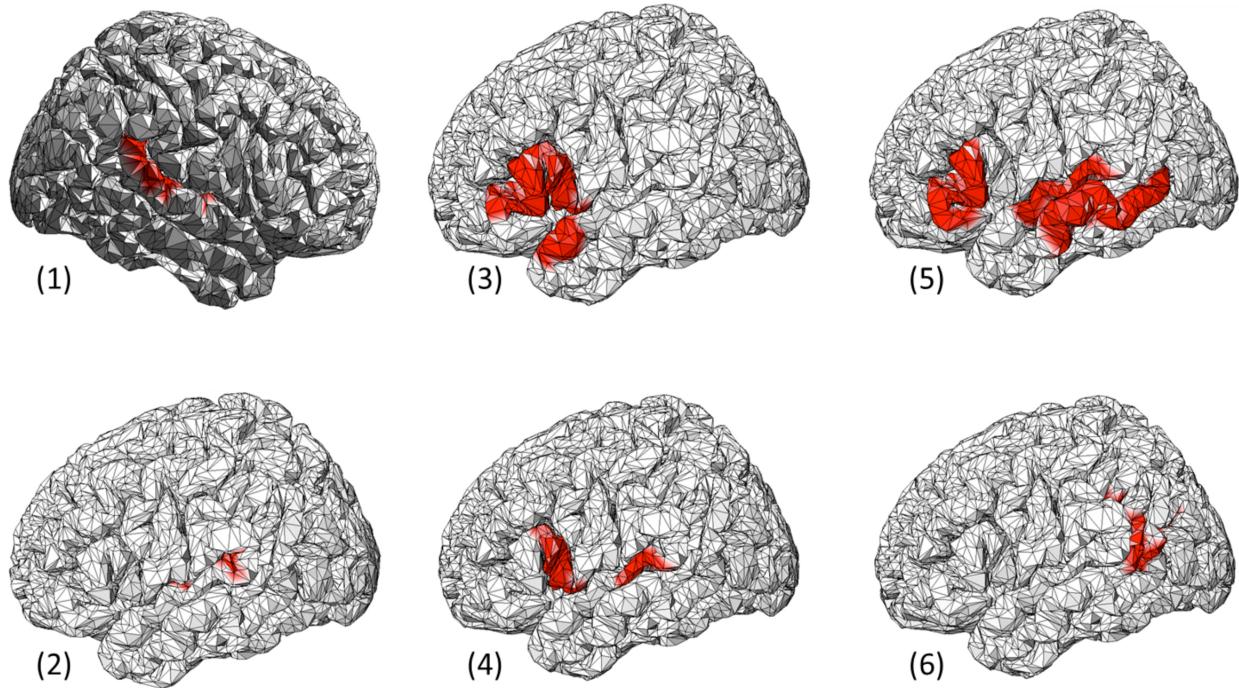


Figure 7.9: Faces of the realistic brain mesh associated with the 2002 model in the head model for dipole localization. All the areas are defined using the Destrieux anatomical atlas. (1) & (2) N100 module. The bilateral Heschel's gyrus is not visible here. (3) ELAN module. (4) LAN module. (5) N400 module. (6) P600 module. (For the precise anatomical regions and their comparison with the 2002 model see Table 2).

anterior, middle, or posterior parts of the superior and middle temporal gyri as required by the 2002 model, but such regions do not have a standard definition and are therefore not part of the Destrieux atlas. Finally, it is important to note that the 2002 model includes the basal ganglia in the P600 brain module but we do not include any subcortical structures in our current head model.

Single average dipoles were associated with each brain area which was part of at least one brain module. When simulating scalp potential topographies, the lead field was computed at every vertex of the top part of the outer scalp mesh. Otherwise, the Brainstorm 10/10 65 channels default electrode positions were used as computation points for the lead field. Conductivities were kept at the values defined by (Oostendorp et al 2000).

### 7.6.2 Processing model: Brain modules and activity timing

We use the connectivity defined in the graph of Figure 3. Due to the serial nature of the 2002 model, the connectivity is interpreted as a graph of temporal precedence. For each brain module  $A$ , the activation times  $T_1(A)$  (“on Time”) and  $T_2(A)$  (“off Time”) are defined as follow.  $T_2(A)$  is systematically defined as  $T_1(A) + D(A)$  where  $D(A)$  is the full width at half extremum of the ERP component associated with  $A$ . For the input module, N100 module representing the auditory sensory areas, the  $T_1$  was set to 20ms as the time required for the auditory input to activate these brain areas (Rupp et al 2002). For the other modules, the  $T_1$  was defined as the “off time” of the preceding module to which 20ms were added to account for transfer delays. Other delays such as those inherent to the dynamics of neural computation (in particular post-synaptic potential buildup delays) are thought of as already lumped into the definition of  $d_A$ . Phase 1 will assess the requirements these impose on neural and schema networks. In the case of multiple inputs (see P600 module), we made the assumption that both preceding processes needed to be done before the next processing could start.

Activation times are given for each module and for each experimental condition in Fig. 7.12. In the

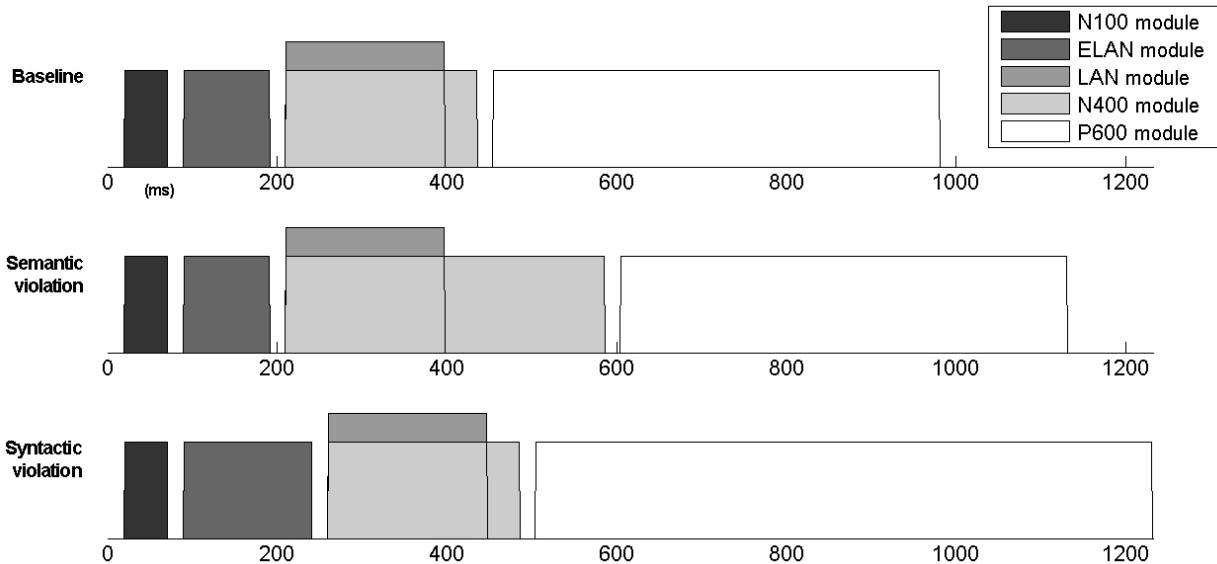


Figure 7.10: Activation durations for each brain module for ERP component sources in the various experimental conditions. The durations of activity for each module were extracted from experimental data as the width of the respective component at half maximum. (LAN boxcar is shown higher so as not to be obscured by N400.)

semantic condition, only the activity of the N400 module changes duration from the normal condition following the reported effect of such violation on the N400 ERP. In the syntactic violation condition, both ELAN and P600 modules activation times are affected since this violation triggers a change in the ELAN and P600 components (according to Hahne & Friederici 2002). Figure 9 presents the  $d_A$  associated with each brain module. They are given by the gamma functions extracted from the empirical ERP data as detailed in section 4.4. However, to follow more closely the hypotheses made by the 2002 model, the onset of activity is now determined by the onset times we just discussed and not by the onset of the ERPs to reflect the serial computation hypothesis.

We do not claim to have the capacity to extract the exact timing of activations or activity level of brain modules from ERP but for now follow the general assumption that the ERP duration reflects the duration of the associated brain processes while the ERP amplitude reflects their activity levels. These claims are pervasive in neurolinguistics and the role of the present work is to show both their strength but also their limitations in the framework of Synthetic ERP. Once again, future work on Synthetic ERP Phase 1 will relax such assumptions since the timing of activations and activity levels of brain areas should then directly result from activity patterns of the neural or schema model.

### 7.6.3 Simulation results 1: Scalp potential topographic maps

For each ERP component, we now compare the scalp potential topography generated by simulation of the baseline level activation of its associated brain modules with an empirical measurement of the scalp potential topography. Each brain area composing a module is represented by a single dipole averaging its curvature and proportional in magnitude to the surface of the area. Appendix F (B) makes explicit the fact that such an averaging would be an acceptable approximation of a brain region in a general case under the hypotheses that this region is small and that its curvature is negligible. The assumption we make here aims therefore at illustrating the problem often ignored by conceptual models that EEG sources need to be considered in their spatial extension - with the exception of very focal activities as in the case of an epileptic focus<sup>2</sup>.

<sup>2</sup>From an historical point of view, identifying an epileptic focus has been one of the driving forces behind the development of EEG source localization methods. This could partially explain why so much of the literature circles around modeling brain activity through the use of one or a few dipoles. This fits the hypothesis that only a few localized patches of cortex are generating

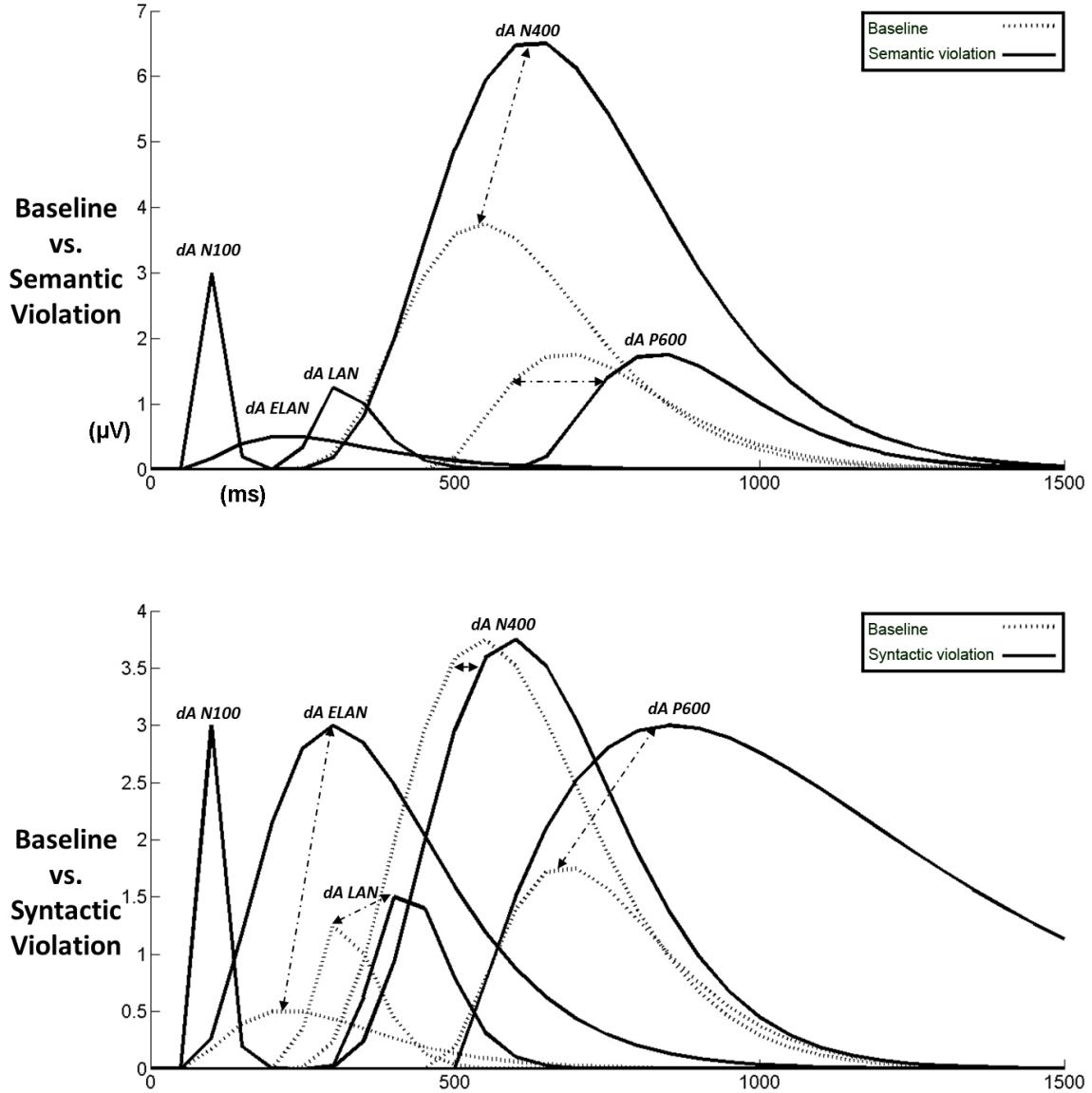


Figure 7.11:  $d_A$  associated with each brain module for the three experimental conditions. In the semantic violation case, the  $d_A$  for the N400 module changes in duration and activation level. This triggers a change in onset time of the P600 due to the serial processing hypothesis. In the syntactic violation condition, the  $d_A$  for the ELAN changes in duration and activation level compared to control as well as P600. The change in ELAN duration impacts all the activation times downstream. (Note that the activation scale differs between the top and the bottom graph). Dotted arrows identify, when needed,  $d_A$  function associated with a same module.

ERP components show variability in their associated scalp distributions just as they show variability in their shape, timing, and magnitude. However, this initial comparison enables us to get a first estimation of the validity of the simulations at the stage where the only assumption results from what could be extracted from the conceptual model concerning the neural substrates associated with each brain module.

As shown in Fig. 7.12 , the simulated topographies did not qualitatively match the empirical ones for all the brain modules. In the case of the N100 module and the LAN module, the results are close to the empirical measurements with the following caveats. The N100 is more right lateralized. The maximum negativity is achieved for the electrode C6 as opposed to Cz for the empirical result. This asymmetry in our simulation could be due to the anatomical asymmetry of the bilateral brain areas defined as the sources of the N100 (see fig. 7.11(1) & (2)). The topography for the simulated LAN is roughly correct with a negativity slightly posterior of the empirical one (minimum for electrode T7/T3 for the simulation compared with F7 for the empirical data). For ELAN, N400, and P600 the simulated topographies are blatantly incorrect. However, we could find for these a face in the anatomical region attributed to each brain module whose associated dipole yields a greatly improved potential topography. For ELAN the result is more left lateralized than in the empirical measurements. The same is true of the N400 simulated topography whose asymmetry differs from the overall symmetry of the empirical one. As for P600, once again the simulated distribution is shifted to the left side of the head compared to the empirical data. This general tendency to be left lateralized could reflect the absence of consideration of the role of the right hemisphere in language processing in the 2002 model. The N400 is the most obvious case and it has been reported that the right hemisphere plays an important role as a source of this component (Maess et al 2006).

This shows the importance of the careful consideration of the geometry of the cortical areas whose activity generates an ERP. A coarse anatomical definition of the brain areas associated with a processing module in a conceptual model does not provide in most cases enough information to simulate an ERP-associated scalp potential topography. Although averaging the curvature of an area can give reasonable results, in most cases it is necessary to search for the correct cortical patch within an area whose curvature and position result in a field that fits the ERP empirical topography. It is important to note however that such a search differs from most of the inverse solution models that look for best fitting dipole localization and orientation without any anatomical constraints. Although insufficient, the brain modules defined in the 2002 model, once associated with a realistic mesh modeling the grey matters folds, considerably constrain the search space to a finite set of dipoles. A yet even more realistic approach would be to model brain regions by a distribution of dipoles over its surface rather than by a single dipole (in the spirit of Bojak et al. as described in 1.3). However, it will be the role of Phase 1 to analyze how to link a neural nework representation of the distributed computation occurring in a brain region to a geometrically accurate model of cortical activity.

#### 7.6.4 Simulation results 2: Synthetic ERPs

Given the preceding results, we modified the head model in the following way. For the N100 and LAN brain modules we kept their associated dipoles defined as average area dipoles. For the ELAN, N400, and P600 brain modules for which the average area dipoles gave incorrect scalp potential topographies, we kept the better fitting single dipoles described above. In order to keep the contribution of the areas comparable, the dipoles for all the areas (including the ones associated with the N100 and LAN brain modules) were given an equal magnitude. This is tantamount to saying that we remove the area surface factor and consider all the dipoles to summarize a patch of cortex of equal surface. In the case of the single average dipoles, that means that we kept the location and orientation but that the dipole is not thought to model the activity of the whole area but some distributed portion of this whole area.

We compare the empirical data extracted from (Hahne & Friederici 2002) and our simulated ERP. For simplicity, we only consider here the empirical ERP time course as recorded at the most significant electrode (Cz for N400, F7 for ELAN, and Pz for P600). In the absence of noise in our system, we simply define our “most significant electrodes” as the ones where the maximum peak amplitudes for a given component are observed (TP7 for N400, F3 for ELAN, and PO3 for P600). Fig. 7.13 provides a visual representation of the differences of localization in the most significant electrodes between the empirical and simulated data. ELAN and P600 appears to be roughly maximal in the same scalp quadrant as for the empirical data (anterior left

---

the EEG signal during an epileptic seizure. However, it is clear that in the more general case as in language processing the number of cortical regions as well as their size would require to revise such view.

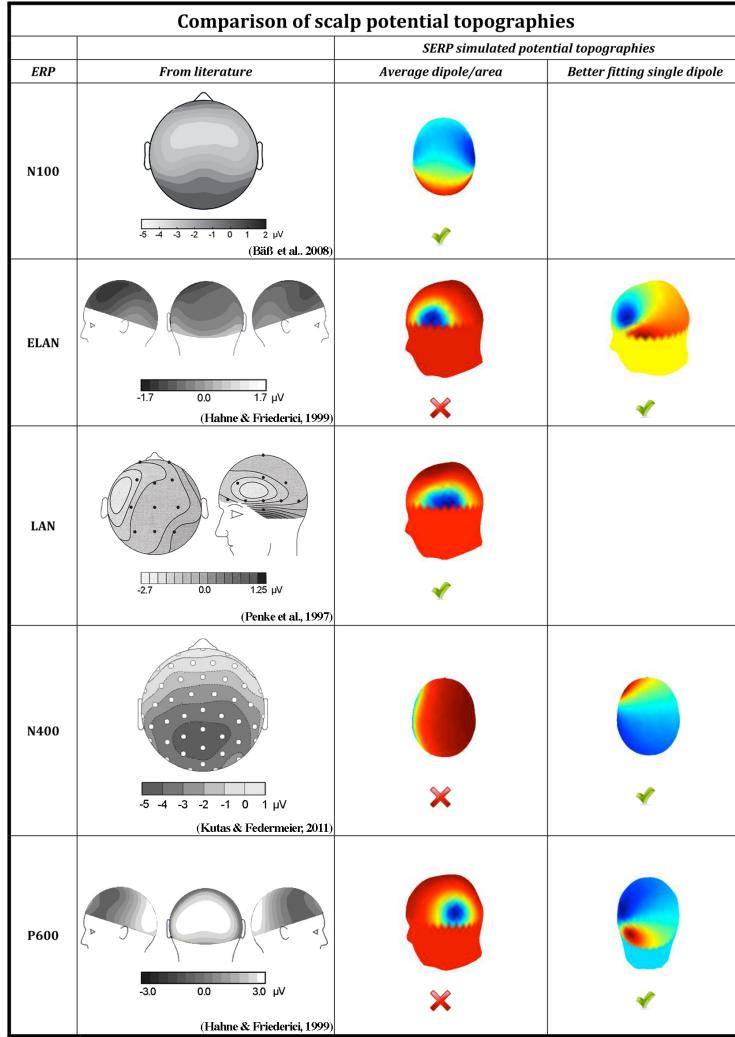


Figure 7.12: Scalp potential topographies. Comparison of empirical data and simulated fields. Each row represents the data and simulation for the ERP effect listed in the leftmost column. The second column from the left depicts the topographies of the potentials associated with the ERP as extracted from the literature (Bass et al 2008, Hahne & Friederici 1999, Kutas & Federmeier 2011, Penke et al 1997). Rather than displaying waveforms as in Figure 3, we display scalp potential topographies as smoothed “heat maps” for the moment at which a given potential reaches its peak. In the absence of a standard way to report these topographies, we did not try to find matching representations and the patterns should be taken as qualitative depictions. The two columns on the right present simulated scalp topographies based on the activation of the brain module associated with the ERP by the 2002 model (lower potential values in blue, higher in red). The “Average dipole/area” column present the simulated topographies based on single average dipole for each area. If the N100 and LAN components coarsely match the empirical data, ELAN, N400, and P600 topographies are incorrect. In the rightmost column, for these incorrect cases, we found a single dipole associated with a face belonging to a brain module whose resulting simulated topography fits qualitatively better with the empirical topography.

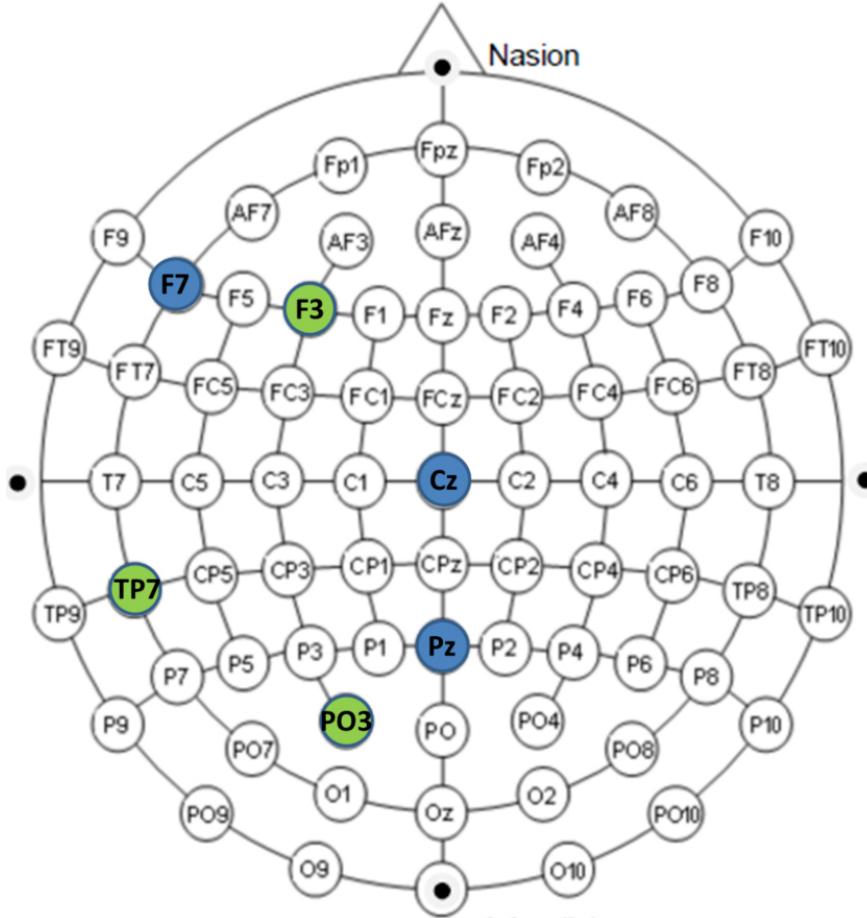


Figure 7.13: Comparison of the empirical most significant electrode position for (Hahne & Friederici 2002) (blue) and Synthetic ERP (green). For the empirical results, the most significant electrodes for the measurements of ERP are: F7 for ELAN, CZ for N400, and Pz for P600. In the case of Synthetic ERP they are F3 for ELAN, TP7 for N400, and PO3 for P600. If the simulated most significant electrodes for ELAN and P600 are roughly similar to their empirical counterparts, the case of N400 shows a clear left lateralization in the simulation absent from the empirical data (see discussion in Simulation results 1: Scalp potential topographic maps).

for ELAN, posterior left for P600). N400 maximum amplitude on the other hand is highly left lateralized in the simulated case when it is central in the empirical data. As mentioned in Simulation results 1, this could tie back to the assumption of purely left lateralized sources for N400. Future work should try and compare this result with those resulting from the inclusion of right lateralized brain areas as N400 sources as suggested by (Maess et al 2006).

From these comparisons, it seems that the level of precision one could expect from Synthetic ERP simulations should clearly be lower than level of granularity of scalp regions provided by the 10/10 65 electrodes cap positions. In linking empirical reports of ERP to computational simulations, an average value over a few electrodes covering a standard brain region might be more appropriate. This states the challenge both for Synthetic ERP phase 1 with a need to assess more precisely what the scalp localization precision could be. But it also should highlight the role computational Synthetic Brain Imaging endeavors to play in assessing the correct level of representation of empirical results providing suitable quantitative summaries for modelers.

Fig. 7.14 presents the comparison of the empirical ERP time course and the simulated ones based on our forward model. In both the empirical and Synthetic ERP case, the solid line represents the control time

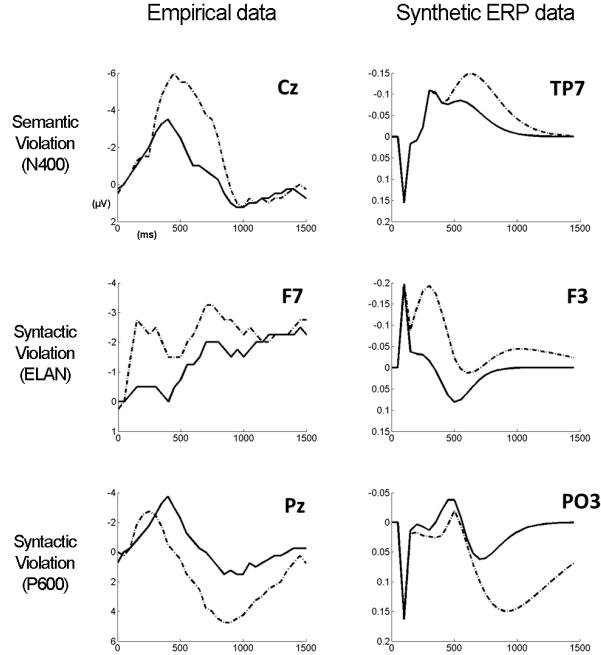


Figure 7.14: Comparison of ERP time course as extracted from empirical data (Hahne and Friederici, 2002) and simulated using our forward model. In each plot, the solid line represents the ERP time course for the baseline condition while the dotted line represents the ERP time course for the experimental condition. The leftmost column gives the experimental condition associated with the row as well as the ERP component associated with such condition. For the empirical data, we selected the recordings at the most significant electrodes. For Synthetic ERP, the “most significant electrodes” are defined as the ones where the maximum peak amplitudes for a given component are simulated.

course while the dotted line represents the time course in the experimental condition mentioned on the right of the figure. Once again we did not aim at simulating the exact potential values but the relative variations between the baseline and the experimental conditions. These simulations raise the following issues.

Looking at the P600 simulation, the first clearly apparent result is that, if we replicate the late positivity, the temporal relation of the P600 component in the baseline and in the syntactic violation condition differs between the empirical and simulated results. Such a difference results from the fact that we did not directly choose the onset time of our brain module’s activations based on the ERP but derived them from the serial processing hypothesis made by the 2002 model. As shown in fig. 7.10, the onset time for the activation of the brain module hypothesized to generate the P600, given the hypothesis of serial processing, does not markedly differ between the baseline and the syntactic violation condition (interestingly such a difference in onset is clear however in the semantic violation condition). The only a minor difference in activation times is due to the increase by 50ms of the duration of the ELAN downstream which delays all the following processes.

Turning to the ELAN component, the simulation shows a large negative deflection peaking in a time window [100300ms] similar to the one presented in the empirical data. However, the general trend of negative going deflection displayed by the empirical time course in both baseline and syntactic violation conditions is not replicated. This points towards the limitations of our approach but also of the empirical reports of ERP results which tend to ignore some features of the measured signal that would however provide interesting benchmark cases for testing a modeling effort.

Finally, for the N400 the model does replicate the increase in duration and magnitude of the N400 component when contrasting the semantic violation condition to the baseline. If we abstract the large initial positive peak with results from the early activation of the N100 brain module, the shape of the empirical N400 is relatively well replicated, including the initial absence of noticeable difference between the anomalous

and baseline conditions until 500ms when a clear negative going deflection can be observed in the semantic violation condition.

## 7.7 From Preliminary Results to Emerging Challenges

### 7.7.1 Source dipole modeling and cortical geometry

We associated the cortical regions defined in the 2002 model to source dipoles by first mapping them onto our realistic brain mesh, and then averaging the normals to the surfaces of these areas to generate a representative dipole orientation. An important challenge concerned the 2002 model’s use of multiple ontologies to define neural substrates of cognitive processes and ERP components. Functional regions were used alongside cytoarchitectonically defined Brodmann areas and coarse gyral ontology. Since the critical factor for modeling the source of the electromagnetic field is the geometry of the grey matter surface, functional regions or Brodmann areas need to be converted to their associated grey matter surface location since only then can the orientation of the area’s surface be retrieved. Moreover, the orientation may vary across a region, indicating the challenge of linking functioning neural circuitry to subregions or even voxels within a given region. We used the Destrieux atlas as a unified ontology providing the correct amount of detail to recast the mixed ontology of the 2002 model into a single brain atlas – an approach which highlights the lack of unified standards for brain region (and subregion).

As has been discussed previously in Section 1.3, the ill-posed nature of the electromagnetic inverse problem means that no unique source model can be obtained without the use of fallible a priori constraints. Accordingly, this approach can only provide a general localization of neural activity, such as the 2002 model’s claim that a contributing source for the N400 resides in BA 45. As shown by Synthetic ERP simulations in Figure 10, such limitations do not allow for faithful simulation of the canonical N400 scalp distribution (among others). Two solutions can palliate these limitations. The first one consists in finding within these areas the cortical patch providing the best fitting solution. If in the present chapter we simply hand picked a better fitting dipole, future work could investigate the possibility of using the apparatus involved in computing the inverse solution while constraining the search space to the source dipoles (or combination of source dipoles) associated with a given brain area. An example of a related approach to EEG source reconstruction including anatomical constraints on their solution space is given by Phillips et al (2002). Another option would be to make direct use of the coordinates where fMRI revealed a significant increase of BOLD signal and posit this as the location of the relevant dipole. However, we have stressed that detailed processing models will in general show that a given epoch of processing for a given language subfunction will in general involve the competition and cooperation multiple neural circuits beyond those restricted to areas of most significant BOLD activity and this observation motivates our work on both Synthetic fMRI and Synthetic ERP, with the long-term goal of developing “doubly causal” models (more in Section 7) that can predict both fMRI and ERP signals under varied task conditions.

### 7.7.2 Activation modeling

We provided a way to use ERP data as a basis for modeling the time course of activations in brain areas that goes beyond a simple modeling of activation as a boxcar function defined on a period directly mapped in onset and offset time of an ERP component defined by some form of anomalous input. Our attempt to simulate the ELAN, N400, and P600 ERP components reported by (Hahne & Friederici 2002) provided new insights into the challenges of relating neurolinguistic data and models to empirical ERP data through synthetic brain imaging. Nonetheless, this does not obviate the need for detailed processing models (Synthetic ERP Phase 1).

### 7.7.3 Quantitative ERP data extraction from literature

We have laid bare the difficulty of extracting quantitative ERP data from the literature in a format suitable for Synthetic ERP computational modeling. The appropriate data format needs to include the shape of the component or EEG trace measured under a variety of experimental conditions. Our ELAN simulation showed not only that a computational model should replicate the negative going component at around 100msec

but also the negative-going trend all through the epoch time. The common methodology of describing a component at the locus of electrodes as opposed to their average over larger scalp region also poses problem, as the electrode level precision seems too small a scale for Synthetic ERP. It also poses problems for the definition of ERP components themselves since a “most significant” electrode location cannot be empirically assigned to a component across subjects and experimental designs. Finally, the N400 simulation highlights the difficulty of clearly defining an ERP component when its particular field contribution is superimposed upon those from other concurrent processes, or when the component is known to have several functionally and scalp-voltage topographically distinct subcomponents as in the case of N100 or N400. We anticipate such superimposition of fields and diversity of subcomponents to be the norm for any sufficiently detailed description of neuroelectromagnetic dynamics associated with cognitive processing, and therefore recognize that qualitative accounts of ERP data the limiting factor for any modeling or comparative exercise. In particular, we stress that many different neural processes may underlie a negativity occurring in a range around 400 ms, and so we may expect progress in neurolinguistics to depend strongly on the ability to discriminate different “N400s” rather than speaking of “the” N400.

## Chapter 8

# Toward a Neurocomputational Model: The Challenge of Brain Anchoring.

*“Je nomme enchantements ces immenses actions jouées entre deux membranes sur la toile de notre cerveau”*  
*(“I name enchantments these immense actions played between two membranes on the web of our brain.”)*

Balzac

Seraphita

### 8.1 An Attempt

In contrast to more artificial intelligence oriented endeavors that also follow such schema theoretic principles such as VISION and HEARSAY, each new step in the design of the SALVIA model is motivated not only by the desire to enrich the computational scope but also by some specific empirical data points that constrain the new development and provide a new computational challenge. The SALVIA model for production focused on the simulation of the variability in the form that scene description take under different time pressure as well as on the simulation of the link between these various form and eye-tracking data. Turning to comprehension, SALVIA incorporates neuropsychological findings on agrammatic aphasics simulating both the coordinated role of world-knowledge, visual, and grammatical information during comprehension, but also the possible for light and heavy semantic constraints to be selectively impaired.

Linking more precisely the SALVIA architecture to the known neuroanatomy of the language system will require incorporating neuroimaging data into the model. In the last decade, the joint use of fMRI and DTI data has led to the identification of at least two different neural pathways involved in language processing (see figs. 8.1 & 8.3).

- A ventral pathway (1) running along the superior temporal gyrus (STG) linking auditory cortex (AC) and its neighboring areas to the inferior frontal gyrus (IFG) and involving the extreme capsule (EC) white matter tracks.
- A dorsal pathway (2) linking the AC through the arcuate fasciculus (AF) to the motor region involved in speech production or to the dorsal portion of the inferior frontal gyrus (IFG) (pars opercularis).
- These two pathways might be supplemented by at least one other dorsal pathway (3) linking AC and IFG indirectly through the angular gyrus (AG) through the superior longitudinal fasciculus (SLF) (although the distinction between SLF and AF is difficult using DTI).

- And finally another ventral pathway (4) could be involved running along the inferior part of the temporal gyrus and connecting the anterior temporal lobe to the anterior part of the IFG through the uncinate fasciculus (UF) (Friederici, 2012, 2009).

The existence and functional role of these various pathways are still debated. Some consensus seems to emerge on a possible role of (2) in linking sensory and motor representations of word forms. (1) has been shown to be involved in processing local syntactic structures (Bahlmann et al., 2008) but is also thought to be involved in various aspect of semantic processing (Saur et al., 2008). This latter role is supported by the evidence gathered from patients suffering from semantic dementia on the specific role that the anterior temporal lobe (ATL) could play as a semantic-hub on which converge the semantic representations encoded in a distributed way in modal cortical regions (Binder and Desai, 2011). Finally, the role of the inferior frontal gyrus in syntactic processing has been emphasized by various conceptual models (Hagoort, 2013; Bornkessel and Schlesewsky, 2006; Hickok and Poeppel, 2004).

Turning to SALVIA, the two-route model of language comprehension can be only for now roughly related to these neuroimaging results. The sensory-motor portion of the language system mentioned in ch. 5, fig. 5.1 that is involved recognizing (or producing) word forms would encompass the dorsal pathway (2). The grammatical working memory could be instantiated in the BA 44 and BA 45 of the inferior frontal gyrus (Broca's area). However recent neuroimaging data using a multi-voxel pattern analysis (MVPA) decoding method to analyze the brain regions specifically involved in processing argument structure constructions pinpointed BA 47 and the anterior portion of BA22 (Allen et al., 2012). This result points out to a possible involvement of one of the ventral path in processing the semantic constraints associated with the constructions.

The question of the relation between syntactic encoding and decoding is computationally partially tackled by SALVIA: both production and comprehension consist in building construction assemblages in grammatical working memory through cooperative computation. The model remains agnostic as to whether the construction instances stored in long term memory and invoked during production and comprehension are the same. Linguistic work on idioms has distinguished between encoding and decoding idioms (Makkai, 1972). A hearer could figure out the meaning of an encoding idiom when she first encounters it although as a speaker she would not have guessed that these expressions are grammatically correct (e.g. "answer the door") while one needs to learn the conventional meaning of an decoding idiom to be able to understand and use it (e.g. "he kicked the bucket" or "he pulled a fast one"). From a usage base perspective, such differences between encoding and decoding can be extended to all constructions with speakers having their own idiosyncratic encoding preferences at the word, idiom, up to argument structure level, while decoding expectations are shaped by the landscape of input that the speaker receives. So if the question of the relation of between the grammatical knowledge stored in long term memory for production and comprehension remains to be better analyzed in SALVIA, the fact that a single brain system supports both the encoding and decoding grammatical working memory finds support in recent behavioral and fMRI results.

Using a paraphrasing protocol in which participants are asked to verbally paraphrase sentences presented in fragments, Kempen et al. (2012) were able to show that during this grammatical multi-tasking paradigm syntactic expectations generated during decoding about what input would follow a given fragment are immediately replaced by expectations stemming from the paraphrase produced. This result provides empirical evidence that both production and comprehension have at least in part access to a shared "grammatical workspace". The computational and theoretical consequences of this finding are discussed in detailed in (Kempen, 2014). fMRI adaptation studies have reported suppression effects in BA 44 and BA 21 after repetition of a similar syntactic structures either in production or in comprehension (Menenti et al., 2011). Such result supports the hypothesis that similar brain regions support comprehension and production but does not touch upon the question of shared processes. A following study therefore used the same paradigm but this time found suppression after cross-modal repetition (i.e. with intermixed comprehension and production of similar syntactic structures), adding a strong support to the idea of shared syntactic processes between production and comprehension (Segaert et al., 2012).

As a model that link production and comprehension, the next step in the development of the SALVIA model will need to address the relations between grammatical encoding and decoding from a computational perspective as well as in relation to possible shared neural substrates. In doing so, SALVIA will start offering neurocomputational insights into the neural processes that endow humans not only with the capacity to speak but also with the capacity to carry out conversations in which production and comprehension enter

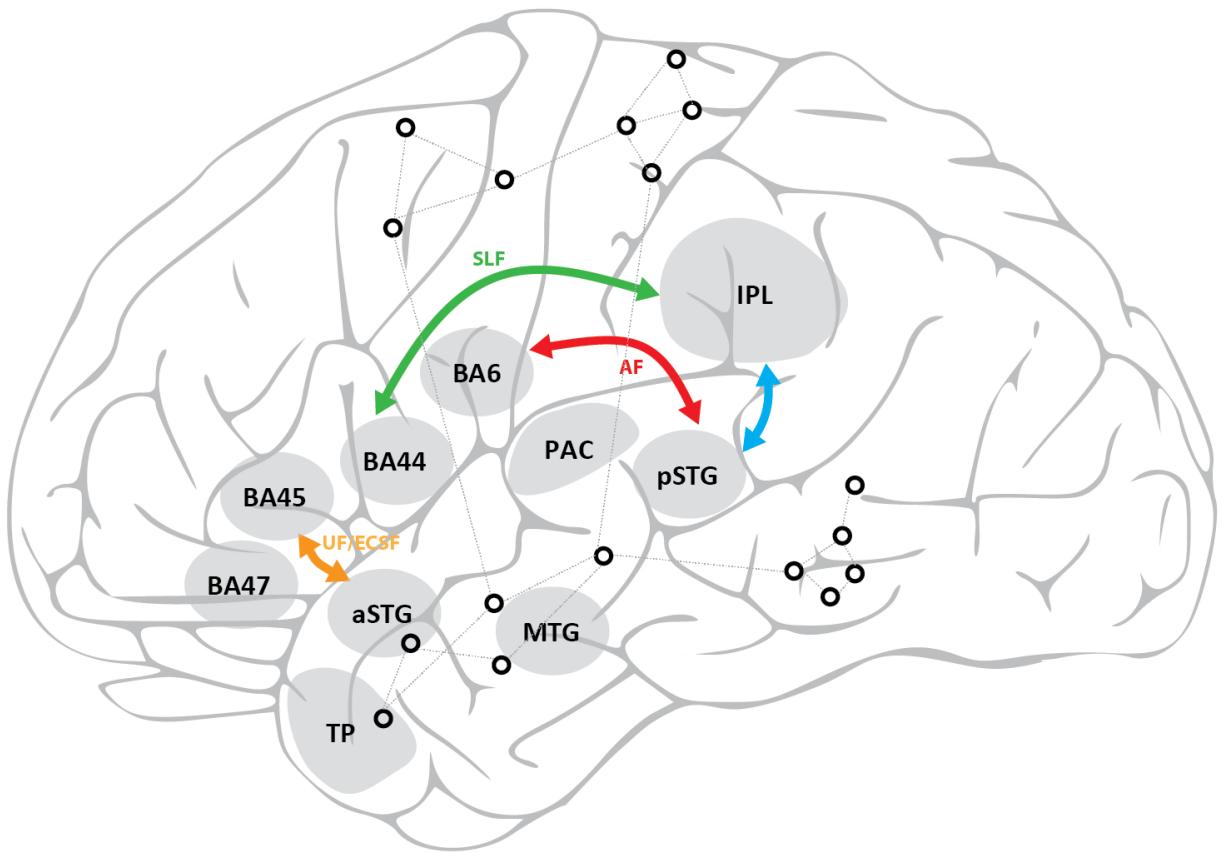


Figure 8.1: Connectivity and organization of the left perisylvian region. Main language relevant anatomical regions and white matter connections. **Arrows:** White matter connections. AF: Arcuate Fasciculus, SLF: Superior Longitudinal Fasciculus, UF: Uncinate Fasciculus, ECSF: Extreme Capsule (Notation UF/ECSF indicates that the distinction between the two is difficult in Diffusion Tensor Imaging (DTI) studies). **Marked regions:** language relevant anatomical regions as they are often noted in the neurolinguistic literature using a mix of various brain ontologies. BA: Broca's area. PAC: Primary Auditory Cortex, STG: Superior Temporal gyrus (anterior (aSTG) and posterior (pSTG)), MTG: Middle Temporal Gyrus, TP: Temporal Pole, IPL: Inferior Parietal Lobule. **Graph overlay:** Conceptual reminder of the distributed nature of the semantic network that involves the motor modal content (top-left cluster), the somato-sensori modal content (top-right cluster), abstract (and/or possibly linguistic/high level vision) content (bottom-left cluster) and visual modal content (bottom-left cluster). This network should only be thought of as a way to make clear that most of the semantic network involved in the comprehension or production of meaning involves a large set of distributed regions including both modal and multi-modal areas.

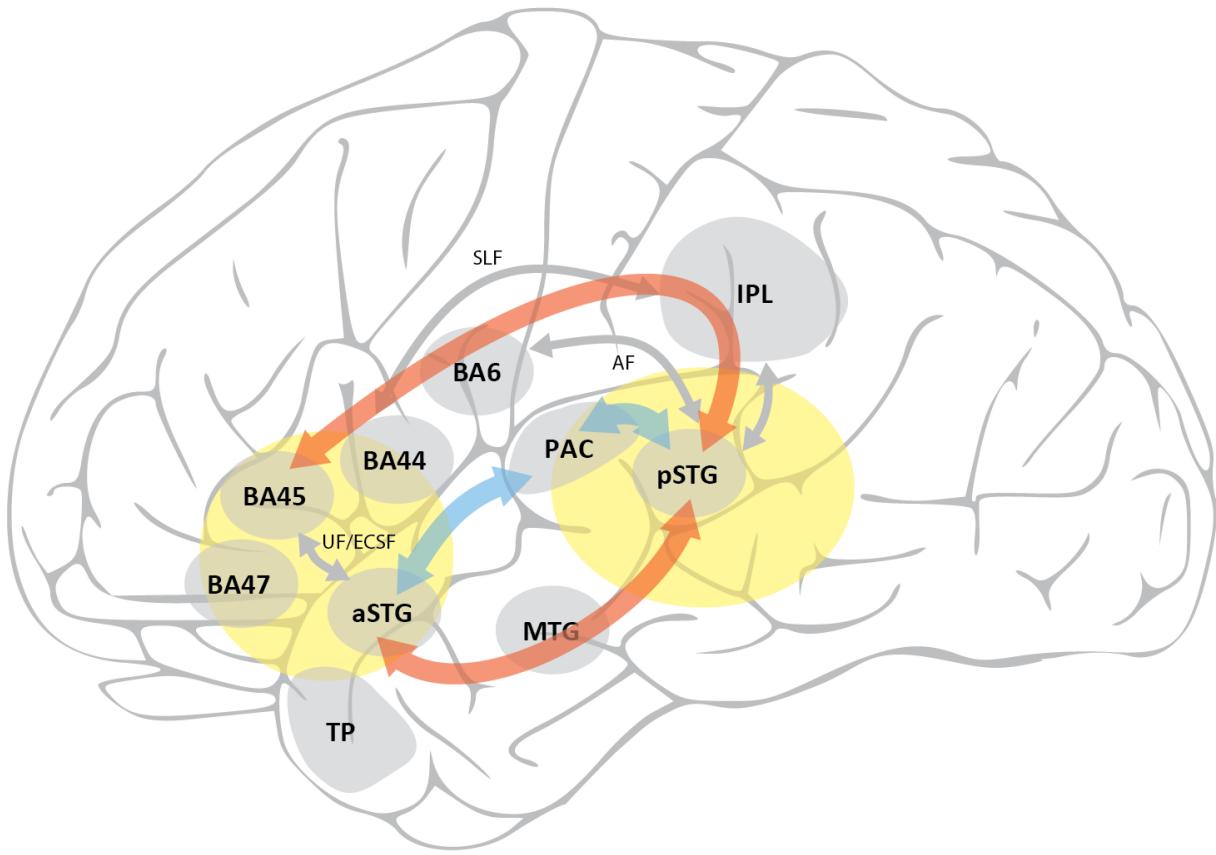


Figure 8.2: A functional neurolinguistic hypotheses for comprehension. Two main functional routes (red arrows). A dorsal route hypothesized to be involved in linguistic form based processes above the word recognition level. This route consists of a system involving the posterior Superior Temporal Gyrus (pSTG) and Broca's area (BA 45) through the Inferior Parietal Lobule (IPL). It is therefore to be distinguished from the system formed by Primary Auditory Cortex (PAC)/pSTG and BA6 connected through AF and that is the more classic system involved in sensory-motor word-form recognition (whose damage result in difficulty in repeating words). A ventral route hypothesized to be more directly involved in meaning-based processes. It consists of a system encompassing pSTG and connecting to anterior Superior Temporal Gyrus (aSTG) through the Middle Temporal Gyrus (MTG). Blue arrows: connections between PAC and both aSTG and pSTG. This system is hypothesized to be involved in retrieval of lexical form (through pSTG and then BA6) and meaning (through aSTG). The distinction between lexical and non-lexical construction meaning is maintained here in order to account for the dissociation between agrammatic aphasia and anomia. Yellow shaded areas denote loci of integration/unification between the concurrent processes taking place in the dorsal and ventral routes. The anterior one corresponds coarsely to Broca's area while the posterior one points to the general region of Wernicke's area. The contention would be that the anterior integration area plays a more direct role in structure building at various level of semantic coarseness (from clause to discourse levels) (Hagoort and van Berkum, 2007; Bornkessel-Schlesewsky and Schlesewsky, 2013) while the posterior integration regions is more directly involved in associating the priming effects and expectations derived from both routes in order to set up a context allowing for efficient integration of upcoming linguistic content (Brouwer et al., 2016)

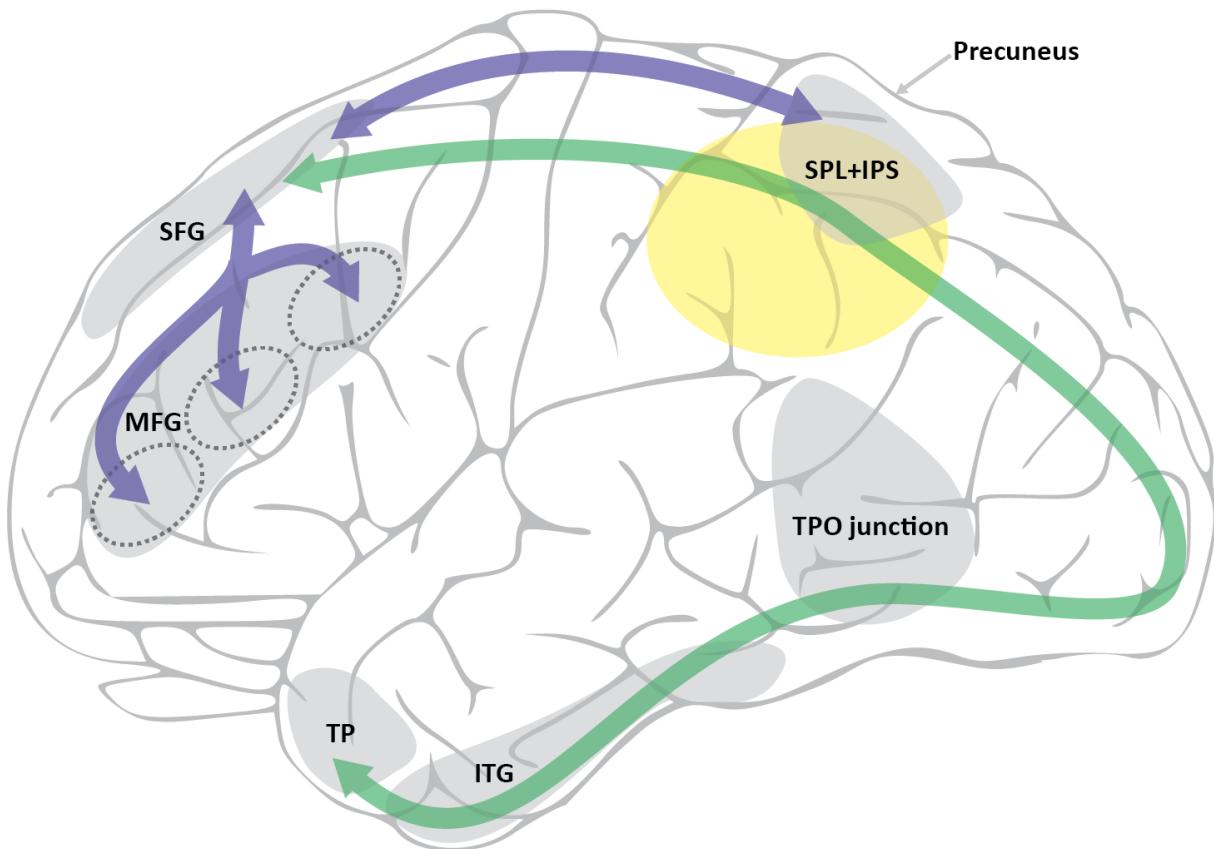


Figure 8.3: Pathways of visual processing and cognitive control. The language system in its role for language use cannot be understood outside of the context of the other main brain systems it interacts with. Here the focus is placed in the visual processing (green arrows) and cognitive control systems (dark blue arrows). Cognitive control as part of a general high level reasoning network involves the anterior Superior Frontal Gyrus (SFG) and the posterior complex composed of the Superior Parietal Lobule (SPL), the Intraparietal Sulcus (IPS) and the Precuneus. In addition to this system generally thought to be involved in high level cognitive tasks, the cognitive control system more directly rests on the Middle Frontal Gyrus (MFG) that display a antero-posterior gradient, that reflects the distal vs. proximal nature of the control goals. The most anterior areas are involved in keeping track of distal plans which then operate a control over the more proximal plans that the more posterior regions track, and this repeats in a cascade fashion, all the way to the motor control level (Koechlin et al., 2003; Badre, 2008; Kouneiher et al., 2009). The visual processing system is split into a dorsal and ventral route following the classic “how” versus “what” distinction. The “what” or visual semantic route involves a system composed of the Inferior Temporal Gyrus and the Temporal Pole (TP), as well as the general region of the Temporo-parieto-occipital junction (TPO). The interactions of the dorsal and ventral routes at the level of the prefrontal cortex is not shown here but is of course necessary for the visual system to guide action plans and fully interact with the control of such plans (through the cognitive control network).

in complex interactions within and between speakers (Garrod and Pickering, 2004; Menenti et al., 2012).

The information contained in the world knowledge system (heavy semantics) is thought to be distributed over many different brain regions. Results on embodied semantics emphasize the role that motor and perceptual system could play in the language comprehension process. Going back to ch. 5, fig. 5.1, world knowledge is thought to encompass motor and perceptual schemas that are used by the visual system to generate top-down expectations. Such schemas are neurally implemented by modal systems (perceptual or motor) although schema theory insists on the fact that perceptual and motor schemas cooperate with each other to guide action and therefore participate in brain functions that involve networks of brain regions. However, we do not restrict world knowledge to these perceptuo-motor schemas (see below, sec. 8.4 for an in depth discussion of the question of embodiment in relation to the brains' capacity to generate meaning).

Using a constructional framework adds the role that grammatical meaning (light semantics knowledge), learned through linguistic interactions and that span multiple time scales (from the knowledge that is relevant only during a given linguistic interaction and forgotten quickly afterwards to the knowledge that becomes part of our long term memory). Whether such knowledge can be reduced or not to modal perceptuo-motor schemas is at the heart of the debate over the embodiment of semantics (Mahon and Caramazza, 2008; Pulvermüller, 2005). But if a purely amodal semantic system has been shown to lead to empirical and theoretical difficulties, the opposite solution consisting of a purely embodied/modal knowledge system tends to underestimate the challenge that abstraction poses and falls short of explaining the variety of meanings that can be produced and understood beyond those carrying motor or perceptual contents (Arbib, 2008; Dove, 2010) (see below sec. 8.4 for a full discussion).

Frames are implemented in SALVIA as a core part of the structuring aspect of world knowledge. They represent our knowledge of how events unfolds, who or what is expected to be involved, and the roles that various participants play (such knowledge is also sometimes refer to as schemas in the psychological literature but we chose to use the frame denomination introduced by Fillmore to avoid confusions (Fillmore, 1976)). Such event knowledge bundles together multiple actions, objects and their interactions that can each link to perceptuo-motor or linguistic-semantic knowledge and has been shown to rely more on medial cortical regions and in particular on the posterior cingulate gyrus and precuneus that work in conjunction with the hippocampal memory system to encode and retrieve these event memories. The hippocampus was initially thought of as outside the language system since the seminal work on the patient H.M. who, after a bilateral partial removal of the hippocampus, was thought to suffer from a pure memory deficit. However, recent work on H.M. have emphasized that upon closer examinations, the hippocampal system could play a key role in language comprehension and production beyond information retrieval and could be involved in processing discourse level co-reference (MacKay et al., 2007; Mackay et al., 1998; MacKay et al., 1998; Skotko et al., 2005).

Finally, anchoring the SemRep as a model of the language-vision interface in the brain will require expanding the classic notion a language system confined to the left perisylvian regions to encompass its connections with perceptuo-motor regions known to be involve in tracking object positions and relations (in particular superior parietal lobe and parahippocampal regions (Vann et al., 2009)), in orienting attention (including both the bottom-up processing of salient features and top-down cognitive saliency biases (Baluch and Itti, 2011), and in controlling the oculomotor saccade system that generates the fixations patterns (Dominey and Arbib, 1992).

Fig. 8.4 summarizes some of the key issues encountered in linking SALVIA to a brain anchored model. The dorsal-ventral distinction appears in both the language and the visual systems, and in each takes a fairly similar functional flavor, v.i.z a separation between form based operations (tied to motor operations in the case of the visual system) as opposed to meaning based operations (tied to the recognition of object's identity and the nature of scenes in the case of vision). A general hypothesis would therefore be that the language system emerged atop the system for sensory-motor integration. This would give the study of the interaction of vision and language a new meaning as it would make contact with the analysis of the evolutionary origin of language. It is however necessary to go beyond the dorsal-ventral distinction. An important question to settle is the relations that exists between the role and gradient displayed by the prefrontal cortex as part of cognitive control systems and the distinctions that exist along the Inferior Frontal Gyrus (BA47, BA45, BA44, and BA6). Novick et al. (2005) as proposed that a direct relation exists between cognitive control and the role of Broca's area (see also Novick et al., 2010). It has also been suggested that (part) of the prefrontal part of the language system could be involved in discourse level operations rather than

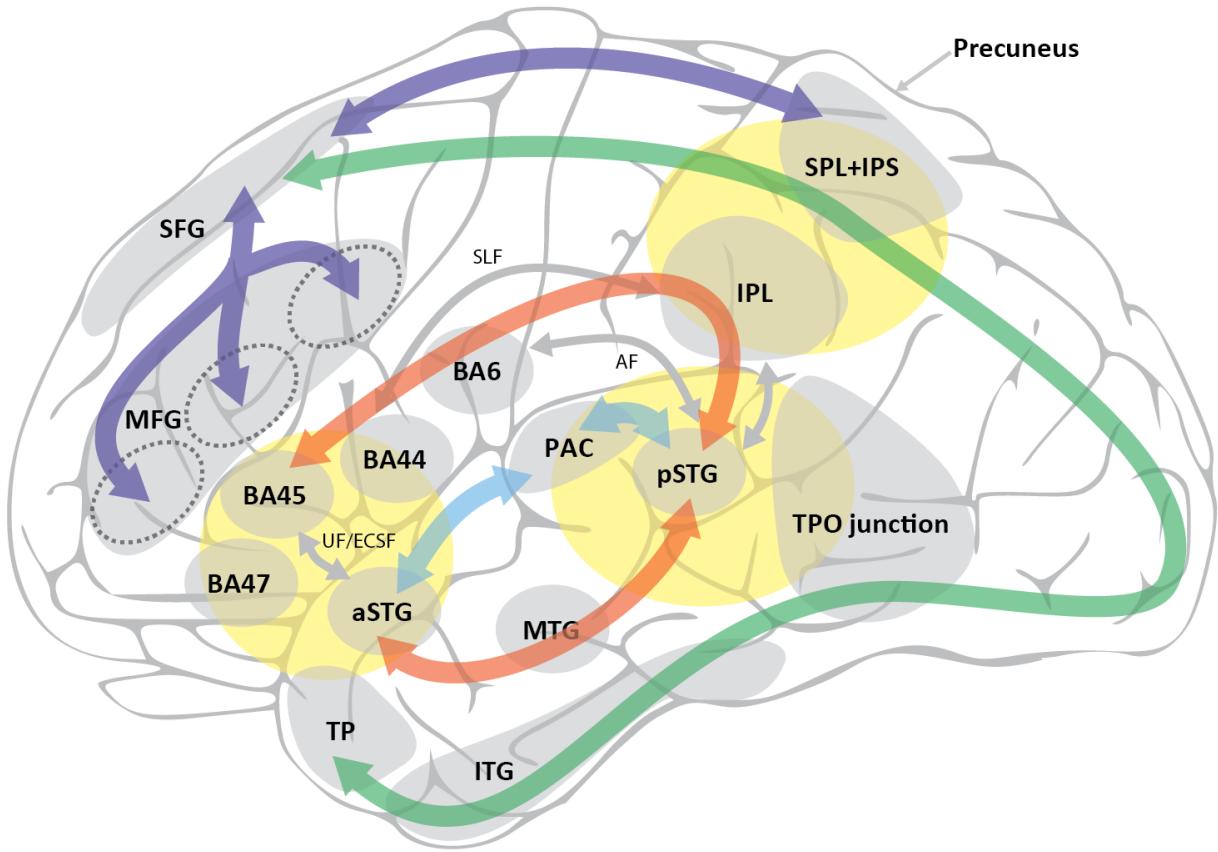


Figure 8.4: From SALVIA to brain anchored model: a first step. General view of the language (fig. 8.2 system embedded within the vision and cognitive control systems (cf. fig. 8.3). (See text for details)

in grammatical processes per se (a distinction that is somewhat blurred in the case of a constructional approach) (Bornkessel-Schlesewsky and Schlesewsky, 2013). In addition to the role of cognitive control, the more general role that reasoning can play around language processing (although maybe not directly as part of language processing) has been emphasized as involving the integration within the prefrontal language region of information stemming from dorsal systems (composed of the Superior Frontal Gyrus (SFG), the Superior Parietal Lobule (SPL) among other) more generally involved in non-linguistic high level inferences (Fedorenko et al., 2012; Fedorenko and Thompson-Schill). Finally, also going beyond the dorsal-ventral distinction, the role of multi-modal region such as the Inferior Parietal Lobule (IPL) could possibly play a crucial role in anchoring the semantic representations in visuo-spatial representations that can exist within different referentials, in particular if IPL is taken in relation to the sub-cortical systems it connects to (see Vann et al., 2009). What in SALVIA was therefore taken as a semantic representation (SemRep) held within a single cognitive-level working memory, future work should re-analyze its nature and how its structure and the processes it is involved in can be reworked to involve the prefrontal regions involved in both building the structure and linking it to both world knowledge and a hierarchy of goals, as well as the temporo-parietal regions that seem better fitted to link the SemRep to visuo-spatial representations, required both to keep track of the referents but also to then link to the attentional saccadic system.

The following three sections will discuss in turn three challenges that face any attempt to anchor computational models in brain data. First the possibility to use quantitative Synthetic Brain Imaging methods to quantitatively link a model's computation to ERP data is discussed. Then, the specific case of the role of Broca's area is discussed, highlighting the difficulties that still exist in understanding this key language brain area. Finally, the problem of meaning is addressed: as a cognitive model SALVIA insists on the interactions between form and meaning, but “meaning” remains an illusive and often ill posed concept in neurolinguistics.

This last section poses the question of semantics as the frontier beyond which much of the interesting things happen but remain largely out of reach.

## 8.2 Quantitatively Linking Phenomenological Models and Computational Models: Toward Synthetic Brain Imaging Approach to Neuro-Computational Modeling of Language Processes

Our long-term goal is to generate models with hypothesized circuitry localized to specific brain regions but competing and cooperating to yield overall behavior, making explicit how prior processing creates states that affect computing in the current epoch. The task is immensely simplified if it can be argued that the human circuitry for the function is a variation on circuitry in the monkey brain for which neurophysiological data are available. This was the case in studies of reach to grasp behavior (Grafton et al., 1998) (Arbib et al 2003) and basic forms of working memory and auditory object processing (Deco et al 2004, Husain et al 2004, Tagamets & Horwitz 1998, Tagamets & Horwitz 2000). In these cases, Synthetic Brain Imaging was applied to predict synaptic activity in circuitry localized in different brain regions as the basis for computing predictions testable by PET or fMRI imaging. But what of ERP data? ERP represents a unique source of brain data for neurolinguistics, well suited to study the fast, time-dependent language process (for which there is no possibility to take a detour through other animal models that would allow for neurophysiology experiments). As indicated above, the strategy is to use computational models to predict the time course  $d_A(t)$  of current equivalent dipoles for relevant areas of the brain, and then use forward modeling to derive predictions of ERP measurements. Unlike previous work on Synthetic Brain Imaging, model neurons must now be linked to cortical geometry so that dipole orientation can be defined in an anatomically sound fashion. Neural network models of language processing seldom link circuits to brain regions. SALVIA, equally, remains a cognitive level model with forays into the neural levels through the tackling of aphasias, but lacking mappings between processes and brain systems. Just as most schema models, SALVIA still lacks strong links to neurobiological data.

### 8.2.1 Testing a Neural Network Model Against ERP Data Using a Forward Approach: Modeling Requirements

The goal of computational neurolinguistics is to build models that not only simulate a linguistic task or behavior (such as decomposing the sound wave of a word into phonemes, assigning the proper syntactic structure to a sequence such words, describing a visual scene, etc.), but also replicates underlying brain processes. Our discussion of the forward model part of Synthetic ERP (Phase 2) showed that ERPs imposes important requirements on neural network modeling. Table 8.5 offers a preliminary presentation of these requirements for various features of neural network models. Such requirements will be at the heart of the future work on Phase 1 of Synthetic ERP.

#### The Minimal Neuron Model for Combining Synthetic Brain Imaging and Synthetic ERP

For Synthetic Brain Imaging, we require the time course of synaptic activity for all neurons in each simulated brain region. This can then be passed through the hemodynamic function and averaged appropriately to yield predictions for PET or fMRI. However, although the model may contain many neurons in both cortical and non-cortical regions, the neurons most relevant to the ERP signal are cortical pyramidal cells and the most relevant aspect of their activity is the PSPs in their apical dendrites. Moreover, the contribution of these dendrites to the dipoles integrated into  $d_A(t)$  for a specified region A depend crucially on their orientation. Thus, whereas the modeling of neurons for Synthetic Brain Imaging can ignore the spatial location of its constituent neurons, modeling of the cortical pyramidal cells for Synthetic ERP must represent not only the location but also the orientation of the neurons. We stress that in general the model will include other cell types from cerebral cortex as well as circuitry in subcortical regions but the activity of these cells is not included in Phase 2 of Synthetic ERP. In our discussion of forward modeling, we introduced a head model which represents the surface of cerebral cortex by a mesh (see sec. 7.5, details in Appendix F.3). Hence, in a computational model of cortical processing to be used with Synthetic ERP, the location of all pyramidal

Model Features	Constraints	Requirements
<b>Neuron model</b>	<ul style="list-style-type: none"> <li>- Current dipole's amplitude is linked to synaptic activity</li> <li>- Neural activity should be simulated with a ms precision</li> </ul>	<ul style="list-style-type: none"> <li>- Allow the quantification of synaptic activity</li> <li>- Account for the impact of synaptic activity characteristic times on processing times.</li> </ul>
<b>Neuron types</b>	Only the synaptic activity of pyramidal cells contribute to EEG signal	For Phase 1, the model must include all cell types required to provide circuitry that performs the stated tasks and can be tested against single-cell data (where applicable). For the current version of Phase 2, anatomical localization and orientation must be defined for the cortical pyramidal cells included in the model.
<b>Brain area model</b>	The cortical geometry is a key parameter of ERP modeling	Map each neural layer representing a given brain region onto a 3D mesh model of the cortical surface geometry using an existing surface atlas.. Neurons in a layer should carry hypotheses on their cell type and connectivity but also on their 3D location on the cortical surface.
<b>Overall network architecture</b>	Signal conduction times between brain regions need to be accounted for.	<ul style="list-style-type: none"> <li>- Incorporate into the large-scale connectivity between neural layers the known white matter connectivity between the brain regions they represent.</li> <li>- Account for the action potential propagation delays in the white matter tracks.</li> </ul>

Figure 8.5: Preliminary requirements imposed on computational neural networks by constraints associated with the use of an EEG forward model. For each feature of a neural network model outlined in the leftmost column, the central column stipulates the new constraints imposed on this feature by the use of the Synthetic ERP phase 2 forward model. The rightmost column converts these constraints into requirements that the feature needs to meet in order to be used in Synthetic ERP. Although we have exemplified Synthetic ERP with examples from computational neurolinguistics, this table offers guidelines of more general applicability. In particular, they should help orient neural network modeling for both systems and cognitive neuroscience towards model types that facilitate contact with ERP data.

cells must be specified either (a) as being below a specific face of the mesh or (b) below some larger “slab” obtained by aggregating a number of contiguous faces and assigning them an averaged orientation. Each cell is then oriented orthogonally to the face or slab to which it is assigned, and we must then employ a neural model that yields the time course of PSPs of the apical dendrites. Given the resulting spatial structure and the fine temporal resolution required to compute the ERP, we need to include axonal propagation time in the model, and for this reason it would seem that spiking models would serve better than rate models. The simulated PSPs of apical dendrites would provide the 3D distribution of dipoles, fixed in orientation but time-varying in amplitude, to drive the forward model. The result would be the Synthetic ERP for the phenomenon captured by the simulated network.

### Cortical Geometry and Surface Atlases

Synthetic ERP is not alone in stressing the importance of cortical geometry in defining the position and orientation of EEG sources (see Table 7.2). However, our approach to the forward problem both within the framework of neurolinguistics and as a tool to move from conceptual to computational models lead us to re-position this issue at the center of the forward problem. The analysis of the current dipole model and of the head model (see sec. 7.5, details in Appendix F.2), as well as the problems raised by the mapping of the brain regions defined by the 2002 model onto a single surface atlas (see sec. 7.6.1) show the issue of cortical geometry to be much richer than that of merely constraining dipole orientation as normal to the cortical surface. In order to quantitatively account within a computational framework for the impact of cortical folds on the EEG signals, we suggest the following questions be addressed within models of cerebral cortex:

- (1) How to quantitatively constrain the electric sources orientation and position by the geometry of the cortical surface?
- (2) What cortical surface should be used? Individual cortical surfaces or population-average templates?
- (3) What surface atlas should be used to parcellate the cortical surface and ensure a standard mapping between the brains of various individuals?

Question (1) is commonly addressed in the forward modeling literature through the use of 3D meshes representing the cortical surface. A dipole can then be constrained in orientation by the normal vector of a given face of the mesh. Its impact on neural network modeling has already been detailed (see sec. 7.7.1).

Question (2) raises the issue of the type of cortical surface used. The lack of inter-individual correspondence in cortical folding makes any attempt to define population-average templates difficult. The well-documented case of the duplication of Heschl’s gyrus in some individual (two gyri instead of one) provides a good example (Leonard et al., 1998). Heschl’s gyrus is part of the primary auditory cortex which, according to the 2002 model, performs acoustic analysis and phoneme identification, linked to the N100 ERP component. How should these variations in Heschl’s gyrus’ morphology be incorporated in the framework of computational modeling? How do they impact the forward modeling of N100? In part, the answer may well depend on the precision with which the ERP data are presented. If the data average over a number of subjects without separation with respect to, at least, gross differences in gyration, then it may be appropriate to use the simpler anatomy as the basis for Synthetic ERP modeling.

Specifically, our simulations are based on a cortical mesh generated from high resolution MRI scans for a single individual (MNI Colin 27, see Appendix F.3). However, the fact that ERPs are usually reported in the literature as population averages begs the question of whether the use of a population-average template would not be more appropriate. Most of the existing EEG source localization software (e.g. Oostenveld et al., 2011; Tadel et al., 2011) address this issue by offering the possibility to either (a) use a cortical mesh extracted from an anatomical MRI scan of the individual from whom the EEG signal has been recorded or (b) use a standard brain mesh (such as MNI Colin 27) that can be warped to match sensors locations or be used with standard sensor locations (which is similar to our choice). However, (a) seems unsatisfactory for computational neurolinguistics since most ERP data reported in the literature do not come with associated cortical meshes. As to (b), the possibility to use a population-template brain mesh is undermined by fact that averaging surfaces from different brains remains a challenge for which no standard solution has yet emerged (Auzias et al., 2011; Dale et al., 1999; Fischl et al., 1999; Lyttelton et al., 2007; Van Essen, 2005). Moreover, such average seems more likely to blur gyration rather than providing a useful standard. Thus,

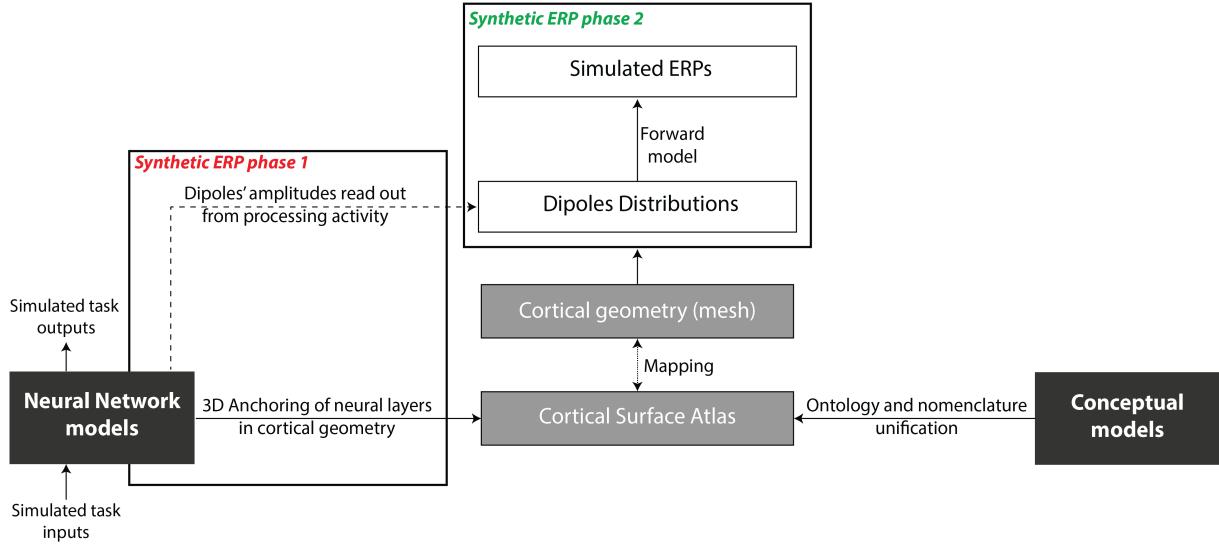


Figure 8.6: Building bridges between computational and conceptual models and quantitatively testing the models against ERP data. Neurolinguistics is rich in conceptual models (right black box) that provide detailed qualitative analysis of brain imaging results, and mostly ERP. The goal is for computational model to use the many insights in both architecture and processing given by the conceptual models while turning those into fully specified models that provide a causal and mechanistic explanation of a behavior (left black box: model should take simulated tasks inputs and generate simulated task outputs.) Conceptual models are however usually build using mixed ontology which makes the comparisons between conceptual model and the integration of their insights into neurocomputational models difficult. It also makes it impossible to quantitatively link computational models to ERP data. The first step should therefore be to project both computational and conceptual models onto a shared cortical surface atlas that can then be mapped onto cortical geometry. From there it is possible to use Synthetic ERP tools to quantitatively assess whether or not a neurocomputational model generates not only the proper behavior but also the proper ERP signal. The work noted as Phase 2 is the one that has been presented above. Phase 1 remains to be fully implemented.

for now, it seems a reasonable strategy to use MNI Colin 27 as the basis for localization and orientation of cortical areas for a neural network model.

Question (3) emphasizes the issues raised by the diversity of brain region atlases used by modelers, the difficulty to quantitatively define the relations between them, and the difficulty to link some of them to cortical geometry. The 2002 model offered a good example of this, using of a mixed ontology including parcellations of the cortical surface based on neural cytoarchitecture (Brodmann areas), cortical folds (gyri and sulci), function (e.g. auditory cortex), and lesion data (e.g. Broca's area)(see Table 7.8). Without quantitative treatment of these various nomenclatures, all pose problems for the simulation of EEG signals. Individual variation in cortical folds was already mentioned in relation to point (2), but in addition the idiosyncratic variations of localization within the cortical folds of Brodmann areas have only recently started to be quantitatively analyzed (Fischl et al., 2008). In particular, linking computational neurolinguistic models to ERP data will require a deeper analysis of recent findings on the individual variations in Broca's area anatomy (Keller et al., 2009).

Finally, we note that the incorporation of empirical data from non-human primate studies into neurolinguistic modeling, such as single cell recordings or white matter connectivity, raises between-species homology challenges (Deacon, 2004; Arbib and Bota, 2004). Neural networks have been extensively used to model the brain functions of non-human primates. Their use in neurolinguistics should be thought of not only in terms of their computational power, but also as a way to make contact with computational models of results from monkey neurophysiology. This approach will incorporate evolutionary hypotheses into neurolinguistic computational models and test them against human neuroimaging data.

### 8.2.2 Models of Language Processing

The vast majority of neural network models associated with language processing are unconstrained by brain imaging or ERP data. A number of interesting models have, for example, used Simple Recurrent Networks which can be trained, e.g., to learn sequences in a way which replicates application of simple constructions in forming sentences. Among the deepest studies of this kind is one that addresses a range of psychological and developmental data (Chang 2002, Chang et al 2006) but it does not address neurophysiological or neurological data. Another model only addresses lexical access but does offer some insight into aphasia (Dell et al 1997). The GODIVA model of sequencing phonemes to form words (Bohland et al., 2010) does address a range of neurophysiological data but is clearly at a lower level of language processing than those that have engaged us in this article. Dominey, extending earlier research modeling linkages between cortex, basal ganglia and brain stem in the control and learning of eye movements (Dominey and Arbib, 1992; Dominey et al., 1995), provides a neurally plausible model of the interaction of syntax and semantics in the parsing of simple sentences (Dominey et al., 2006, 2009). The model employs recognition of the sequence of function words in a sentence to provide access to the syntactic construction that assigns the content words to their semantic roles<sup>1</sup>. Previous attempts at relating artificial neural network activity to language-related ERP include models of the dynamics of neural masses (David et al., 2011; Yvert et al., 2012) but do not simulate the information processing whereby the brain performs a given linguistic task (recall Table 7.2). Another approach employs simple Hebbian networks but does not analyze the impact and role of forward models on ERP signals simulation (Garagnani et al., 2008).

Of course, a major obstacle to neurophysiologically realistic models of language processing is that only humans possess language in the sense of an open-ended lexicon and a grammar that allows the flexible production and comprehension of utterances that convey novel meanings in diverse domains of discourse. As noted earlier, there are two strategies for development of fine-scale models of language processing that follow from this:

- 1) One is to employ evolutionary hypotheses to create models that combine (i) modules whose detailed neural circuitry can be related to that of non-humans executing a *similar* subfunction and (ii) modules executing *human-specific* subfunctions for which the circuitry can be structured on the basis of the evolutionary hypotheses (Aboitiz, 2012) (see Aboitiz et al 2010, Arbib 2006, Arbib 2010, Arbib 2012 for examples of such hypotheses).

<sup>1</sup>The model was then extended to handle much more complex sentences comprehension and production, using reservoir computing approaches that keep intact the basic functional architecture but are less directly linked to brain systems (Hinaut and Dominey, 2013; Hinaut et al., 2015).

- 2) The other is to abandon the use of simulated neurons as the unit of processing and instead use networks of interacting schema instances to work at a scale slightly coarser than that of neurons, but far more detailed than that of brain regions. Schema theory has been employed successfully over the years in modeling visually guided behaviors in frogs, rats, monkeys and humans where we make crucial hypotheses as to the localization of schemas in different parts of the brain. It was introduced to neurolinguistics by Arbib and Caplan (1979); and we further developed links between schema theory and language (aphasia, acquisition, production) in Arbib et al. (1987). Here, we have extended schema theory to define Template Construction Grammar and the SALVIA architecture as a mechanism linking vision and language in the description of visual scenes, following in the footsteps of the work started by Arbib and Lee (2008, 2007), but as we saw this model includes no hypotheses on cerebral localization of schemas (though we have taken a small step in this direction by discussing some aspects of aphasia). Crucially, schema-theoretic models in their current form cannot support Synthetic ERP for at least two reasons:
  - a) Spatial: To employ Synthetic ERP, various schema subnetworks would have to be assigned to different faces or slabs of a cortical mesh or linked to subcortical regions;
  - b) Temporal: To compute  $d_A(t)$ , “schema time” would have to be mapped onto “neural time,” and competition and cooperation between cortical schema instances would have to be mapped to patterns of activation of apical dendrite PSPs.

Solving (a) lies within the remit we have established above for neural network modeling, whereas (b) poses new challenges, but challenges that must be met since schema theory serves as a bridge between the language of psychology and the language of neuroscience, and this “translation” needs to be extended to link ERPs as a tool for psycholinguistic observation to a deeper understanding of the patterns of information processing distributed across the brain.

Nonetheless, some links and challenges for the spatial problem can be made explicit. A phoneme recognition schema could be relatively well localized around Heschl’s gyrus. However, semantic memory can be represented as a widely distributed schema network, making explicit the problem of mapping onto the cortical geometry required for ERP modeling.

Our earlier work on Synthetic Brain Imaging was motivated by the need to link neural models (inspired in part by data from primate neurophysiology) to the results of human brain imaging. The work presented in ch. 7 addresses the challenge of linking such models to ERP data (and, although we have not discussed it here, the data of magnetoencephalography). We saw that whereas Dynamic Causal Modeling (DCM) addresses the issue “What aggregated measures of underlying neural activity could cause the observed ERP recordings?” our new Synthetic ERP methodology is doubly causal, based on two phases:

- Phase 1 addresses the question “What patterns of interaction in neural circuitry could cause the observed behavior (and, where available, explain single-cell recordings)?” whereas
- Phase 2 addresses the question “Could the aggregate activity of neurons in the circuitry so modeled cause the observed ERP recordings?”

We have seen that Phase 2 is well-defined and has much in common with DCM, but enforces the assumption that ERPs be calculated using a forward model on the basis of dipoles whose orientation is provided by orientation of the corresponding region of cortex, rather than having arbitrary orientations based on one of the many possible solutions to the inverse problem posed by observed ERP data.

However, our work on Phase 1 is at an earlier stage of development. The key point is that, in addition to the requirements of neural network models used for Synthetic Brain Imaging that they can be tested against single-cell data (where available) as well as behavioral data, the neural networks used for Synthetic ERP must include neuroanatomically realistic placement and orientation of cortical pyramidal neurons. However, work on Phase 1 cannot (in general) succeed if only cortical regions are modeled subcortical structures such as basal ganglia and thalamus may play a crucial role.

All this poses exciting challenges for future work in neural networks that is to more fully contribute to computational modeling for systems and cognitive neuroscience. In particular, future work in neurolinguistics will depend on both new approaches to the structuring of empirical data and on the development of novel computational models of language processing.

Finally, we noted the power of schema network models in explaining behaviors for which the pool of relevant neuron-level data is impoverished, but then raised the daunting question of how to link activity of schema networks to anatomically localized and oriented dipoles, in both space and time.

## 8.3 The Elusive Role of Broca's Area in Comprehension and Production

"Tan"

Tan

It is hard to overestimate the influence that the findings of Paul Broca have had on the course of cognitive neuroscience. Simply stated, his characterization of the specifically linguistic disorder displayed by his patient Leborgne (or "Tan", after the only syllable he could produce following his brain lesion), and its association with a lesion to what appeared to be a relatively well-defined brain region within the left inferior frontal cortex, set the stage for more than a century (and counting) of neurolinguistic work seeking to associate brain regions with aspects of language processing. Broca had found a "motor center for speech" in what is now called Broca's area. Wernicke, Lichtheim, Geschwind among others would add to this picture other brain areas associated with other types of linguistic faculties summarized in the classic Lichtheim model. At all times there were researchers that suggested that such localist views might be inappropriate (we can mention Freud, Goldstein, Luria, Caplan & Arbib etc). But modular views of the brain championed by cognitive scientists such as Fodor and the rise of functional imaging (PET, fMRI) with its use of contrast methods both strongly reinforced this paradigm. Whether we adhere or not to this method of enquiry, it remains that it did focus the attention of neurolinguists on some key brain regions, and in particular on Broca's area, resulting in a large body of data about its possible function that offers an interesting basis to analyze what it is that any neurocomputational theory of language processing needs to tackle. I want here to review empirical data that shed some light onto the role that Broca's area might play in spoken language comprehension and production. I will organize these data points according the three landmark functions that have been put forward for Broca's area: syntax processing, verbal working memory, and cognitive control. In addition, I will mention the recent surge in interest in connectivity analyses, thought of as a pathway to function from structure.

I will eschew here the question of whether or not it is epistemologically justified to look for a function associated with a given brain region since this would lead me well beyond the boundaries of this essay. Suffices to say that since people have tried to answer this question in the case of Broca's area, it is worth looking at the data they have gathered while remaining open-minded as to how to interpret it.

### 8.3.1 Definition

Although I will not dive too deep into the many issues related to the definition of a brain region such as Broca's area, it is worth saying a word about this problem. Although historically lesion-based, the current canonical neurolinguistic definition identifies Broca's area with the posterior two-third of the left inferior frontal gyrus (lIFG) composed of pars triangularis (PTr) and pars opercularis (PO). This characterization in terms of cortical geometry is usually taken as equivalent to definitions based on cytoarchitecture which defines Broca's area as encompassing BA44 and BA45. However recent work has shown that the boundaries of Brodmann areas only coarsely correspond to the cortical folding geometry (Keller et al., 2009). In addition connectivity-approaches to brain region definition generate maps that do not directly overlap with either of these two preceding ones (Anwander et al., 2007; Catani et al., 2005; Friederici, 2009; Glasser and Rilling, 2008; Saur et al., 2008). Finally recent work looking not only at cytoarchitecture but also at transmitter receptor distribution and local connectivity has gone beyond Brodmann coarse cellular level analysis and revealed important microstructural gradation within Broca's area (ie streams along which cytoarchitectural

features change from one to another (Sanides, 1964)) (Amunts and Zilles, 2012). Understanding how these different mappings relate to one another is most likely one of the key that will allow neurolinguistics to develop better theories, in particular when considered within an evolutionary neurohomology perspective (Arbib and Bota, 2003; Bota and Arbib, 2004). However, it is clear that for now high level resolution remains inaccessible to most of the neurolinguistic experimental results, including those drawn from neuropsychology and functional neuroimaging. For this reason I will settle here with the canonical definition of Broca's area as encompassing left PTr and PO (or equivalently BA44 and BA45), although occasionally BA6 and BA47 as well as the frontal operculum (FO) will be brought into the picture.

### 8.3.2 Syntax

Although the initial work of Broca defined lIFG as a motor speech center, for more than 50 years it is with syntax processing that this region became associated to. This was the result of (1) the general focus on syntax as the core property of language following the birth of generative grammar; (2) the empirical findings that patients suffering with Broca's aphasia (expressive, or non-fluent aphasia) produce asyntactic speech (Gleason et al., 1975; Goodglass and Berko, 1960; Goodglass, 1968, 1976; Kean, 1977); (3) the empirical observation that asyntactic production in Broca's aphasics was accompanied with an difficulty to make use of syntactic cues during comprehension (Berndt et al., 1996; Caramazza and Zurif, 1976; Grodzinsky and Marek, 1988; Sherman and Schweickert, 1989). The parallel between asyntactic production and comprehension led to the characterization by cognitive neuropsychologists of the clinical impairments of Broca's aphasics as "agrammatic" (agrammatic aphasia), i.e. the lesion had specifically impaired the capacity to process grammar be it for production or comprehension. Importantly lesion to Broca's area was thought to be directly associated with Broca's aphasia (defined as a symptom complex), therefore Broca's area was implicitly associated with the role of processing syntactic information (as at least a necessary player in a syntactic processing system) for both production and comprehension.

Two main neurocognitive theories emerged and still subsist that support and refine the idea that Broca's area is dedicated to the specifically linguistic task of processing syntax.

#### Grodzinsky

put forward the idea that Broca's area is specifically involved in processing syntactic movements as defined within generative theories of grammar (at least the versions that still consider that there are such things as movements) (Grodzinsky and Santi, 2008; Grodzinsky, 2000). Empirical support for this theory comes directly from the difference in comprehension performances of Broca's aphasics on sentences that contain or not syntactic movements. Neuroimaging studies have also shown increased activation of Broca's area for sentences that involve long compared to short distance movement ("Diego hates the gift that the girl with the yellow clipboard bought" vs. "Diego hates the gift that the girl bought") (Santi and Grodzinsky, 2007a,b). More recently, using an fMRI adaptation paradigm Santi and Grodzinsky (2010) have further reported that BA45 selectively shows adaptation for movement types (object movement "The boy that the girl chased [...] vs. subject movement "The boy that [...] chased the girl") while BA44 shows adaptation for both movement types and embedding positions (center embedded vs. right branching). Grodzinsky's hypothesis however suffers from multiple flaws. The main point of criticism, shared with most of the classic neurocognitive analysis that attribute a syntactic processing role to Broca's area, is that it is now clear that Broca's aphasia as a clinical symptom complex does not directly map onto lesion to Broca's area. It has been known for quite a while that a subject can display the clinical symptoms of expressive aphasia without having lesions to the lIFG (Dick et al., 2001) and reversely that a lesion to Broca's area does not necessarily cause Broca's aphasia (Hamilton and Martin, 2005; Mohr et al., 1978; Mohr, 1976; Novick et al., 2009; Thompson-Schill et al., 2002). In addition, it is well known that sentences that involve movements are more taxing on the working memory than those that do not, introducing a confound in terms of working memory load. The fMRI results that are brought in support of this theory are also somewhat contradicted by the fact that Broca's area shows a movement distance effect only for semantically unconstrained sentences ("The lawyer that the banker irritated filed a hefty lawsuit" where both lawyer and banker are as likely to be the subject of the verb "irritate") but not for semantically constrained ones ("the thief that the policeman arrested was known to carry a knife" where the thief is much more likely to be the patient of "arrested")

than the policeman) (Caplan et al., 2008).

### Friederici

Friederici proposed that Broca's area supports two separated grammatical processing roles. She hypothesized that FO in relation to the anterior temporal lobe (aTL) is involved into local phrase structure building (e.g. building of a NP) while PO in relation to the posterior superior temporal sulcus (pSTS) supports "hierarchical structure processing" (Friederici, 2009, 2011). In an fMRI experiment involving learning an artificial grammar defined using consonant vowels sequences, and that included hierarchical dependencies - AnBn- or no hierarchical dependency - (AB)n, Bahlmann et al. showed that grammatical violations in both condition elicited an increased activation in FO but only violations of the hierarchical grammar elicited increased activation in PO (Bahlmann et al., 2008). This result was replicated using natural language sentences and PO was found to be increasingly active when global hierarchical structure had to be processed (center embedded relatives) while controlling for working memory load effects (Makuuchi et al., 2009). Previous studies had also reported increased BOLD activation in FO for sentences containing syntactic violations compared to correct sentences (Friederici et al., 2003). Friederici's hypothesis has been criticized on different grounds. First of all, the use of artificial grammars suggests that the results point at best to a non-language specific role of PO in hierarchical structure building. Second Makuuchi et al. (2009) use a quite idiosyncratic definition of grammatical hierarchical structure (which seems restricted mainly to center-embedding). More importantly, there is no clear account given of what would distinguish local from more global hierarchical structure processing nor is it specified how the brain would perform such distinction (Bornkessel-Schlesewsky and Schlesewsky, 2013).

Overall, the endeavors that have sought to directly define the role of Broca's area as related to some linguistic form of language specific syntactic processing have suffered from the collapse of the support they once had from neuropsychological evidence. These do not allow direct mapping between aphasic symptoms (including agrammatism) and lesion localization. Neuroimaging studies have not so far provided any significant result that would back up this idea. In addition, focusing on syntax forces to ignore the piling evidence that link Broca's area to non-linguistic functions (e.g. action planning, cognitive control, etc.).

### 8.3.3 Verbal Working Memory and Articulatory Rehearsal

Broca's area has been shown to be consistently involved in various forms of verbal working memory. PET studies have shown increased activation for Broca's area during verbal working memory task in which subject have to memorize words list (Awh et al., 1996; Smith et al., 1996; Smith and Jonides, 1997). fMRI studies have replicated these findings (Buchsbaum et al., 2001, 2005). From this point of view, Broca's area is thought to function in concert with pSTS to support phonological rehearsal as well as word recognition through sensory-motor mapping (with Broca's area hosting articulatory motor codes used both for speech production and speech perception) (Hickok and Poeppel, 2004). Some connectivity analyses (DTI) however have suggested that it is BA6 that could be more directly involved in sensory-motor mapping and not Broca's area, since only BA6 is directly connected to pSTS (through the arcuate fasciculus (AF)) (Glasser and Rilling, 2008) while BA44 could be connected to pSTS only through the inferior parietal lobule (IPL) (Catani et al., 2005). This is significant since most the theories of verbal working memory suggest some system of direct reverberation between motor and sensory areas, which would imply a direct connection between them. Therefore it is possible that upon closer look and given the caveat regarding the definition of Broca's area, a more premotor region such as BA6 would be involved in articulatory rehearsal. The question of the involvement of Broca's area in verbal working memory has often been pointed to as a possible explanation for its increased activation during complex syntactic task (or for the agrammatism that can result from its lesion). Producing or comprehending complex sentences could require a functioning verbal working memory as a buffer for the linguistic output to be uttered (in the case of production) or as a necessary resource for reanalysis and repair of misparsed utterances (in the case of comprehension). Caplan et al. have investigated this issue and found that activation in Broca's area increased with the syntactic complexity of the sentences even in the during articulatory suppression (ie while the subject had to keep in memory a sequence of words) (Caplan et al., 2000). More recently however Rogalsky et al. (2008) have repeated a similar experiment and have reported that the articulatory suppression eliminates the effect of

syntactic complexity on activation level in PO but not in PTr (while both show an increased activation effect without articulatory suppression). Moreover, the positive correlation between activation level and syntactic complexity in PTr became non-significant when subjects had to perform a simple finger-tapping task while they were listening to the sentences but the effect reappeared in PO. The authors therefore suggested that PO participates in comprehension only through its role in verbal rehearsal though a sensory-motor system that is vocal tract related (undisturbed by finger movement pattern production) and that the role of PTr in comprehension might not be language specific (since it shows interference with a simple finger-tapping task).

### 8.3.4 Cognitive Control

A more recent type of approach to the role of Broca's area in language processing seems to be born from joint the collapse of the syntax processing hypothesis and the concomitant rise of empirical evidence pointing to the general role of prefrontal cortex in cognitive control i.e. in the general process of shaping the type of task- and context- dependent responses required by most complex human behaviors (Badre, 2008; Badre and Wagner, 2007; Koechlin et al., 2003; Miller and Cohen, 2001). More specifically, cognitive control focuses on how the PFC allows for the fast remapping of stimulus-response associations based on context or goals, rapidly integrating various sources of information to "sculpt the response space" (Fletcher et al., 2000). Moving away from the language specific perspective on Broca's area, a new line of empirical work, spearheaded by Jared Novick (Novick et al., 2005), has emerged that aims at finding a reconciliation between the cognitive control and the neurolinguistic literature. One of the key hypothesis is that Broca's area is more specifically involved in conflict resolution processes both during typical cognitive control tasks (such as the Stroop task, or the flanker task in which subjects needs to respond to the direction of a central arrow ignoring the direction of the arrows that surrounds it) or during language processing. In the case of language, conflict resolution is thought to be required during comprehension to deal with polysemy (at the semantic or syntactic level), referent resolution, or garden-path effects and, during production, to resolve the competition between multiple semantically related options (with a strong focus on lexical access and the need to handle the competition between semantically related words).

Neuroimaging data suggest that the same brain regions in IIFG are active during syntactic and non-syntactic conflict resolution task. January et al. have reported that similar patterns of BOLD activation were found in BA44 and BA45 within individuals for both a Stroop task and ambiguous sentence comprehension (e.g. ambiguous prepositional phrase as in "The girl saw the man with the binocular") (January et al., 2009). Similarly, Ye et al. have compared BOLD activation patterns for the Stroop and flanker tasks, with those of a linguistic task involving a conflict between world knowledge and syntactic information (as in "The policeman kept the thief at the station" compared to "The thief kept the policeman at the station" which would be more compatible with the passive construction). They found a co-localization of the effects in IIFG (Ye and Zhou, 2009). Developmental research has also led researchers to see a parallel between the development of cognitive control capacities to override automatic responses to stimuli, the development of the prefrontal cortex, and the capacity of the child to override the initial wrong parse of a garden-path sentence. In a visual world paradigm experiment, upon hearing the sentence "Put the frog on the napkin in the box" while being visually presented with a set up that includes a frog on a napkin, an empty napkin, and a box, adults will initially fixate the empty napkin as a possible destination for the frog but upon hearing the final prepositional phrase "in the box" will revise their parse and start fixating the box and will eventually perform the correct action. Five-year old children on the other hand, will show the same initial pattern of fixation (they are led down the garden-path) but more than half the time do not revise the initial wrong parse and for example will place the frog first on a napkin and then in the box (so called Kindergarten-path effect!) (Trueswell et al., 1999). Such results seems quite remote from a direct analysis of the role of Broca's area in language comprehension but illustrates novel attempts at understanding the role of PFC that try to integrate many different types of data points (including developmental data). Neuropsychological case studies of patient with focal damage to IIFG have also been put forward to support the cognitive control role of Broca's area. Patient IG with a circumscribed lesion to IIFG did not show any sign of agrammatism or of Broca's aphasia, but showed a decline in cognitive control capacities (measured using proactive interference effect). Patient IG also displayed language production deficits compared to control in picture-naming task that involved naming objects with low name agreement (many equivalent lexical options competing) while no deficit was observed for high name agreement objects. He similarly displayed production deficits when

asked to produce many words associated with a superordinate categories (e.g. Animals, which provide few constraints and therefore generates a large set of lexical items competing for production) compared to a matched condition in which he had to produce words associated to a subordinate category (e.g. Farm animals). These results suggest that IG deficit resides in the incapacity to produce words when multiple lexical items are competing. However, under low-competition condition IG's production is normal (Novick et al., 2009). Similar results are reported about patient ANG, also suffering from a focal lesion to IIFG, who was shown to be selectively impaired when asked to complete sentences with multiple possible continuations (Robinson et al., 1998, 2005) (see also (Schnur et al., 2009) for a meta-analysis of the performances of 12 patients with circumscribed IIFG lesions during picture-naming task with semantic interference). Turning to comprehension, patient IG was tested on the “frog task” described above and made errors similar to those of five-year old children showing reduced capacities to revise parsing. Importantly though, IG made no error in the case when the sentence was not ambiguous (“Put the frog that’s on the napkin in box”) which clearly shows that his deficit is not in parsing per se but in disengaging from a wrong parse. Work on word-sense disambiguation also showed that patients with IIFG damage tend to do worse than control subjects in tasks that require fast lexical ambiguity resolution (Bedny et al., 2007).

These results are far from being sufficient to generate a clear picture of what Broca’s area does. In particular one of their main drawbacks is that they do not rest on a clear theory of what cognitive control is. This is particularly flagrant when compared to the extremely precise theoretical claims of Grodzinsky described above. However, this approach has the interesting characteristic of trying to unify under the umbrella of cognitive control the various empirical results about Broca’s area that have emerged from different neuroscientific fields (e.g. neurolinguistics but also motor control). In addition, this perspective can be seen as generally compatible with neurolinguistic theories that suggest a more general information unification role for Broca’s area (Bornkessel-Schlesewsky and Schlesewsky, 2013; Hagoort, 2005).

### 8.3.5 Multi-Stream Integration

A last interesting and quickly growing type of empirical data that sheds a light on the role of Broca’s area in language processing stems from white-matter and functional connectivity analyses. The past ten years have seen the quick accumulations of studies highlighting the fact that the language system, coarsely defined as the left perisylvian area, was in fact composed of multiple anatomical and functional pathways that run between pSTS and Broca’s areas. At least two different DTI studies demonstrated the existence of two ventral pathways (Anwander et al., 2007; Saur et al., 2008). Those two DTI studies suggest the existence of a single dorsal pathway but Catani et al. reported the existence of two different dorsal pathways with one connecting directly pSTS and PTr while the other connects indirectly pSTS and PO through IPL (Catani et al., 2005) while Glasser et al. also reported two dorsal pathways connecting STG to PO and MTG to PTr respectively (Glasser and Rilling, 2008). Although the specific functions of these ventral and dorsal pathways are unknown, there is a tendency to assign semantic processing to the ventral pathways and syntactic processing to the dorsal pathways (and possibly verbal working memory to one of the dorsal pathway) (Friederici, 2011). Such empirical advances in understanding the general connectivity pattern within the language system provides support to the idea of that language processing rests on distributed computation in a multi-stream architecture. Such a view favors an interpretation of the role of Broca’s area as a more general controller shaping the interactions of the various processing streams based on task demands or pragmatic goals during both comprehension and production.

### 8.3.6 Interaction Between Production and Comprehension in Broca’s Area

A review of the literature on the role of Broca’s area in language production and comprehension makes quite obvious that:

1. Studies of comprehension overwhelmingly dominate the literature. Studies of language production are usually reduced to studies of lexical access during naming protocols, or verbal working memory studies. Such a domination of comprehension studies stems in part from the great difficulty to design controlled experimental protocols to investigate language production (and from the fact that any motor act tends to contaminate functional imaging data).

2. This lack of studies of production is usually not considered a problem since Broca's area is often thought to support equally production and comprehension. In the case of Broca's area I would argue that this view stems from the same key ingredients that 50 years ago gave birth to the "Broca's area as syntax processing system" interpretation, i.e. from the influence of generative grammar that considers syntax to be independent of the modality of use (production or comprehension) and from the results on Broca's aphasics that had suggested that any deficit in syntax processing during production was paralleled by a deficit in syntax processing during comprehension. Both pushed for an interpretation of the core language system as rather amodal.

Although the "syntax center" interpretation of Broca's area has now fallen in disfavor, the idea that its function is similar in production and comprehension remains intact. The verbal working memory view and its related motor-theory of speech perception both involve Broca's area as a core articulatory motor center but both also highlight its reciprocal use during word production and speech perception. Similarly the control theory literature insists on the fact that the role of Broca's area in conflict resolution and response shaping is equally required for production and comprehension. Some recent work has tried to directly analyze what are the brain systems that play a similar role in production and comprehension. A first fMRI adaptation study has reported similar location of suppression effects in BA 44 and BA 21 after repetition of a similar syntactic structures either in production or in comprehension (Menenti et al., 2011). This suggests that production and comprehension overlap in those brain regions but does not touch upon the question of whether they actually share processes. A following study therefore used the same paradigm but this time found suppression effect in the same areas after cross-modal repetition (ie with interleaved comprehension and production of similar syntactic structures), adding a stronger support to the idea of shared syntactic processes between production and comprehension (Segaert et al., 2012). Such shared processing systems for production and comprehension have been hypothesized to play an important role to support linguistic alignment during conversation (Menenti et al., 2012; Pickering and Garrod, 2007).

This type of fMRI adaptation studies makes a first step towards better understanding the interactions between production and comprehension in Broca's area but so far this issue remains largely understudied.

## 8.4 The Neuroscience of Semantic Processing: A Frontier

*"Le sens trop précis rature  
Ta vague littérature."*

Mallarmé

Toute l'âme résumée, in *Poésie*

### 8.4.1 Overview of the Problem

Here I will present some important current neuroscientific theories of semantic processing and discuss some of the key conceptual axes along which they can be organized. It would be dangerous to start such an endeavor without saying first a word of about the meaning of the expression "semantic processing". "It needs to be at least briefly clarified since I would contend that this expression is at the center of an important misunderstanding in the (cognitive) neuroscience literature, especially when language is brought into the picture. Cognitive scientists tend to use "semantics" to describe conceptual or world knowledge, its storage in a specific type of memory system, and its use in various types of cognitive processes. For example seeing an apple would trigger the activation of an apple representation (the format of which is irrelevant here), that would correspond to the semantic (conceptual) knowledge about apples an individual has gathered through it repeated interactions with the world, knowledge that could undergo various type of semantic processes (supporting either object recognition, scene understanding, the generation of a verbal description, etc.). Said briefly, for much of cognitive science work, "semantic processes" is somewhat equivalent to

“conceptual processes”. However, from a linguistics perspective, semantics refer to something quite different. As Pylkkanen et al. (2009) put it, in linguistic theory “semantics’ refer to the composition operations that serve to construct the meaning of an expression and world knowledge’ to our non-linguistic knowledge about the world that, for example, determines whether an utterance describes a plausible situation or not”. During language comprehension and production, semantics here would correspond to a specific component of language, and “semantic processing” to semantic operations such as compositionality, complement and aspectual coercion, scope, quantifiers, negations etc. Conceptual knowledge does not appear in this picture. I will mostly focus on semantic processing as defined by cognitive scientists, but I will also discuss, when relevant, semantic processing as defined by linguists.

I propose to organize the multi-faceted problem in the following way. In the first section, I will review some of the main current theories of semantic processing based of the type of neural semantic architecture they hypothesize. As for most of other brain functions, a large part of the theoretical debate over the nature of semantic processing in cognitive neuroscience is tightly linked to a debate over how the semantic content is organized in the brain. In the remaining sections I will reduce the scope and focus on theories of semantic processing that are related to aspects of language processing. In a second section I will survey theories based on the nature of the semantic representations they hypothesize play a role in language comprehension. In a third section I will briefly point to some of the few theories that have addressed the question of semantic processing as defined from the linguist’s perspective, therefore providing a richer picture to the type of semantic operations that support language comprehension. Finally, I will discuss briefly some of the few current theories of semantic processing that are supported by a computational model.

### 8.4.2 Neural Organization of the Semantic System

At the core of most the theories about the nature of the semantic processes carried out by our nervous system lies the assumption that semantic function is deeply routed into the structure of the semantic system. Theories differ widely in terms of how distributed the semantic system is hypothesized to be, which has direct consequences in terms of what types of processes it supports.

#### Distributed-only

Distributed-only theories consider that semantic processes is not localized in any specific brain regions or sub-system but rather is largely distributed across the brain. Based on an early review of the neuroimaging literature, Thompson-Schill (2003) concluded indeed that “the search for the neuroanatomical locus of semantic memory has simultaneously led us nowhere and everywhere”, and put forward a position in which semantic knowledge was essentially stored and processed on large networks organized by attributes (e.g. color and form knowledge is assigned to ventral regions while object-manipulation knowledge is assigned to the premotor cortex) and possibly by categories. If she assigns nonperceptual conceptual knowledge to anterior regions of the temporal cortex, she defines it in term of verbal knowledge, insisting that no system is found to gather and integrate the information of those distributed networks (with prefrontal regions and in particular left IFG being involved only in general purpose selection or control, not restricted to semantic processing). Such picture of the semantic system supports a view of semantic processing in which meaning building result from the complex coordination of activity between large distributed networks, without any specific semantic control center. Early work by Pulvermüller can be seen as offering a more precise theoretical foundation for the distributed only view (Pulvermüller, 2005). This work tended to insist on the key role that distributed Hebbian cell assemblies would play in semantic processing, downplaying the role of potential amodal centers. However, his view seems to have evolved towards adopting a distributed-plus-hubs theory of semantic processing (see below) (Pulvermüller, 2013a,b). To my knowledge, the distributed-only view does not find much support currently, Pulvermüller remaining in some aspects the closest supporter of this theory, mainly due to its insistence on the role that general Hebbian wiring processes play in the organization of the semantic system.

#### Distributed + Hub

Against the distributed only view, it has been proposed that the semantic system rests on an architecture that incorporates both distributed representations in modal networks plus a single hub on which those

distributed representations converge and that hosts amodal representations (Patterson et al., 2007). This “hub-and-spoke” model of the semantic system places the hub in the anterior temporal lobe (aTL) whose role is to provide semantic representations that associates multiple modal features (but are thought of amodal, and not as heteromodal) in a task-independent way. For example, this hub can host a neuronal population that simultaneously links to a neural population that encodes a color (e.g. “red”, possibly encoded in visuo-ventral networks) and to a neural population that encodes for a sensori-motor phonemic code (e.g. “/red/”, possibly encoded in auditory-premotor perisylvian networks), to generate an amodal representation that can then be accessed, regardless of the task (e.g. processing the color of a red balloon in a visual scene, or talk about the color of the apple you just ate). Compared to the distributed-only model, the convergent architecture of the hub-and-spoke model offers a way to understand how generalization across concepts can occur. The neural populations in the hub can serve as a basis to link concepts that have similar semantic significance while differing in terms of their sensori-motor features. This can explain how our semantic system is able to build and manipulate rather complex semantic categories (Rosch et al., 1976; Rosch and Mervis, 1975), a capacity that is difficult to tackle using a distributed-only perspective. The model rests principally (but not only) on two types of results: empirical results from patients suffering from semantic dementia (SD) and simulation results. Patients suffering from SD typically show evidence of degeneration of the aTL bilaterally. Importantly they present symptoms of anomia (Hodges et al., 1995, 1992), difficulties in picture categorization (Rogers and Patterson, 2007), difficulties in detecting unusual body features in animal drawings (Rogers et al., 2004b), and show tendencies to use entry level category features even when irrelevant during tasks that require them to draw from memory (Bozeat et al., 2003). These results support the idea of a supramodal semantic role for aTL. Importantly no impairment affecting specifically the semantic system at a category level has been found following the deterioration of any other brain region. In addition, such over the board semantic impairments resulting from local degeneration of brain tissue seems hard to reconcile with a distributed-only view. Computationally, simulations have shown that a neural network built following the convergent architecture of the type hypothesized by the hub-and-spoke model was able to build representations that generalize across features (Rogers et al., 2004a), while lesion to the model’s hub resulted in semantic deficits comparable to those observed in SD patients.

It seems that the distributed+ hub model accounts for much more data points than the distributed-only model. Thompson-Schill did assign non-perceptual conceptual representation to the aTL, however, in her view, those seem to be associated with verbal representations. This points to a particularly difficult issue regarding the nature of semantic processing, namely whether the amodal representations are linguistic in nature. Data on SD show that if such patients are usually anomic, they also display difficulties in apparently non-linguistic tasks (drawing, categorizing images, etc). However, in absence of a clear model of how language structures our conceptual knowledge and how language processes interact with non-linguistic processes (such as visual processes), it is quite difficult to rule out the possibility that what are called amodal category level representations stored in aTL are really representations more directly linked to, and used in close contact with, lexical representations.

## Distributed + Hubs

More than twenty years ago, it had already been noted that theories putting too much emphasis on the integrative role of a single semantic hub were likely to be inadequate. In particular, the issue was already raised that those might, inadvertently or not, overemphasize language related aspects of semantic processing. This was in particular the position put forward by Damasio who, based mainly on neuroanatomical and lesion data, proposed the “convergence zones” theory of semantic processing (Damasio, 1989a,b). According to this theory, the semantic system is organized around multiple brain areas on which various types of semantic information converge and whose role is to bind them into more complex and stable representations (features are bound into objects, objects, into events). The associative role of convergence zones is similar to that proposed for the aTL in the hub-and-spoke model. However, a key hypothesis of Damasio’s theory is that there are multiple convergence zones, each specialized in binding specific types of lower level features in a way that generates evolutionary relevant categories (Damasio et al., 2004, 1996; Tranel et al., 1997). At its core, the convergence zone theory links the question of semantic processing directly to the binding problem, which places semantics in close relation to questions regarding the phenomenology of conscious experience and farther from questions more uniquely related to our language faculty. This is why the convergence zone

theory had denied to aTL a role as a unique amodal semantic center, arguing that patients with bilateral damages to aTL maintain a coherent perceptual experience (Damasio et al., 1987, 1985; Damasio, 1985). The convergence zone theory can therefore be seen as the first distributed + hubs theory in which semantic processing takes place on top of an architecture that is both distributed while also encompassing multiple hubs. Those hubs serve to build semantic categories relevant for the interface of semantic knowledge with other cognitive systems.

Recently, such distributed+hubs type of theory has regained traction against the distributed+hub and the distributed-only model of the semantic architecture. Taking together evidence from neuroimaging, neuropsychology, and neuroanatomy, Binder and Desai (2011) have proposed that the semantic system is organized around two large amodal regions (temporal lobe and inferior parietal cortex) that functions has semantic hubs on which multiple perceptual processing stream converge. Just as in the convergence zone theory and the hub-and-spoke, these hubs enable the formation of more abstract supramodal representations of the perceptual experiences that can then be used as a basis of our many cognitive functions that rely on conceptual knowledge. Importantly, these two hubs do not exhaust the architecture of the semantic system in which, as in all the previous models, modal areas also participate and encode semantic knowledge about modal features and qualities. Finally, Binder and Desai also make hypotheses about the role of brain regions that are not *sensu stricto* hosting semantic representations but are directly involved in semantic processes that make use of such representations. Regarding the definition of the hubs, the inclusion of the temporal lobe is unsurprising given that it has been generally associated with semantic processing in various modalities (language, vision, audition) and is coarsely in line with the hub-and-spoke theory. However, according to Binder and Desai, the temporal pole is not involved in specific amodal type of semantic processing, as hypothesized by the distributed+hub theory. The authors suggest that its unique connectivity pattern with hippocampal and parahippocampal regions makes it more likely to be involved in episodic memory encoding and retrieval. Interestingly, they point out that the alleged convergence of neural pathways onto the temporal pole (suggesting its role as a hub), if it holds for non-human primate neuroanatomical data (Felleman and Van Essen, 1991), might be largely overestimated in human (Orban et al., 2004). The inferior parietal cortex (IPC) is hypothesized to correspond to the second hub. The evidence supporting this inclusion stem both from neuroimaging studies that have found signatures of semantic effects within the angular gyrus (AG) (e.g. Humphries et al., 2007) and from data suggesting that IPC (but more generally the posterior perisylvian regions) could correspond to a “what” auditory stream (Rauschecker and Scott, 2009). Binder and Desai hypothesize that AG might be involved in encoding semantic representation of events unfolding in time and space. The theory also provides insights into the role that the prefrontal cortex could play during semantic processing. Coherent with the many theories and empirical results pointing to an involvement of the inferior frontal gyrus (IFG) in cognitive control (as part of a more general prefrontal network) (Badre, 2008; Botvinick et al., 2001; Koechlin et al., 2003), as well as with the data suggesting that, during language processing, IFG seems to be specifically involved in selection during ambiguity resolution processes (Novick et al., 2005, 2010; Rodd et al., 2005), the authors propose that IFG is involved in the selection of the relevant semantic information. On the other hand, they point out that the superior frontal gyrus (SFG), due to its preferential connections both to ventromedial PFC involved in emotion and reward processing and to the lateral PFC involved in cognitive control, could play a key role in translating affective states into plans for semantic knowledge retrieval.

In summary, Binder and Desai proposed a theory that does not fundamentally differ from that put forward by Damasio. They suggest that the semantic processing takes place on top of an architecture that is both distributed but contains two hubs on which information converges and that host supramodal representations that binds together sensori-motor semantic features that are stored in modal areas. The fact that convergence zones are necessary is shared by all the theories with the exception of the distributed-only view, theory that does not receive much support anymore. Most of the debate focuses on the number of convergence zones, their neural substrates, and their functions. The potential neural substrates of such hubs is actually quite uncontroversial since those are usually associated with classic amodal/heteromodal areas with the exception of PFC whose role is usually construed as control/executive. I would suggest that the number of “convergence zones” is actually a function of the definition that is chosen for “semantic processing”: the more complex are the psychological functions placed under the umbrella of semantic processing, the more numerous the convergence zones. The hub-and-spoke theory has quite a narrow focus on non-contextual category level semantic processing in tasks involving mostly language and/or vision, Binder and Desai review data including

a more diverse set of tasks and a more diverse set of empirical methods but remain within the realm of category-level semantic processing, finally Damasio links semantic processes to our capacity to build a stable subjective experience of the world. This progression in scope could explain the increase in the number of convergence zones hypothesized.

### 8.4.3 Embodiment, Disembodiment, Weak-embodiment

I will now narrow the focus to theories that analyze the relation between semantic and language processing. The role played by sensori-motor representations in generating meaning from utterances is a fundamental dimension that has for the past 15 years profoundly organized the landscape of such theories. Proponents of embodied theories of semantic processing insist on the role that sensori-motor programs can play decoupled from any action in the world to ground cognitive functions in modern humans, and in particular to ground language comprehension. Concepts of simulation or emulation are often used to refer such off-line uses of sensori-motor schemas during comprehension. Within such embodied theories coexist strong and weak claims. Strong embodiment views consider that simulations are the sole and ultimate support of language understanding (e.g. Feldman, 2010; Gallese † and Lakoff, 2005; Pulvermüller, 2005). To some extent the Perceptual Symbol System of L. W. Barsalou (1999) can be included within this category although his focus is on the symbol grounding problem rather than on the role of modal simulations. Weak embodiment or hybrid views on the other hand simply claim that motor/modal simulations can enrich and/or change the phenomenological quality of comprehension but are not necessary (e.g Binder and Desai, 2011; Dove, 2010; Meteyard et al., 2010). Conversely, disembodied theories claim that any involvement of the sensori-motor system during language comprehension is epiphenomenal or corollary to the language comprehension process in which it plays no direct causal role (Caramazza et al., 1990; Mahon and Caramazza, 2008).

Although it will not be possible to review these theories in details since this would be far outside the scope of this essay, I will try to indicate some key aspects of a few of them, to shed some light on their strength and weaknesses as well as on the type of data used to support them.

#### Strong Embodied Theories

Strong embodied theories uphold the idea that sensori-motor programs used in non-linguistic perceptual and motor tasks also form the ultimate “stuff” that composes the meaning of utterances. Focusing in particular on the comprehension of action-related words or propositions, these theories rest on empirical results that link comprehension to sensori-motor processes that would be involved in performing such action. Although not exclusively, such empirical results derive from studies analyzing interactions between sensori-motor behavioral performances and sentence comprehension (e.g. Borreggine and Kaschak, 2006; Glenberg and Kaschak, 2002; Sato et al., 2013; Stanfield and Zwaan, 2001; Zwaan et al., 2004, 2002) or increase BOLD responses in regions that somatotopically code for related effectors during the processing of utterances reporting an action or, in the case of utterances reporting a perceptual quality, an increase in BOLD activity in perceptual regions involved in processing such quality (e.g. Aziz-Zadeh et al., 2006; Hauk et al., 2004; Kemmerer and Gonzalez-Castillo, 2010, and the numerous other related neuroimaging results).

#### Disembodied View of Semantic Processing

Actively supported by Mahon and Caramazza (2008), the disembodied view of semantic processing rests in great part on the neuropsychological evidence that suggest that patients suffering brain lesions to sensori-motor areas can still correctly understand sentences that refer to motor acts. In particular, this theory points out that apraxic patients who are unable to use an object can still correctly name this object or pantomimes associated with the use of this object (Mahon and Caramazza, 2005; Negri et al., 2007; Rothi et al., 1991). In addition, proponents of this theory often mentioned that neuroimaging data does not provide evidence for a causal role played by sensori-motor cortices in comprehension. For this reason performances of patients with damaged motor areas remain key in the debate between strongly embodied and disembodied views.

A few recent results are worth mentioning. Grossman et al. (2008) studied patients with amyotrophic lateral sclerosis and found that atrophy of the motor cortex hinders comprehension of action verbs but not of nouns. However, contradicting results have been reported in (Arévalo et al., 2012) where the authors analyzed the performances of patients that had left-hemisphere strokes. Patients were asked whether a given

action word matched the action depicted in a picture or not and no correlation was found between the type of action (what body part it involves) and lesion to body-part-related motor areas. Papeo et al. (2010) reported a double-dissociation in patients with left-hemisphere strokes between the capacity to pantomime an action and the capacity to produce and understand the verb that refers to it. These results seem therefore to overall point towards an at best optional role of the motor system in language comprehension. Another line of work has been focusing on patients suffering from Parkinson's Disease (PD), looking at their performances in understanding action-related linguistic content compared to matched normal subjects. Comparisons between PD patients on medication and off medication are also carried out. Boulenger et al. (2008) compared the performances of PD patients ON and OFF medication in a lexical decision task using a masked repetition priming paradigm. They found an interaction between the type of word (object or action words) and the OFF or ON medication status of the patients with the OFF patients being slower to detect action words than object words, while no such difference was detected in ON medication patients. As pointed out in (Mahon and Caramazza, 2008), this experiment is however difficult to interpret since there is a confound between grammatical and semantic content of words (all the action related words were verbs and all the non-action related words were nouns). Fernandino et al. (2012) also analyzed PD patients linguistic performances using a semantic similarity judgment task. However, as pointed out by Kemmerer in (Kemmerer et al., 2013), they are flaws in the analysis of their data and if they found that PD patients were less accurate at judging action verbs than abstract verbs, the authors did not compare the accuracy between-group i.e. between matched controls and PD patients for the judgment of action verbs. Only this comparison would have allowed them to properly conclude whether or not PD patients are impaired in their judgment of action related verbs. Kemmerer et al. (2013) therefore recently published a study that addressed such issues. Their only solid result was that PD patients were slower overall when performing semantic similarity judgment on action-related words compared to matched controls (a difference that they claim is not due slow motor response although no proper control for this was included). And no difference was observed between patients ON medication and those OFF medication. In addition, no effect of PD on accuracy was observed. Of course, since the impact on PD on the capacity of a patient to carry out motor simulations is unknown, these results cannot be said to directly contradict the strong embodiment view. Nevertheless, taken together, neurological results seem to point towards a weak embodiment view.

### **Weak-Embodiment or Hybrid Views**

Weak-embodiment or hybrid views contend that both embodied and disembodied representations are jointly at play during semantic processes. The hub-and-spoke and the distributed+hubs theories reviewed in the previous section typically falls within this category as both amodal and modal brain regions form the semantic system whose processing flexibility results from the coexistence of category level and perceptual feature level representations. In addition, the feature and unitary semantic space hypothesis (FUSS) of Vigliocco and colleagues offer another example of a theory that brings together modal and amodal representations (Meteyard et al., 2010). FUSS, implemented as a statistical model, hypothesizes that word meanings are grounded in conceptual featural representations while offering the option to organize some of those according to the modality they belong to. In addition, it hypothesizes that such conceptual featural representations are also bundled into lexico-semantic representations that serve as an interface between conceptual knowledge and the language system (Vigliocco et al., 2004). Finally, Dove (2010) proposes a hybrid semantic processing theory that extends the concept of simulation to incorporate simulation over linguistic representations. In this view, part of our conceptual knowledge is represented in terms of sensori-motor simulations based on perceptuo-motor programs that can also support our physical interactions with the world, as hypothesized by embodied theories. However, Dove suggests that another part of our conceptual knowledge rests on sensori-motor simulations of language processing. These simulations make use semantic representations that are "dis-embodied" in the sense that they are dynamically multimodal. This position is quite appealing as it accounts for the role of amodal representations while maintaining the idea that simulation of sensori-motor programs forms the basis of semantic processing. This allows the anchoring of both types of semantic processes (modal and amodal) into the fundamental role that our nervous systems have evolved to play: controlling our sensori-motor coupling with the environment.

#### 8.4.4 Access vs. Composition: Problems of Linguistic Semantic Processing

It seems to me that one of the methodological problems that plagues an important part of the empirical studies linking semantic and language processing, particularly in the embodiment literature, is the restriction of what is meant by comprehension to something like (or at least in a great part defined by) lexical access. The focus on the comprehension of individual words as opposed to utterances restricts semantic issues to that of linking a word form to a word meaning and do not address crucial problems that linguists places under the category of “semantic processing”, problems such as compositionality, quantification, negation, and coercion. Indeed, none of the neurocognitive theories reviewed above gives us much information about how the brain perform semantic processing in the linguistic sense. For this reason, I would like to very briefly highlight a few theories that try to account for such issues<sup>2</sup>.

Pylkknen and colleagues have, in the past ten years, focused on carrying out neuroimaging studies that specifically target some of the core semantic operations as hypothesized by linguistic theories. In particular they have focused on complement coercion (Brennan and Pylkkänen, 2008; Pylkkänen and McElree, 2007) and simple composition (Bemis and Pylkkänen, 2011). Based on these experiments, Pylkknen has claimed that the ventromedial PFC plays a key role in semantic compositional operations, bringing at the forefront of semantic processes a brain region that so far was not considered as part of a semantic system per se (but had been extensively studied in relation to decision making processes (Bechara and Damasio, 2005)). In a similar vein, Piango studied coercion in aphasics showing that Broca’s aphasics seemed unimpaired for this type of semantic processes unlike Wernicke’s aphasics (Piñango and Zurif, 2001; Piñango, 2006). Both Piango and Pylkknen offer sketches of theories of semantic processing that heavily rely on linguistic theories, of Jackendoff (2002) for the former, and on generative grammar for the latter. Although these studies point to important aspects of semantic processing, they are mostly a direct application of linguistic theories to neurocognitive neuroscience. Therefore, so far, most the theoretical questions regarding of how compositional semantics is carried out by the brain remain relatively untouched.

Taking a very different approach, the neural theory of language (NTL) (Feldman and Narayanan, 2004) focuses on conceptual metaphor as a key theoretical construct to explain the complex relation between proposition level meanings and sensori-motor processes. It is to my knowledge the only theory that attempts to build a neurocognitive model of language processing that incorporates semantic processing at both the linguistic and conceptual level (Gallese † and Lakoff, 2005). Since this theory is also computationally implemented I will describe it in the next and final section.

#### 8.4.5 Computational Theories

To conclude this discussion of the neuroscience of semantic processing, I would like to consider existing computational theories. In the first section, I have already mentioned the neural network implementation of a distributed+hub model of the semantic system (Rogers et al., 2004a). This model follows the general connectionist principles outlined by McClelland (McClelland et al., 2010; Rumelhart and McClelland, 1986) to show how category level representations can be learned in a neural layer that receives convergent information about various types of features. In the second section I have mentioned the FUSS model that uses statistical procedures to learn organized conceptual features representations that are associated with lexical items (Vigliocco et al., 2004).

In the first section, Pulvermüller was mentioned as one of the proponent of what could be considered a distributed-only theory. His group has offered some computational modeling work to support his theoretical claims. In particular, his theory rests on the idea that semantic processing as well as the organization of the semantic system can be accounted to by a general Hebbian associative theory of learning. Semantic representations are supported by a network of distributed neural populations that, through repeated coactivation, have become bundled together into Hebbian cell assemblies (Braitenberg, 1978). The creation and use of such cell assemblies has been computationally investigated and was shown be capable of supporting some generalization over surface co-occurrence patterns in word sequences (Pulvermüller and Knoblauch, 2009) and linkages between sensory and motor areas during word acquisition (Garagnani et al., 2008). So far however, this line of work does not seem to scale up easily to account for more complex semantic processes.

<sup>2</sup>And there are many more studies that would be worth mentioning here (see for example Aravena et al., 2012; Raposo et al., 2009; Tettamanti et al., 2008; Tomasino et al., 2010). But I would like to limit the discussion to work that has been associated to some theoretical claims about semantic processing.

As mentioned in the previous section, the Neural Theory of Language (Feldman and Narayanan, 2004) offers a theory of semantic processing unique in scope. At its core, the computational elements that are the X-Schemas and the Embodied Construction Grammar (ECG), offer the computational counterpart to the strong embodiment claim of Gallese † and Lakoff (2005). This framework seeks to explain language comprehension in terms of sensory-motor simulations on which linguistic meaning can be directly anchored in the case of concrete action sentences. In the case of abstract sentences, the idea is that the pervasive use of metaphorical mappings from an abstract target domain (e.g. International economics) onto an embodied source domain (as in “The liberalization plan stumbled”) makes such simulations possible. The concept of X-schemas (Narayanan, 1999) was developed as a way to computationally represent the hierarchy of premotor structures that package some of the motor control into a limited set of parameters and can be used either to direct action in the world or to carry out offline simulations that in turn form the basis of language comprehension. Narayanan was successful in showing how a system that contains (1) abstract world knowledge about a target domain (knowledge of international economics coded as a Belief Network), (2) sensory-motor knowledge represented as a network of X-schemas, and (3) metaphorical mappings between the two, linking belief values to X-Schemas parameters, could generate correct inferences, when presented with a newspaper headline such as “Liberalization plan stumbling”: that there is an ongoing economic plan, that it is facing difficulties, and that it is likely to fail. Such inferences are possible because the system can use X-Schemas to simulate the effect of stumbling on a WALK schema and map the resulting state (falling state unless a lot of force is applied) to the concept of difficulty and failure in the target domain of economic policy. Expanding on this core semantic processing computational model later work introduced ECG as a way to use grammatical constructions to map more generally linguistic content into conceptual schemas that can set the parameters of the X-Schemas which then carry the sensory-motor simulations necessary to generate the proper inferences (Bergen and Chang, 2005; Feldman, 2010). This work clearly illustrates how motor schemas can be harnessed to support the inferences that have to necessarily accompany language comprehension including in abstract domains. However, it also shows that sensory-motor simulations are useful only in that they can interact with the abstract knowledge of the source domain. The question of the nature of brain structures that support the metaphorical mappings and the constructions (and their assemblage) are left rather unspecified, opening an interesting path of research. Notwithstanding this limitation, when compared to models directly implemented in neural networks and the limited advances they have offered, NTL clearly highlights the power of working at a more symbolic level in order to capture the more sophisticated aspects of semantic processing.

#### 8.4.6 Conclusion

Whether by focusing on the structure of the semantic system or by analyzing, in the case of language comprehension, the types of representations involved in generating meaning from utterances, it appears that neurocognitive theories of semantic processing tend to support the ideas that: (1) convergent brain structures play a key role in shaping our conceptual knowledge in particular with regard to the emergence of category level representations, a fact supported by connectionist simulations; (2) both embodied and disembodied representations are involved in semantic processes (in particular in those that support language comprehension). However, I pointed out the fact that a large part of the work on semantic processing during language comprehension focuses on lexical access, with fewer attempts to analyze how the brain handles semantic operations such as semantic composition. This might reflect a long-lasting tendency to tackle the question of semantic processing from the point of view of semantic memory, most likely stemming from the idea that “execution” of semantic operations is supported by domain general executive systems. This, overall, results in certain shortage of theories that address the question of how semantic representations are manipulated. A few computational models of semantic processing have nevertheless tried to offer some insights into this problem. As the comparison between Puvermuller’s and NTL highlights, one of the main lessons that can be drawn from such endeavors is that symbolic-level models seem more fitted to the task than neural-level implementations, although the question of how the former could be anchored in brain systems remains unsolved. I cannot here do justice to all the theories that have been put forward to explain how our brain generates and manipulates semantic contents. In particular, I have eschewed altogether the theories that address semantic processing from a situated cognition perspective (Clark, 2006; Engel et al., 2013; Pylyshyn, 2001). I can justify this choice by simply saying that those theories tend to be quite limited

in terms of the neuroscience data they incorporate. But their insistence on the role that the interactions between our body, its physical environment, and other bodies play in supporting our semantic processing abilities might reveal invaluable for any neurocognitive theory that will try to address our semantic abilities from an evolutionary perspective, and that will place the question of how we make use of meaningful contents at their core.

# Bibliography

- Abelson, R. P. (1981). Psychological status of the script concept. *American Psychologist*, 36(7):715–729.
- Aboitiz, F. (2012). Gestures, vocalizations, and memory in language origins. *Frontiers in Evolutionary Neuroscience*, 4:2.
- Allen, K., Pereira, F., Botvinick, M., and Goldberg, A. E. (2012). Distinguishing grammatical constructions with fMRI pattern analysis. *Brain and Language*, 123(3):174–182.
- Altmann, G. T. and Kamide, Y. (1999). Incremental interpretation at verbs: restricting the domain of subsequent reference. *Cognition*, 73(3):247–264.
- Altmann, G. T. M. and Mirković, J. (2009). Incrementality and Prediction in Human Sentence Processing. *Cognitive Science*, 33(4):583–609.
- Amunts, K. and Zilles, K. (2012). Architecture and organizational principles of Broca’s region. *Trends in Cognitive Sciences*, 16(8):418–426.
- Ansell, B. J. and Flowers, C. R. (1982). Aphasic adults’ use of heuristic and structural linguistic cues for sentence analysis. *Brain and language*, 16(1):61–72.
- Anwander, A., Tittgemeyer, M., von Cramon, D. Y., Friederici, A. D., and Knösche, T. R. (2007). Connectivity-Based Parcellation of Broca’s Area. *Cerebral Cortex (New York, N.Y.: 1991)*, 17(4):816–825.
- Aravena, P., Delevoye-Turrell, Y., Deprez, V., Cheylus, A., Paulignan, Y., Frak, V., and Nazir, T. (2012). Grip force reveals the context sensitivity of language-induced motor activity during “action words” processing: evidence from sentential negation. *PloS one*, 7(12):e50287.
- Arbib, M. and Bota, M. (2003). Language evolution: neural homologies and neuroinformatics. *Neural Networks: The Official Journal of the International Neural Network Society*, 16(9):1237–1260.
- Arbib, M. A. (1981). Perceptual structures and distributed motor control. *Comprehensive Physiology*.
- Arbib, M. A. (1989). *The Metaphorical Brain 2: Neural Networks and Beyond*. John Wiley & Sons, Inc., New York, NY, USA, 2nd edition.
- Arbib, M. A. (2008). From grasp to language: Embodied concepts and the challenge of abstraction. *Journal of Physiology-Paris*, 102(1–3):4–20.
- Arbib, M. A. (2010). Mirror system activity for action and language is embedded in the integration of dorsal and ventral pathways. *Brain and Language*, 112(1):12–24.
- Arbib, M. A. (2012). *How the brain got language: the mirror system hypothesis*, volume 16. Oxford University Press.
- Arbib, M. A. (2016a). Primates, computation, and the path to language. *Physics of Life Reviews*, 16:105–122.
- Arbib, M. A. (2016b). Toward the Language-Ready Brain: Biological Evolution and Primate Comparisons. *Psychonomic Bulletin & Review*.

- Arbib, M. A., Billard, A., Iacoboni, M., and Oztop, E. (2000). Synthetic brain imaging: grasping, mirror neurons and imitation. *Neural Networks: The Official Journal of the International Neural Network Society*, 13(8-9):975–997.
- Arbib, M. A., Bischoff, A., Fagg, A. H., and Grafton, S. T. (1994). Synthetic PET: Analyzing large-scale properties of neural networks. *Human Brain Mapping*, 2(4):225–233.
- Arbib, M. A. and Bonaiuto, J. (2008). From grasping to complex imitation: mirror systems on the path to language. *Mind & Society*, 7(1):43–64.
- Arbib, M. A. and Bota, M. (2004). Response to Deacon: Evolving mirror systems: homologies and the nature of neuroinformatics. *Trends in Cognitive Sciences*, 8(7):290–291.
- Arbib, M. A. and Caplan, D. (1979). Neurolinguistics must be computational. *Behavioral and Brain Sciences*, 2(03):449–460.
- Arbib, M. A., Conklin, E. J., and Hill, J. A. C. (1987). *From schema theory to language*. Oxford University Press.
- Arbib, M. A. and Lee, J. (2007). Vision and Action in the Language-Ready Brain: From Mirror Neurons to SemRep. In Mele, F., Ramella, G., Santillo, S., and Ventriglia, F., editors, *Advances in Brain, Vision, and Artificial Intelligence*, number 4729 in Lecture Notes in Computer Science, pages 104–123. Springer Berlin Heidelberg.
- Arbib, M. A. and Lee, J. (2008). Describing visual scenes: towards a neurolinguistics based on construction grammar. *Brain Research*, 1225:146–162.
- Arévalo, A. L., Baldo, J. V., and Dronkers, N. F. (2012). What do brain lesions tell us about theories of embodied semantics and the human mirror neuron system? *Cortex*, 48(2):242–254.
- Auzias, G., Colliot, O., Glaunes, J., Perrot, M., Mangin, J.-F., Trouve, A., and Baillet, S. (2011). Diffeomorphic Brain Registration Under Exhaustive Sulcal Constraints. *IEEE Transactions on Medical Imaging*, 30(6):1214 –1227.
- Awh, E., Jonides, J., Smith, E. E., Schumacher, E. H., Koeppen, R. A., and Katz, S. (1996). Dissociation of storage and rehearsal in verbal working memory: Evidence from positron emission tomography. *Psychological Science*, pages 25–31.
- Aziz-Zadeh, L., Wilson, S. M., Rizzolatti, G., and Iacoboni, M. (2006). Congruent embodied representations for visually presented actions and linguistic phrases describing actions. *Current biology*, 16(18):1818–1823.
- Badre, D. (2008). Cognitive control, hierarchy, and the rostro-caudal organization of the frontal lobes. *Trends in cognitive sciences*, 12(5):193–200.
- Badre, D. and Wagner, A. D. (2007). Left ventrolateral prefrontal cortex and the cognitive control of memory. *Neuropsychologia*, 45(13):2883–2901.
- Bahlmann, J., Schubotz, R. I., and Friederici, A. D. (2008). Hierarchical artificial grammar processing engages Broca’s area. *Neuroimage*, 42(2):525–534.
- Ballard, D. H., Hayhoe, M. M., Pook, P. K., and Rao, R. P. (1997). Deictic codes for the embodiment of cognition. *The Behavioral and Brain Sciences*, 20(4):723–742; discussion 743–767.
- Baluch, F. and Itti, L. (2011). Mechanisms of top-down attention. *Trends in Neurosciences*, 34(4):210–224.
- Barrès, V. and Lee, J. (2013). Template Construction Grammar: from visual scene description to language comprehension and agrammatism. *Neuroinformatics*, pages 1–28.
- Barrès, V., Simons III, A., and Arbib, M. (2013). Synthetic event-related potentials: A computational bridge between neurolinguistic models and experiments. *Neural Networks*, 37:66–92.

- Barres, V. J. (2017). Template Construction Grammar: A Schema-Theoretic Computational Construction Grammar. In *2017 AAAI Spring Symposium Series*.
- Bartlett, F. C. (1932). Remembering: An experimental and social study. *Cambridge: Cambridge University*.
- Bates, E. and MacWhinney, B. (1989). Functionalism and the competition model. *The crosslinguistic study of sentence processing*, pages 3–73.
- Bechara, A. and Damasio, A. R. (2005). The somatic marker hypothesis: A neural theory of economic decision. *Games and economic behavior*, 52(2):336–372.
- Bedny, M., Hulbert, J. C., and Thompson-Schill, S. L. (2007). Understanding words in context: The role of Broca's area in word comprehension. *Brain research*, 1146:101–114.
- beim Graben, P., Gerth, S., and Vasishth, S. (2008). Towards dynamical system models of language-related brain potentials. *Cognitive neurodynamics*, 2(3):229–255.
- Beim Graben, P., Jurish, B., Saddy, D., and Frisch, S. (2004). Language processing by dynamical systems. *International Journal of Bifurcation and Chaos*, 14(02):599–621.
- beim Graben, P., Liebscher, T., and Kurths, J. (2007). Neural and cognitive modeling with networks of leaky integrator units. In *Lectures in Supercomputational Neurosciences*, pages 195–223. Springer.
- Beim Graben, P., Pinotsis, D., Saddy, D., and Potthast, R. (2008). Language processing with dynamic fields. *Cognitive Neurodynamics*, 2(2):79–88.
- Bemis, D. K. and Pylkkänen, L. (2011). Simple composition: a magnetoencephalography investigation into the comprehension of minimal linguistic phrases. *The Journal of Neuroscience*, 31(8):2801–2814.
- Bergen, B. and Chang, N. (2005). Embodied construction grammar in simulation-based language understanding. *Construction grammars: Cognitive grounding and theoretical extensions*, pages 147–190.
- Berndt, R. S. and Caramazza, A. (1999). How “regular” is sentence comprehension in Broca's aphasia? It depends on how you select the patients. *Brain and Language*, 67(3):242–247.
- Berndt, R. S., Mitchum, C. C., and Haendiges, A. N. (1996). Comprehension of reversible sentences in “agrammatism”: a meta-analysis. *Cognition*, 58(3):289–308.
- Beuls, K. and Steels, L. (2013). Agent-Based Models of Strategies for the Emergence and Evolution of Grammatical Agreement. *PLoS ONE*, 8(3):e58960.
- Binder, J. R. and Desai, R. H. (2011). The neurobiology of semantic memory. *Trends in Cognitive Sciences*, 15(11):527–536.
- Bloomfield, L. (1962). Language, 1933. *Holt, New York*.
- Boas, H. C. and Sag, I. A. (2012). *Sign-Based Construction Grammar*. CSLI Publications/Center for the Study of Language and Information.
- Bock, K., Irwin, D. E., and Davidson, D. J. (2004). Putting first things first.
- Bock, K. and Levelt, W. (2002). Language production. *Psycholinguistics: Critical concepts in psychology*, 5:405.
- Bogacz, R., Usher, M., Zhang, J., and McClelland, J. L. (2007). Extending a biologically inspired model of choice: multi-alternatives, nonlinearity and value-based multidimensional choice. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 362(1485):1655–1670.
- Bohland, J. W., Bullock, D., and Guenther, F. H. (2010). Neural representations and mechanisms for the performance of simple speech sequences. *Journal of Cognitive Neuroscience*, 22(7):1504–1529.

- Borji, A. and Itti, L. (2013). State-of-the-art in visual attention modeling.
- Bornkessel, I. and Schlesewsky, M. (2006). The extended argument dependency model: a neurocognitive approach to sentence comprehension across languages. *Psychological Review*, 113(4):787–821.
- Bornkessel-Schlesewsky, I. and Schlesewsky, M. (2013). Reconciling time, space and function: A new dorsal–ventral stream model of sentence comprehension. *Brain and Language*, 125(1):60–76.
- Borreggine, K. L. and Kaschak, M. P. (2006). The action–sentence compatibility effect: It’s all in the timing. *Cognitive Science*, 30(6):1097–1112.
- Bota, M. and Arbib, M. (2004). Integrating databases and expert systems for the analysis of brain structures. *Neuroinformatics*, 2(1):19–58.
- Botvinick, M. M., Braver, T. S., Barch, D. M., Carter, C. S., and Cohen, J. D. (2001). Conflict monitoring and cognitive control. *Psychological review*, 108(3):624.
- Boulenger, V., Mechtaouf, L., Thobois, S., Broussolle, E., Jeannerod, M., and Nazir, T. A. (2008). Word processing in Parkinson’s disease is impaired for action verbs but not for concrete nouns. *Neuropsychologia*, 46(2):743–756.
- Bozeat, S., Lambon Ralph, M. A., Graham, K. S., Patterson, K., Wilkin, H., Rowland, J., Rogers, T. T., and Hodges, J. R. (2003). A duck with four legs: Investigating the structure of conceptual knowledge using picture drawing in semantic dementia. *Cognitive Neuropsychology*, 20(1):27–47.
- Braitenberg, V. (1978). Cell assemblies in the cerebral cortex. In *Theoretical approaches to complex systems*, pages 171–188. Springer.
- Brennan, J. and Pylkkänen, L. (2008). Processing events: behavioral and neuromagnetic correlates of Aspectual Coercion. *Brain and Language*, 106(2):132–143.
- Brouwer, H., Crocker, M. W., Venhuizen, N. J., and Hoeks, J. C. (2016). A Neurocomputational Model of the N400 and the P600 in Language Processing. *Cognitive Science*.
- Brouwer, H., Fitz, H., and Hoeks, J. (2012). Getting real about Semantic Illusions: Rethinking the functional role of the P600 in language comprehension. *Brain Research*, 1446:127–143.
- Brown-Schmidt, S. and Tanenhaus, M. K. (2006). Watching the eyes when talking about size: An investigation of message formulation and utterance planning. *Journal of Memory and Language*, 54(4):592–609.
- Bryant, J. E. (2008). *Best-fit Constructional Analysis*. PhD thesis.
- Buchsbaum, B. R., Hickok, G., and Humphries, C. (2001). Role of left posterior superior temporal gyrus in phonological processing for speech perception and production. *Cognitive Science*, 25(5):663–678.
- Buchsbaum, B. R., Olsen, R. K., Koch, P., and Berman, K. F. (2005). Human dorsal and ventral auditory streams subserve rehearsal-based and echoic processes during verbal working memory. *Neuron*, 48(4):687–697.
- Caplan, D., Alpert, N., Waters, G., and Olivieri, A. (2000). Activation of Broca’s area by syntactic processing under conditions of concurrent articulation. *Human brain mapping*, 9(2):65–71.
- Caplan, D., Baker, C., and Dehaut, F. (1985). Syntactic determinants of sentence comprehension in aphasia. *Cognition*, 21(2):117–175.
- Caplan, D., Stanczak, L., and Waters, G. (2008). Syntactic and thematic constraint effects on blood oxygenation level dependent signal correlates of comprehension of relative clauses. *Journal of Cognitive Neuroscience*, 20(4):643–656.
- Caramazza, A., Capasso, R., Capitani, E., and Miceli, G. (2005). Patterns of comprehension performance in agrammatic Broca’s aphasia: A test of the Trace Deletion Hypothesis. *Brain and Language*, 94(1):43–53.

- Caramazza, A., Hillis, A. E., Rapp, B. C., and Romani, C. (1990). The multiple semantics hypothesis: Multiple confusions? *Cognitive Neuropsychology*, 7(3):161–189.
- Caramazza, A. and Zurif, E. B. (1976). Dissociation of algorithmic and heuristic processes in language comprehension: Evidence from aphasia. *Brain and Language*, 3(4):572–582.
- Catani, M., Jones, D. K., and ffytche, D. H. (2005). Perisylvian language networks of the human brain. *Annals of Neurology*, 57(1):8–16.
- Chang, F., Dell, G. S., and Bock, K. (2006). Becoming syntactic. *Psychological review*, 113(2):234.
- Chang, N., De Beule, J., and Micelli, V. (2012). Computational Construction Grammar: Comparing ECG and FCG. In Steels, L., editor, *Computational Issues in Fluid Construction Grammar*, volume 7249 of *Lecture Notes in Computer Science*, pages 259–288. Springer Berlin / Heidelberg.
- Chang, N. C.-L. (2008). *Constructing grammar: A computational model of the emergence of early constructions*. ProQuest.
- Chomsky, N. (1995). *The minimalist program*, volume 28. Cambridge Univ Press.
- Chomsky, N. (2002). *Syntactic structures*. Walter de Gruyter.
- Christianson, K., Hollingworth, A., Halliwell, J. F., and Ferreira, F. (2001). Thematic Roles Assigned along the Garden Path Linger. *Cognitive Psychology*, 42(4):368–407.
- Christianson, K. and Luke, S. G. (2011). Context Strengthens Initial Misinterpretations of Text. *Scientific Studies of Reading*, 15(2):136–166. WOS:000287705500002.
- Cisek, P. (2006). Integrated Neural Processes for Defining Potential Actions and Deciding between Them: A Computational Model. *The Journal of Neuroscience*, 26(38):9761–9770.
- Cisek, P. and Kalaska, J. F. (2010). Neural Mechanisms for Interacting with a World Full of Action Choices. *Annual Review of Neuroscience*, 33(1):269–298.
- Clark, A. (2006). Language, embodiment, and the cognitive niche. *Trends in Cognitive Sciences*, 10(8):370–374.
- Cooper, R. and Shallice, T. (2000). Contention scheduling and the control of routine activities. *Cognitive neuropsychology*, 17(4):297–338.
- Cooper, R. P., Schwartz, M. F., Yule, P., and Shallice, T. (2005). The simulation of action disorganisation in complex activities of daily living. *Cognitive Neuropsychology*, 22(8):959–1004.
- Cooper, R. P. and Shallice, T. (2006). Hierarchical schemas and goals in the control of sequential behavior. *Psychological Review*, 113:887–916.
- Corbacho, F., Nishikawa, K. C., Weerasuriya, A., Liaw, J.-S., and Arbib, M. A. (2005a). Schema-based learning of adaptable and flexible prey- catching in anurans II. Learning after lesioning. *Biological Cybernetics*, 93(6):410–425.
- Corbacho, F., Nishikawa, K. C., Weerasuriya, A., Liaw, J.-S., and Arbib, M. A. (2005b). Schema-based learning of adaptable and flexible prey-catching in anurans I. The basic architecture. *Biological Cybernetics*, 93(6):391–409.
- Corbacho, F. J. and Arbib, M. A. (1995). Learning to Detour. *Adaptive Behavior*, 3(4):419–468.
- Cottrell, G. W. (1985). Implications of Connectionist Parsing for Aphasia. *Proceedings of the Annual Symposium on Computer Application in Medical Care*, page 237.
- Crocker, M. W., Knoeferle, P., and Mayberry, M. R. (2010). Situated sentence processing: The coordinated interplay account and a neurobehavioral model. *Brain and Language*, 112(3):189–201. WOS:000276320900007.

- Croft, W. (2001). *Radical Construction Grammar: Syntactic Theory in Typological Perspective*. Oxford University Press.
- Croft, W. (2005). Logical and typological arguments for Radical Construction Grammar. *Construction Grammars: Cognitive grounding and theoretical extensions*, pages 273–314.
- Croft, W. and Cruse, D. A. (2004). *Cognitive linguistics*. Cambridge University Press.
- Dale, A. M., Fischl, B., and Sereno, M. I. (1999). Cortical surface-based analysis. I. Segmentation and surface reconstruction. *NeuroImage*, 9(2):179–194.
- Damasio, A. R. (1985). Disorders of complex visual processing: agnosias, achromatopsia, Balint's syndrome, and related difficulties of orientation and construction. *Principles of behavioural neurology*, 1:259–288.
- Damasio, A. R. (1989a). The Brain Binds Entities and Events by Multiregional Activation from Convergence Zones. *Neural Computation*, 1(1):123–132.
- Damasio, A. R. (1989b). Time-locked multiregional retroactivation: A systems-level proposal for the neural substrates of recall and recognition. *Cognition*, 33(1):25–62.
- Damasio, A. R., Damasio, H., Tranel, D., Welsh, K., and Brandt, J. (1987). Additional neural and cognitive evidence in patient DRB. *Society for Neuroscience*, 13:1452.
- Damasio, A. R., Eslinger, P. J., Damasio, H., Van Hoesen, G. W., and Cornell, S. (1985). Multimodal amnesia syndrome following bilateral temporal and basal forebrain damage. *Archives of Neurology*, 42(3):252.
- Damasio, H., Grabowski, T. J., Tranel, D., and Hichwa, R. D. (1996). A neural basis for lexical retrieval. *Nature*.
- Damasio, H., Tranel, D., Grabowski, T., Adolphs, R., and Damasio, A. (2004). Neural systems behind word and concept retrieval. *Cognition*, 92(1–2):179–229.
- David, O., Kiebel, S. J., Harrison, L. M., Mattout, J., Kilner, J. M., and Friston, K. J. (2006). Dynamic causal modeling of evoked responses in EEG and MEG. *NeuroImage*, 30(4):1255–1272.
- David, O., Maess, B., Eckstein, K., and Friederici, A. D. (2011). Dynamic Causal Modeling of Subcortical Connectivity of Language. *The Journal of Neuroscience*, 31(7):2712–2717.
- Deacon, T. (2004). Monkey homologues of language areas: computing the ambiguities. *Trends in Cognitive Sciences*, 8(7):288–290; discussion 290–291.
- Dick, F., Bates, E., Wulfeck, B., Utman, J. A., Dronkers, N., and Gernsbacher, M. A. (2001). Language deficits, localization, and grammar: Evidence for a distributive model of language breakdown in aphasic patients and neurologically intact individuals. *Psychological Review*, 108(4):759–788.
- Dominey, P., Arbib, M., and Joseph, J.-P. (1995). A Model of Corticostriatal Plasticity for Learning Oculomotor Associations and Sequences. *Journal of Cognitive Neuroscience*, 7(3):311–336.
- Dominey, P. F. and Arbib, M. A. (1992). A Cortico-Subcortical Model for Generation of Spatially Accurate Sequential Saccades. *Cerebral Cortex*, 2(2):153–175.
- Dominey, P. F. and Boucher, J.-D. (2005). Learning to talk about events from narrated video in a construction grammar framework. *Artificial Intelligence*, 167(1–2):31–61.
- Dominey, P. F., Hoen, M., Blanc, J.-M., and Lelekov-Boissard, T. (2003). Neurological basis of language and sequential cognition: Evidence from simulation, aphasia, and ERP studies. *Brain and Language*, 86(2):207–225.
- Dominey, P. F., Hoen, M., and Inui, T. (2006). A neurolinguistic model of grammatical construction processing. *Journal of Cognitive Neuroscience*, 18(12):2088–2107.

- Dominey, P. F., Inui, T., and Hoen, M. (2009). Neural network processing of natural language: II. Towards a unified model of corticostriatal function in learning sentence comprehension and non-linguistic sequencing. *Brain and Language*, 109(2–3):80–92.
- Dove, G. (2010). On the need for Embodied and Dis-Embodied Cognition. *Frontiers in Psychology*, 1:242.
- Draper, B., Collins, R., Brolio, J., Hanson, A., and Riseman, E. (1988). The Schema System. Technical report, University of Massachusetts, Amherst, MA, USA.
- Ehinger, K. A., Hidalgo-Sotelo, B., Torralba, A., and Oliva, A. (2009). Modelling search for people in 900 scenes: A combined source model of eye guidance. *Visual Cognition*, 17(6-7):945–978.
- Engel, A. K., Maye, A., Kurthen, M., and König, P. (2013). Where's the action? The pragmatic turn in cognitive science. *Trends in Cognitive Sciences*, 17(5):202–209.
- Erman, L. D., Hayes-Roth, F., Lesser, V. R., and Reddy, D. R. (1980). The Hearsay-II Speech-Understanding System: Integrating Knowledge to Resolve Uncertainty. *ACM Comput. Surv.*, 12(2):213–253.
- Fedorenko, E., Duncan, J., and Kanwisher, N. (2012). Language-Selective and Domain-General Regions Lie Side by Side within Broca's Area. *Current Biology*.
- Fedorenko, E. and Thompson-Schill, S. L. Reworking the language network. *Trends in Cognitive Sciences*.
- Feldman, J. (2010). Embodied language, best-fit analysis, and formal compositionality. *Physics of Life Reviews*, 7(4):385–410.
- Feldman, J. and Narayanan, S. (2004). Embodied meaning in a neural theory of language. *Brain and language*, 89(2):385–392.
- Feldman, J. A. and Ballard, D. H. (1982). Connectionist models and their properties. *Cognitive science*, 6(3):205–254.
- Felleman, D. J. and Van Essen, D. C. (1991). Distributed hierarchical processing in the primate cerebral cortex. *Cerebral cortex*, 1(1):1–47.
- Fernandino, L., Conant, L. L., Binder, J. R., Blindauer, K., Hiner, B., Spangler, K., and Desai, R. H. (2012). Parkinson's disease disrupts both automatic and controlled processing of action verbs. *Brain and Language*.
- Ferreira, F. (2003). The misinterpretation of noncanonical sentences. *Cognitive Psychology*, 47(2):164–203.
- Ferreira, F. and Patson, N. D. (2007). The 'Good Enough' Approach to Language Comprehension. *Language and Linguistics Compass*, 1(1-2):71–83.
- Filipe, S. and Alexandre, L. A. (2013). From the human visual system to the computational models of visual attention: a survey. *Artificial Intelligence Review*, pages 1–47.
- Fillmore, C. J. (1976). Frame Semantics and the Nature of Language. *Annals of the New York Academy of Sciences*, 280(1):20–32.
- Fischl, B., Rajendran, N., Busa, E., Augustinack, J., Hinds, O., Yeo, B. T. T., Mohlberg, H., Amunts, K., and Zilles, K. (2008). Cortical Folding Patterns and Predicting Cytoarchitecture. *Cerebral Cortex*, 18(8):1973–1980.
- Fischl, B., Sereno, M. I., and Dale, A. M. (1999). Cortical surface-based analysis. II: Inflation, flattening, and a surface-based coordinate system. *NeuroImage*, 9(2):195–207.
- Fletcher, P. C., Shallice, T., and Dolan, R. J. (2000). "Sculpting the response space"—an account of left prefrontal activation at encoding. *Neuroimage*, 12(4):404–417.
- Fodor, J. A. (1983). *The modularity of mind: An essay on faculty psychology*. MIT press.

- Frazier, L. and Fodor, J. D. (1978). The sausage machine: A new two-stage parsing model. *Cognition*, 6(4):291–325.
- Friederici, A. D. (2002). Towards a neural basis of auditory sentence processing. *Trends in Cognitive Sciences*, 6(2):78–84.
- Friederici, A. D. (2009). Pathways to language: fiber tracts in the human brain. *Trends in Cognitive Sciences*, 13(4):175–181.
- Friederici, A. D. (2011). The brain basis of language processing: from structure to function. *Physiological Reviews*, 91(4):1357–1392.
- Friederici, A. D. (2012). The cortical language circuit: from auditory perception to sentence comprehension. *Trends in Cognitive Sciences*, 16(5):262–268.
- Friederici, A. D., Rüschemeyer, S.-A., Hahne, A., and Fiebach, C. J. (2003). The role of left inferior frontal and superior temporal cortex in sentence comprehension: localizing syntactic and semantic processes. *Cerebral cortex*, 13(2):170–177.
- Frintrop, S., Rome, E., and Christensen, H. I. (2010). Computational visual attention systems and their cognitive foundations: A survey. *ACM Transactions on Applied Perception (TAP)*, 7(1):6.
- Fromkin, V. A. (1984). *Speech errors as linguistic evidence*, volume 77. Walter de Gruyter.
- Gallese †, V. and Lakoff, G. (2005). The Brain's concepts: the role of the Sensory-motor system in conceptual knowledge. *Cognitive Neuropsychology*, 22(3-4):455–479.
- Garagnani, M., Wennekers, T., and Pulvermüller, F. (2008). A neuroanatomically grounded Hebbian-learning model of attention-language interactions in the human brain. *The European Journal of Neuroscience*, 27(2):492–513.
- Garrett, M. F. (1980). Levels of processing in sentence production. *Language production*, 1:177–220.
- Garrod, S. and Pickering, M. J. (2004). Why is conversation so easy? *Trends in Cognitive Sciences*, 8(1):8–11.
- Glasser, M. F. and Rilling, J. K. (2008). DTI tractography of the human brain's language pathways. *Cerebral Cortex*, 18(11):2471–2482.
- Gleason, J. B., Goodglass, H., Green, E., Ackerman, N., and Hyde, M. R. (1975). The retrieval of syntax in Broca's aphasia. *Brain and Language*, 2:451–471.
- Gleitman, L. R., January, D., Nappa, R., and Trueswell, J. C. (2007). On the give and take between event apprehension and utterance formulation. *Journal of Memory and Language*, 57(4):544–569.
- Glenberg, A. M. and Kaschak, M. P. (2002). Grounding language in action. *Psychonomic Bulletin & Review*, 9(3):558–565.
- Goldberg, A. E. (1995). *Construction grammar*. Wiley Online Library.
- Goodglass, H. (1968). Studies on the grammar of aphasics. *Developments in applied psycholinguistics research*. New York: Macmillan, pages 177–208.
- Goodglass, H. (1976). Agrammatism. *Studies in neurolinguistics*, 1:237–260.
- Goodglass, H. and Berko, J. (1960). Agrammatism and inflectional morphology in English. *Journal of Speech, Language and Hearing Research*, 3(3):257.
- Grafton, S. T., Fagg, A. H., and Arbib, M. A. (1998). Dorsal premotor cortex and conditional movement selection: a PET functional mapping study. *Journal of Neurophysiology*, 79(2):1092–1097.
- Griffin, Z. M. and Bock, K. (2000). What the eyes say about speaking. *Psychological science*, 11(4):274–279.

- Grodzinsky, Y. (2000). The neurology of syntax: Language use without Broca's area. *Behavioral and Brain Sciences*, 23(01):1–21.
- Grodzinsky, Y. and Marek, A. (1988). Algorithmic and heuristic processes revisited. *Brain and language*, 33(2):216–225.
- Grodzinsky, Y., Piñango, M. M., Zurif, E., and Drai, D. (1999). The Critical Role of Group Studies in Neuropsychology: Comprehension Regularities in Broca's Aphasia. *Brain and Language*, 67(2):134–147.
- Grodzinsky, Y. and Santi, A. (2008). The battle for Broca's region. *Trends in Cognitive Sciences*, 12(12):474–480.
- Grossberg, S. (1977). Pattern formation by the global limits of a nonlinear competitive interaction in n dimensions. *Journal of Mathematical biology*, 4(3):237–256.
- Grossberg, S. (1978). A theory of visual coding, memory, and development. *Formal theories of visual perception*, pages 7–26.
- Grossberg, S. (1982). How does a brain build a cognitive code? In *Studies of mind and brain*, pages 1–52. Springer.
- Grossman, M., Anderson, C., Khan, A., Avants, B., Elman, L., and McCluskey, L. (2008). Impaired action knowledge in amyotrophic lateral sclerosis. *Neurology*, 71(18):1396–1401.
- Hagoort, P. (2005). On Broca, brain, and binding: a new framework. *Trends in Cognitive Sciences*, 9(9):416–423.
- Hagoort, P. (2013). MUC (Memory, Unification, Control) and beyond. *Frontiers in Psychology*, 4.
- Hagoort, P. and van Berkum, J. (2007). Beyond the sentence given. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 362(1481):801–811.
- Hahne, A. and Friederici, A. D. (2002). Differential task effects on semantic and syntactic processes as revealed by ERPs. *Cognitive Brain Research*, 13(3):339–356.
- Hamilton, A. C. and Martin, R. C. (2005). Dissociations among tasks involving inhibition: A single-case study. *Cognitive, Affective, & Behavioral Neuroscience*, 5(1):1–13.
- Hauk, O., Johnsrude, I., and Pulvermüller, F. (2004). Somatotopic Representation of Action Words in Human Motor and Premotor Cortex. *Neuron*, 41(2):301–307.
- Henderson, J. and Ferreira, F. (2013). *The interface of language, vision, and action: Eye movements and the visual world*. Psychology Press.
- Hickok, G. and Poeppel, D. (2004). Dorsal and ventral streams: a framework for understanding aspects of the functional anatomy of language. *Cognition*, 92(1-2):67–99.
- Hinaut, X. and Dominey, P. F. (2013). Real-Time Parallel Processing of Grammatical Structure in the Fronto-Striatal System: A Recurrent Network Simulation Study Using Reservoir Computing. *PLoS ONE*, 8(2):e52946.
- Hinaut, X., Lance, F., Droin, C., Petit, M., Pointeau, G., and Dominey, P. F. (2015). Corticostriatal response selection in sentence production: Insights from neural network simulation with reservoir computing. *Brain and Language*, 150:54–68.
- Hodges, J. R., Graham, N., and Patterson, K. (1995). Charting the progression in semantic dementia: Implications for the organisation of semantic memory. *Memory*, 3(3-4):463–495.
- Hodges, J. R., Patterson, K., Oxbury, S., and Funnell, E. (1992). Semantic dementia progressive fluent aphasia with temporal lobe atrophy. *Brain*, 115(6):1783–1806.

- Hopfield, J. J. (1982). Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the national academy of sciences*, 79(8):2554–2558.
- Huetting, F., Rommers, J., and Meyer, A. S. (2011). Using the visual world paradigm to study language processing: A review and critical evaluation. *Acta psychologica*, 137(2):151–171.
- Humphries, C., Binder, J. R., Medler, D. A., and Liebenthal, E. (2007). Time course of semantic processes during sentence comprehension: an fMRI study. *Neuroimage*, 36(3):924–932.
- Itti, L. and Arbib, M. A. (2006). Attention and the minimal subscene. In Arbib, M. A., editor, *Action to Language via the Mirror Neuron System*, pages 289–346. Cambridge University Press, Cambridge.
- Itti, L., Koch, C., and Niebur, E. (1998). A model of saliency-based visual attention for rapid scene analysis. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 20(11):1254 –1259.
- Jackendoff, R. (1997). *The architecture of the language faculty*. Number 28. MIT Press.
- Jackendoff, R. (2002). *Foundations of language: Brain, meaning, grammar, evolution*. Oxford University Press.
- January, D., Trueswell, J. C., and Thompson-Schill, S. L. (2009). Co-localization of Stroop and syntactic ambiguity resolution in Broca’s area: Implications for the neural basis of sentence processing. *Journal of Cognitive Neuroscience*, 21(12):2434–2444.
- Joshi, A. K. and Schabes, Y. (1997). Tree-adjoining grammars. In *Handbook of formal languages*, pages 69–123. Springer.
- Kahneman, D., Treisman, A., and Gibbs, B. J. (1992). The reviewing of object files: Object-specific integration of information. *Cognitive Psychology*, 24(2):175–219.
- Kamide, Y., Altmann, G. T., and Haywood, S. L. (2003). The time-course of prediction in incremental sentence processing: Evidence from anticipatory eye movements. *Journal of Memory and Language*, 49(1):133–156.
- Kay, P. and Fillmore, C. J. (1999). Grammatical Constructions and Linguistic Generalizations: the What’s X doing Y? Construction.
- Kean, M.-L. (1977). The linguistic interpretation of aphasic syndromes: Agrammatism in Broca’s aphasia, an example. *Cognition*, 5(1):9–46.
- Keller, S. S., Crow, T., Foundas, A., Amunts, K., and Roberts, N. (2009). Broca’s area: Nomenclature, anatomy, typology and asymmetry. *Brain and Language*, 109(1):29–48.
- Kemmerer, D. (2000a). Grammatically relevant and grammatically irrelevant features of verb meaning can be independently impaired. *Aphasiology*, 14(10):997–1020.
- Kemmerer, D. (2000b). Selective impairment of knowledge underlying prenominal adjective order: evidence for the autonomy of grammatical semantics. *Journal of Neurolinguistics*, 13(1):57–82.
- Kemmerer, D. (2003). Why can you hit someone on the arm but not break someone on the arm?—a neuropsychological investigation of the English body-part possessor ascension construction. *Journal of Neurolinguistics*, 16(1):13–36.
- Kemmerer, D. and Gonzalez-Castillo, J. (2010). The Two-Level Theory of verb meaning: An approach to integrating the semantics of action with the mirror neuron system. *Brain and Language*, 112(1):54–76.
- Kemmerer, D., Miller, L., and Tranel, D. (2013). An investigation of semantic similarity judgments about action and non-action verbs in Parkinson’s disease: implications for the Embodied Cognition Framework. *Frontiers in Human Neuroscience*, 7:146.

- Kemmerer, D., Tranel, D., and Zdanczyk, C. (2009). Knowledge of the Semantic Constraints on Adjective Order Can Be Selectively Impaired. *Journal of Neurolinguistics*, 22(1):91–108.
- Kemmerer, D. and Wright, S. (2002). Selective impairment of knowledge underlying un- prefixation: further evidence for the autonomy of grammatical semantics. *Journal of Neurolinguistics*, 15(3-5):403–432. WOS:000174002300011.
- Kempen, G. (2014). Prolegomena to a neurocomputational architecture for human grammatical encoding and decoding. *Neuroinformatics*, 12(1):111–142.
- Kempen, G., Olsthoorn, N., and Sprenger, S. (2012). Grammatical workspace sharing during language production and language comprehension: Evidence from grammatical multitasking. *Language and Cognitive Processes*, 27(3):345–380.
- Kim, A. and Osterhout, L. (2005). The independence of combinatory semantic processing: Evidence from event-related potentials. *Journal of Memory and Language*, 52(2):205–225. WOS:000227044100003.
- Knight, K. (1989). Unification: A multidisciplinary survey. *ACM Computing Surveys (CSUR)*, 21(1):93–124.
- Knoeferle, P. (2016). Characterising visual context effects. *Visually Situated Language Comprehension*, 93:227.
- Knoeferle, P. and Crocker, M. W. (2006). The Coordinated Interplay of Scene, Utterance, and World Knowledge: Evidence From Eye Tracking. *Cognitive Science*, 30(3):481–529.
- Knoeferle, P. and Crocker, M. W. (2008). The coordinated processing of scene and utterance: evidence from eye tracking. In *Advances in Cognitive Science*.
- Knoeferle, P., Crocker, M. W., Scheepers, C., and Pickering, M. J. (2005). The influence of the immediate visual context on incremental thematic role-assignment: evidence from eye-movements in depicted events. *Cognition*, 95(1):95–127.
- Knoeferle, P., Pyykkönen-Klauck, P., and Crocker, M. W. (2016). *Visually Situated Language Comprehension*. John Benjamins Publishing Company.
- Koch, C. and Ullman, S. (1987). Shifts in selective visual attention: towards the underlying neural circuitry. In *Matters of Intelligence*, pages 115–141. Springer.
- Koechlin, E., Ody, C., and Kouneiher, F. (2003). The architecture of cognitive control in the human prefrontal cortex. *Science (New York, N.Y.)*, 302(5648):1181–1185.
- Kos, M., Vosse, T., Brink, D. v. d., and Hagoort, P. (2010). About edible restaurants: conflicts between syntax and semantics as revealed by ERPs. *Frontiers in Language Sciences*, 1:222.
- Kouneiher, F., Charron, S., and Koechlin, E. (2009). Motivation and cognitive control in the human prefrontal cortex. *Nature Neuroscience*, 12(7):939.
- Kuchinsky, S. E. (2009). *From Seeing to Saying: Perceiving, Planning, Producing*. ERIC.
- Kudo, T. (1984). The effect of semantic plausibility on sentence comprehension in aphasia. *Brain and language*, 21(2):208–218.
- Kukona, A. and Tabor, W. (2011). Impulse processing: A dynamical systems model of incremental eye movements in the visual world paradigm. *Cognitive science*, 35(6):1009–1051.
- Kulkarni, R., Rothstein, S., and Treves, A. (2016). A Neural Network Perspective on the Syntactic-Semantic Association between Mass and Count Nouns. *Journal of Advances in Linguistics*, 6(2).
- Kuperberg, G. R. (2007). Neural mechanisms of language comprehension: challenges to syntax. *Brain Research*, 1146:23–49.

- Kuperberg, G. R., Kreher, D. A., Sitnikova, T., Caplan, D. N., and Holcomb, P. J. (2007). The role of animacy and thematic relationships in processing active English sentences: Evidence from event-related potentials. *Brain and Language*, 100(3):223–237. WOS:000245828100001.
- Langacker, R. W. (1986). An Introduction to Cognitive Grammar. *Cognitive Science*, 10(1):1–40.
- Lashley, K. S. (1951). The problem of serial order in behavior.
- Lee, J. (2012). *Linking eyes to mouth: a schema-based computational model for describing visual scenes*. PhD thesis, University of Southern California.
- Leonard, C. M., Puranik, C., Kuldau, J. M., and Lombardino, L. J. (1998). Normal variation in the frequency and location of human auditory cortex landmarks. Heschl's gyrus: where is it? *Cerebral Cortex*, 8(5):397–406.
- Levelt, W. J. (1993). *Speaking: From intention to articulation*, volume 1. MIT press.
- Levin, B. (1993). *English verb classes and alternations: a preliminary investigation*. University of Chicago Press.
- Linebarger, M. C., Schwartz, M. F., and Saffran, E. M. (1983). Sensitivity to grammatical structure in so-called agrammatic aphasics. *Cognition*, 13(3):361–392.
- Luria, A. (1974). Language and brain: Towards the basic problems of neurolinguistics. *Brain and Language*, 1(1):1–14.
- Lyons, D. M. and Arbib, M. A. (1989). A formal model of computation for sensory-based robotics. *Robotics and Automation, IEEE Transactions on*, 5(3):280–293.
- Lyttelton, O., Boucher, M., Robbins, S., and Evans, A. (2007). An unbiased iterative group registration template for cortical surface analysis. *NeuroImage*, 34(4):1535–1544.
- MacKay, D. G., Burke, D. M., and Stewart, R. (1998). H.M.'s Language Production Deficits: Implications for Relations between Memory, Semantic Binding, and the Hippocampal System. *Journal of Memory and Language*, 38(1):28–69.
- MacKay, D. G., James, L. E., Taylor, J. K., and Marian, D. E. (2007). Amnesic H.M. exhibits parallel deficits and sparing in language and memory: Systems versus binding theory accounts. *Language and Cognitive Processes*, 22(3):377–452.
- Mackay, D. G., Stewart, R., and Burke, D. M. (1998). H.M. Revisited: Relations between Language Comprehension, Memory, and the Hippocampal System. *J. Cognitive Neuroscience*, 10(3):377–394.
- Mahon, B. Z. and Caramazza, A. (2005). The orchestration of the sensory-motor systems: Clues from neuropsychology. *Cognitive Neuropsychology*, 22(3-4):480–494.
- Mahon, B. Z. and Caramazza, A. (2008). A critical look at the embodied cognition hypothesis and a new proposal for grounding conceptual content. *Journal of Physiology-Paris*, 102(1-3):59–70. WOS:000258012400008.
- Makkai, A. (1972). Idiom structure in English. Technical report.
- Makuuchi, M., Bahlmann, J., Anwander, A., and Friederici, A. D. (2009). Segregating the core computational faculty of human language from working memory. *Proceedings of the National Academy of Sciences*, 106(20):8362–8367.
- Mange, D. and Tomassini, M. (1998). *Bio-inspired computing machines: Towards novel computational architectures*. PPUR presses polytechniques.
- Marr, D. (1982). *Vision: A computational approach*. Freeman & Co., San Francisco.

- Mayberry, M., Crocker, M. W., and Knoeferle, P. (2006). A Connectionist Model of the Coordinated Interplay of Scene, Utterance, and World Knowledge.
- McClelland, J. L. (1993). Toward a theory of information processing in graded, random, and interactive networks.
- McClelland, J. L., Botvinick, M. M., Noelle, D. C., Plaut, D. C., Rogers, T. T., Seidenberg, M. S., and Smith, L. B. (2010). Letting structure emerge: connectionist and dynamical systems approaches to cognition. *Trends in cognitive sciences*, 14(8):348–356.
- McClelland, J. L. and Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception: I. An account of basic findings. *Psychological review*, 88(5):375.
- Menenti, L., Gierhan, S. M. E., Segaert, K., and Hagoort, P. (2011). Shared Language Overlap and Segregation of the Neuronal Infrastructure for Speaking and Listening Revealed by Functional MRI. *Psychological Science*, 22(9):1173–1182.
- Menenti, L., Pickering, M. J., and Garrod, S. C. (2012). Toward a neural basis of interactive alignment in conversation. *Frontiers in Human Neuroscience*, 6.
- Meteyard, L., Cuadrado, S. R., Bahrami, B., and Vigliocco, G. (2010). Coming of age: A review of embodiment and the neuroscience of semantics. *Cortex; a Journal Devoted to the Study of the Nervous System and Behavior*.
- Metusalem, R., Kutas, M., Urbach, T. P., Hare, M., McRae, K., and Elman, J. L. (2012). Generalized event knowledge activation during online sentence comprehension. *Journal of Memory and Language*, 66(4):545–567.
- Miller, E. K. and Cohen, J. D. (2001). An integrative theory of prefrontal cortex function. *Annual review of neuroscience*, 24(1):167–202.
- Minsky, M. (1974). A Framework for Representing Knowledge.
- Miyake, A., Carpenter, P. A., and Just, M. A. (1994). A capacity approach to syntactic comprehension disorders: making normal adults perform like aphasic patients. *Cognitive Neuropsychology*, 11(6):671–717.
- Miyake, A., Carpenter, P. A., and Just, M. A. (1995). Reduced resources and specific impairments in normal and aphasic sentence comprehension. *Cognitive Neuropsychology*, 12(6):651–679.
- Mohanan, T. and Wee, L. (1999). *Grammatical semantics: Evidence for structure in meaning*. CSLI.
- Mohr, J. P. (1976). Broca's area and Broca's aphasia. *Stud Neurolinguis*, 1:201–236.
- Mohr, J. P., Pessin, M. S., Finkelstein, S., Funkenstein, H. H., Duncan, G. W., and Davis, K. R. (1978). Broca aphasia Pathologic and clinical. *Neurology*, 28(4):311–311.
- Moore, R. C. (2000). Improved left-corner chart parsing for large context-free grammars. In *Proceedings of the Sixth International Workshop on Parsing Technologies*, pages 171–182.
- Myachykov, A., Thompson, D., Scheepers, C., and Garrod, S. (2011). Visual Attention and Structural Choice in Sentence Production Across Languages. *Language and Linguistics Compass*, 5(2):95–107.
- Narayanan, S. (1999). Moving Right Along: A Computational Model of Metaphoric Reasoning about Events. Technical report, CiteSeerX.
- Navalpakkam, V. and Itti, L. (2005). Modeling the influence of task on attention. *Vision Research*, 45(2):205–231.
- Negri, G. A., Rumianti, R. I., Zadini, A., Ukmar, M., Mahon, B. Z., and Caramazza, A. (2007). What is the role of motor simulation in action and object recognition? Evidence from apraxia. *Cognitive Neuropsychology*, 24(8):795–816.

- Nieuwland, M.S. and Berkum, J.J.A. (2005). Testing the limits of the semantic illusion phenomenon: ERPs reveal temporary semantic change deafness in discourse comprehension. *Cognitive Brain Research*, 24(3):691–701.
- Novick, J. M., Kan, I. P., Trueswell, J. C., and Thompson-Schill, S. L. (2009). A case for conflict across multiple domains: memory and language impairments following damage to ventrolateral prefrontal cortex. *Cognitive Neuropsychology*, 26(6):527–567.
- Novick, J. M., Trueswell, J. C., and Thompson-Schill, S. L. (2005). Cognitive control and parsing: Reexamining the role of Broca's area in sentence comprehension. *Cognitive, Affective, & Behavioral Neuroscience*, 5(3):263–281.
- Novick, J. M., Trueswell, J. C., and Thompson-Schill, S. L. (2010). Broca's area and language processing: Evidence for the cognitive control connection. *Language and Linguistics Compass*, 4(10):906–924.
- Oostenveld, R., Fries, P., Maris, E., and Schoffelen, J.-M. (2011). FieldTrip: Open Source Software for Advanced Analysis of MEG, EEG, and Invasive Electrophysiological Data. *Computational Intelligence and Neuroscience*, 2011:1–9.
- Orban, G. A., Van Essen, D., and Vanduffel, W. (2004). Comparative mapping of higher visual areas in monkeys and humans. *Trends in cognitive sciences*, 8(7):315–324.
- O'Regan, J. K. (1992). Solving the "real" mysteries of visual perception: The world as an outside memory. *Canadian Journal of Psychology/Revue Canadienne de Psychologie*, 46(3):461–488.
- Osterhout, L., Kim, A., and Kuperberg, G. (2007). The neurobiology of sentence comprehension. Technical report, CiteSeerX.
- Oztop, E. and Arbib, M. A. (2002). Schema design and implementation of the grasp-related mirror neuron system. *Biological Cybernetics*, 87(2):116–140.
- Papeo, L., Negri, G. A. L., Zadini, A., and Ida Rumia, R. (2010). Action performance and action-word understanding: Evidence of double dissociations in left-damaged patients. *Cognitive Neuropsychology*, 27(5):428–461.
- Patson, N. D., Darowski, E. S., Moon, N., and Ferreira, F. (2009). Lingering Misinterpretations in Garden-Path Sentences: Evidence From a Paraphrasing Task. *Journal of Experimental Psychology-Learning Memory and Cognition*, 35(1):280–285. WOS:000262095400021.
- Patterson, K., Nestor, P. J., and Rogers, T. T. (2007). Where do you know what you know? The representation of semantic knowledge in the human brain. *Nature Reviews Neuroscience*, 8(12):976–987.
- Paul, H. (1970). The sentence as the expression of the combination of several ideas. *AL Blumenthal (Ed. and Trans.), Language and psychology: Historical aspects of psycholinguistics*, pages 34–37.
- Piaget, J. (1965). The stages of the intellectual development of the child. *Educational psychology in context: Readings for future teachers*, pages 98–106.
- Pickering, M. J. and Garrod, S. (2007). Do people use language production to make predictions during comprehension? *Trends in Cognitive Sciences*, 11(3):105–110.
- Piñango, M. M. (2006). Understanding the architecture of language: the possible role of neurology. *Trends in Cognitive Sciences*, 10(2):49–51.
- Piñango, M. M. and Zurif, E. B. (2001). Semantic operations in aphasic comprehension: implications for the cortical organization of language. *Brain and Language*, 79(2):297–308.
- Pinker, S. (1989). *Learnability and cognition: The acquisition of argument structure*, volume xiv of *Learning, development, and conceptual change*. The MIT Press, Cambridge, MA, US.
- Pinker, S. (2000). *Words and Rules: The Ingredients of Language*. HarperCollins.

- Pirmoradian, S. and Treves, A. (2011). BLISS: an artificial language for learnability studies. *Cognitive Computation*, 3(4):539–553.
- Pirmoradian, S. and Treves, A. (2012). A talkative Potts attractor neural network welcomes BLISS words. *BMC Neuroscience*, 13(Suppl 1):P21.
- Pirmoradian, S., Treves, A., and SISSA, C. N. S. (2013). Encoding words into a Potts attractor network. In *Proceedings of the thirteenth neural computation and psychology workshop (ncpw13) on computational models of cognitive processes, world scientific press, singapore*, pages 29–42.
- Pollard, C. and Sag, I. A. (1994). *Head-driven phrase structure grammar*. University of Chicago Press.
- Pulvermüller, F. (2005). Brain mechanisms linking language and action. *Nature Reviews Neuroscience*, 6(7):576–582.
- Pulvermüller, F. (2013a). How neurons make meaning: brain mechanisms for embodied and abstract-symbolic semantics. *Trends in Cognitive Sciences*, 17(9):458–470.
- Pulvermüller, F. (2013b). Semantic embodiment, disembodiment or misembodiment? In search of meaning in modules and neuron circuits. *Brain and language*.
- Pulvermüller, F. and Knoblauch, A. (2009). Discrete combinatorial circuits emerging in neural networks: a mechanism for rules of grammar in the human brain? *Neural networks*, 22(2):161–172.
- Pylkkänen, L. and McElree, B. (2007). An MEG study of silent meaning. *Journal of Cognitive Neuroscience*, 19(11):1905–1921.
- Pylkkanen, L., Oliveri, B., and Smart, A. J. (2009). Semantics vs. world knowledge in prefrontal cortex. *Language and Cognitive Processes*, 24(9):1313–1334. WOS:000271904800001.
- Pylyshyn, Z. W. (2000). Situating vision in the world. *Trends in Cognitive Sciences*, 4(5):197–207.
- Pylyshyn, Z. W. (2001). Visual indexes, preconceptual objects, and situated vision. *Cognition*, 80(1–2):127–158.
- Quigley, M., Conley, K., Gerkey, B., Faust, J., Foote, T., Leibs, J., Wheeler, R., and Ng, A. Y. (2009). ROS: an open-source Robot Operating System. In *ICRA workshop on open source software*, volume 3, page 5. Kobe, Japan.
- Rao, R. P., Zelinsky, G. J., Hayhoe, M. M., and Ballard, D. H. (2002). Eye movements in iconic visual search. *Vision research*, 42(11):1447–1463.
- Raposo, A., Moss, H. E., Stamatakis, E. A., and Tyler, L. K. (2009). Modulation of motor and premotor cortices by actions, action words and action sentences. *Neuropsychologia*, 47(2):388–396.
- Rauschecker, J. P. and Scott, S. K. (2009). Maps and streams in the auditory cortex: nonhuman primates illuminate human speech processing. *Nature Neuroscience*, 12(6):718–724.
- Rensink, R. A. (2000). The Dynamic Representation of Scenes. *Visual Cognition*, 7(1-3):17–42.
- Resnik, P. (1992). Left-corner parsing and psychological plausibility. In *Proceedings of the 14th conference on Computational linguistics-Volume 1*, pages 191–197. Association for Computational Linguistics.
- Robinson, G., Blair, J., and Cipolotti, L. (1998). Dynamic aphasia: an inability to select between competing verbal responses? *Brain*, 121(1):77–89.
- Robinson, G., Shallice, T., and Cipolotti, L. (2005). A failure of high level verbal response selection in progressive dynamic aphasia. *Cognitive Neuropsychology*, 22(6):661–694.
- Rodd, J. M., Davis, M. H., and Johnsrude, I. S. (2005). The neural mechanisms of speech comprehension: fMRI studies of semantic ambiguity. *Cerebral Cortex*, 15(8):1261–1269.

- Rogalsky, C., Matchin, W., and Hickok, G. (2008). Broca's area, sentence comprehension, and working memory: an fMRI study. *Frontiers in human neuroscience*, 2.
- Rogers, T. T., Lambon Ralph, M. A., Garrard, P., Bozeat, S., McClelland, J. L., Hodges, J. R., and Patterson, K. (2004a). Structure and Deterioration of Semantic Memory: A Neuropsychological and Computational Investigation. *Psychological Review*, 111(1):205–235.
- Rogers, T. T., Lambon Ralph, M. A., Hodges, J. R., and Patterson, K. (2004b). Natural selection: The impact of semantic impairment on lexical and object decision. *Cognitive Neuropsychology*, 21(2-4):331–352.
- Rogers, T. T. and Patterson, K. (2007). Object categorization: reversals and explanations of the basic-level advantage. *Journal of Experimental Psychology: General*, 136(3):451.
- Rosch, E. and Mervis, C. B. (1975). Family resemblances: Studies in the internal structure of categories. *Cognitive Psychology*, 7(4):573–605.
- Rosch, E., Mervis, C. B., Gray, W. D., Johnson, D. M., and Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive Psychology*, 8(3):382–439.
- Rosenkrantz, D. J. and Lewis, P. M. (1970). Deterministic left corner parsing. In *Switching and Automata Theory, 1970., IEEE Conference Record of 11th Annual Symposium on*, pages 139–152. IEEE.
- Rothi, L. J., Ochiai, T., and Heilman, K. M. (1991). A Cognitive Neuropsychological Model of Limp Praxis. *Cognitive Neuropsychology*, 8(6):443–458. WOS:A1991GW56100003.
- Rumelhart, D. E. and McClelland, J. L. (1986). Parallel distributed processing: explorations in the microstructure of cognition. Volume 1. Foundations.
- Russo, E., Pirmoradian, S., and Treves, A. (2011). Associative latching dynamics vs. syntax. In *Advances in Cognitive Neurodynamics (II)*, pages 111–115. Springer Netherlands.
- Russo, E. and Treves, A. (2011). An uncouth approach to language recursivity. *Biolinguistics*, 5(1-2):133–150.
- Russo, E. and Treves, A. (2012). Cortical free-association dynamics: Distinct phases of a latching network. *Physical Review E*, 85(5):051920.
- Saffran, E. M., Schwartz, M. F., and Linebarger, M. C. (1998). Semantic Influences on Thematic Role Assignment: Evidence from Normals and Aphasics. *Brain and Language*, 62(2):255–297.
- Sanchez, E., Mange, D., Sipper, M., Tomassini, M., Pérez-Uribe, A., and Stauffer, A. (1997). Phylogeny, ontogeny, and epigenesis: Three sources of biological inspiration for softening hardware. *Evolvable Systems: From Biology to Hardware*, pages 33–54.
- Sanides, F. (1964). Structure and function of the human frontal lobe. *Neuropsychologia*, 2(3):209–219.
- Santi, A. and Grodzinsky, Y. (2007a). Taxing working memory with syntax: Bihemispheric modulations. *Human brain mapping*, 28(11):1089–1097.
- Santi, A. and Grodzinsky, Y. (2007b). Working memory and syntax interact in Broca's area. *NeuroImage*, 37(1):8–17.
- Santi, A. and Grodzinsky, Y. (2010). fMRI adaptation dissociates syntactic complexity dimensions. *NeuroImage*, 51(4):1285–1293.
- Sato, M., Schafer, A. J., and Bergen, B. (2013). One word at a time: Mental representations of object shape change incrementally during sentence processing.
- Saur, D., Kreher, B. W., Schnell, S., Kümmeler, D., Kellmeyer, P., Vry, M.-S., Umarova, R., Musso, M., Glauche, V., Abel, S., Huber, W., Rijntjes, M., Hennig, J., and Weiller, C. (2008). Ventral and dorsal pathways for language. *Proceedings of the National Academy of Sciences of the United States of America*, 105(46):18035–18040.

- Schnur, T. T., Schwartz, M. F., Kimberg, D. Y., Hirshorn, E., Coslett, H. B., and Thompson-Schill, S. L. (2009). Localizing interference during naming: Convergent neuroimaging and neuropsychological evidence for the function of Broca's area. *Proceedings of the National Academy of Sciences*, 106(1):322–327.
- Schwartz, M. F., Linebarger, M. C., Saffran, E. M., and Pate, D. S. (1987). Syntactic transparency and sentence interpretation in aphasia. *Language and Cognitive Processes*, 2(2):85–113.
- Sedivy, J. C., K Tanenhaus, M., Chambers, C. G., and Carlson, G. N. (1999). Achieving incremental semantic interpretation through contextual representation. *Cognition*, 71(2):109–147.
- Segaert, K., Menenti, L., Weber, K., Petersson, K. M., and Hagoort, P. (2012). Shared Syntax in Language Production and Language Comprehension—An fMRI Study. *Cerebral Cortex*, 22(7):1662–1670.
- Sherman, J. C. and Schweickert, J. (1989). Syntactic and semantic contributions to sentence comprehension in agrammatism. *Brain and Language*, 37(3):419–439.
- Shieber, S. M. (1986). *An introduction to unification-based approaches to grammar*. Microtome Publishing.
- Shieber, S. M. (2003). *An introduction to unification-based approaches to grammar*. Microtome Publishing.
- Simon, H. A. (1962). The architecture of complexity. *Proceedings of the American philosophical society*, 106(6):467–482.
- Simon, H. A. (1972). Theories of bounded rationality. *Decision and organization*, 1(1):161–176.
- Simon, H. A. (1977). The organization of complex systems. In *Models of discovery*, pages 245–261. Springer.
- Sipper, M., Sanchez, E., Mange, D., Tomassini, M., Pérez-Uribe, A., and Stauffer, A. (1997). A phylogenetic, ontogenetic, and epigenetic view of bio-inspired hardware systems. *IEEE Transactions on Evolutionary Computation*, 1(1):83–97.
- Skotko, B. G., Andrews, E., and Einstein, G. (2005). Language and the medial temporal lobe: Evidence from H.M.'s spontaneous discourse. *Journal of Memory and Language*, 53(3):397–415.
- Slobin, D. I. (1996). From “thought and language” to “thinking for speaking”. *Rethinking linguistic relativity*, 17:70–96.
- Smith, E. E. and Jonides, J. (1997). Working memory: A view from neuroimaging. *Cognitive psychology*, 33(1):5–42.
- Smith, E. E., Jonides, J., and Koeppe, R. A. (1996). Dissociating verbal and spatial working memory using PET. *Cerebral Cortex*, 6(1):11–20.
- Smolensky, P. (1990). Tensor product variable binding and the representation of symbolic structures in connectionist systems. *Artificial intelligence*, 46(1-2):159–216.
- Spranger, M., Pauw, S., Loetzsch, M., and Steels, L. (2012). Open-ended procedural semantics. In *Language grounding in robots*, pages 153–172. Springer.
- Spranger, M. and Steels, L. (2015). Co-Acquisition of Syntax and Semantics—An Investigation in Spatial Language. IJCAI.
- Stanfield, R. A. and Zwaan, R. A. (2001). The Effect of Implied Orientation Derived from Verbal Context on Picture Recognition. *Psychological Science*, 12(2):153–156.
- Steels, L. (1999). The talking heads experiment.
- Steels, L. (2011). *Design patterns in fluid construction grammar*, volume 11. John Benjamins.
- Steels, L. and De Beule, J. (2006). Unify and merge in fluid construction grammar. In Vogt, P., Sugita, Y., Tuci, E., and Nehaniv, C., editors, *Symbol Grounding and Beyond, Proceedings*, volume 4211, pages 197–223. Springer-Verlag Berlin, Berlin. WOS:000242307400016.

- Steenstrup, M., Arbib, M. A., and Manes, E. G. (1983). Port automata and the algebra of concurrent processes. *Journal of Computer and System Sciences*, 27(1):29–50.
- Svantner, J., Farka\vs, I., and Crocker, M. (2012). Modeling utterance-driven visual attention during situated comprehension. *Neural Network World*, 22(2):85.
- Tabor, W. and Hutchins, S. (2004). Evidence for self-organized sentence processing: digging-in effects. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 30(2):431.
- Tadel, F., Baillet, S., Mosher, J. C., Pantazis, D., and Leahy, R. M. (2011). Brainstorm: A User-Friendly Application for MEG/EEG Analysis. *Computational Intelligence and Neuroscience*, 2011.
- Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., and Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science (New York, N.Y.)*, 268(5217):1632–1634.
- Tettamanti, M., Manenti, R., Della Rosa, P. A., Falini, A., Perani, D., Cappa, S. F., and Moro, A. (2008). Negation in the brain: modulating action representations. *Neuroimage*, 43(2):358–367.
- Thompson-Schill, S. L. (2003). Neuroimaging studies of semantic memory: inferring “how” from “where”. *Neuropsychologia*, 41(3):280–292.
- Thompson-Schill, S. L., Jonides, J., Marshuetz, C., Smith, E. E., D’Esposito, M., Kan, I. P., Knight, R. T., and Swick, D. (2002). Effects of frontal lobe damage on interference effects in working memory. *Cognitive, Affective, & Behavioral Neuroscience*, 2(2):109–120.
- Tomasello, M. (2009). *Constructing a language: A usage-based theory of language acquisition*. Harvard University Press.
- Tomasino, B., Weiss, P. H., and Fink, G. R. (2010). To move or not to move: imperatives modulate action-related verb processing in the motor system. *Neuroscience*, 169(1):246–258.
- Tomuro, N. (1999). *Left-corner parsing algorithm for unification grammars*. PhD thesis, DePaul University.
- Tourville, J. A. and Guenther, F. H. (2011). The DIVA model: A neural theory of speech acquisition and production. *Language and cognitive processes*, 26(7):952–981.
- Tranel, D., Damasio, H., and Damasio, A. R. (1997). A neural basis for the retrieval of conceptual knowledge. *Neuropsychologia*, 35(10):1319–1327.
- Treisman, A. M. and Gelade, G. (1980). A feature-integration theory of attention. *Cognitive psychology*, 12(1):97–136.
- Treves, A. (2005). Frontal latching networks: a possible neural basis for infinite recursion. *Cognitive Neuropsychology*, 22(3-4):276–291.
- Triesch, J., Ballard, D. H., Hayhoe, M. M., and Sullivan, B. T. (2003). What you see is what you need. *Journal of vision*, 3(1).
- Trueswell, J. C., Sekerina, I., Hill, N. M., and Logrip, M. L. (1999). The kindergarten-path effect: Studying on-line sentence processing in young children. *Cognition*, 73(2):89–134.
- Tsetsos, K., Usher, M., and McClelland, J. L. (2011). Testing multi-alternative decision models with non-stationary evidence. *Frontiers in neuroscience*, 5:63.
- Ueno, T., Saito, S., Rogers, T. T., and Ralph, M. A. L. (2011). Lichtheim 2: Synthesizing Aphasia and the Neural Basis of Language in a Neurocomputational Model of the Dual Dorsal-Ventral Language Pathways. *Neuron*, 72(2):385–396. WOS:000296224000018.
- Usher, M. and McClelland, J. L. (2001). The time course of perceptual choice: the leaky, competing accumulator model. *Psychological review*, 108(3):550.

- Usher, M. and McClelland, J. L. (2004). Loss aversion and inhibition in dynamical models of multialternative choice. *Psychological review*, 111(3):757.
- Van Eecke, P. and Beuls, K. (2017). Meta-Layer Problem Solving for Computational Construction Grammar.
- Van Essen, D. C. (2005). A Population-Average, Landmark- and Surface-based (PALS) atlas of human cerebral cortex. *NeuroImage*, 28(3):635–662.
- van Herten, M., Kolk, H. H., and Chwilla, D. J. (2005). An ERP study of P600 effects elicited by semantic anomalies. *Cognitive Brain Research*, 22(2):241–255.
- Van Trijp, R., Steels, L., Beuls, K., and Wellens, P. (2012). Fluid construction grammar: The new kid on the block. In *Proceedings of the Demonstrations at the 13th Conference of the European Chapter of the Association for Computational Linguistics*, pages 63–68.
- Vann, S. D., Aggleton, J. P., and Maguire, E. A. (2009). What does the retrosplenial cortex do? *Nat Rev Neurosci*, 10(11):792–802.
- Vigliocco, G., Vinson, D. P., Lewis, W., and Garrett, M. F. (2004). Representing the meanings of object and action words: The featural and unitary semantic space hypothesis. *Cognitive Psychology*, 48(4):422–488.
- Vosse, T. and Kempen, G. (2000). Syntactic structure assembly in human parsing: a computational model based on competitive inhibition and a lexicalist grammar. *Cognition*, 75(2):105–143.
- Vosse, T. and Kempen, G. (2009). In defense of competition during syntactic ambiguity resolution. *Journal of Psycholinguistic Research*, 38(1):1–9.
- Wellens, P. and Steels, L. (2011). Organizing constructions in networks. *Design Patterns in Fluid Construction Grammar*. John Benjamins, Amsterdam.
- Wilson, S. M., Dronkers, N. F., Ogar, J. M., Jang, J., Growdon, M. E., Agosta, F., Henry, M. L., Miller, B. L., and Gorno-Tempini, M. L. (2010a). Neural correlates of syntactic processing in the nonfluent variant of primary progressive aphasia. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, 30(50):16845–16854.
- Wilson, S. M., Henry, M. L., Besbris, M., Ogar, J. M., Dronkers, N. F., Jarrold, W., Miller, B. L., and Gorno-Tempini, M. L. (2010b). Connected speech production in three variants of primary progressive aphasia. *Brain: A Journal of Neurology*, 133(Pt 7):2069–2088.
- Wolfe, J. M. (1994). Guided search 2.0 a revised model of visual search. *Psychonomic bulletin & review*, 1(2):202–238.
- Wundt, W. (1970). The psychology of the sentence. *AL Blumenthal (Ed. and Trans.), Language and psychology: Historical aspects of psycholinguistics*, pages 20–31. Original work published 1900.
- Yarbus, A. L., Haigh, B., and Riggs, L. A. (1967). *Eye movements and vision*, volume 2. Plenum press New York.
- Ye, Z. and Zhou, X. (2009). Conflict control during sentence comprehension: fMRI evidence. *Neuroimage*, 48(1):280–290.
- Yvert, G., Perrone-Bertolotti, M., Baciu, M., and David, O. (2012). Dynamic Causal Modeling of Spatiotemporal Integration of Phonological and Semantic Processes: An Electroencephalographic Study. *The Journal of Neuroscience*, 32(12):4297–4306.
- Zurif, E. B. and Piñango, M. M. (1999). The Existence of Comprehension Patterns in Broca's Aphasia. *Brain and Language*, 70(1):133–138.
- Zwaan, R. A., Madden, C. J., Yaxley, R. H., and Aveyard, M. E. (2004). Moving words: dynamic representations in language comprehension. *Cognitive Science*, 28(4):611–619.
- Zwaan, R. A., Stanfield, R. A., and Yaxley, R. H. (2002). Language Comprehenders Mentally Represent the Shapes of Objects. *Psychological Science*, 13(2):168–171.

## Appendix A

# Cognitive Architecture Schema Theory (COAST): Formalism and Implementation

### A.1 Overview

This computational work rests on a Python implemented formalism for Schema Theory: (Cognitive Architecture Schema Theory (COAST)). In this framework, cognitive models are built as System-of-Systems (SoS). The systems have both a fixed architectural organization based on the nature of the connection network they form, and a dynamic functional architecture whose state evolves on the basis of self-organization principles based fueled mainly by cooperative-computation: A cognitive system is both organized (the brain is not a random graph of concurrent processes but is heavily shaped by phylogeny, ontogeny, and epigeny<sup>1</sup>), with a functional fluidity in how this architecture is contextually used to best fit the goals of its host organism. Each (sub-)system carries a state, a function, as well as input and output ports. At each time step, a system reads values from its input ports, performs operations based on its state and function, update its state and post values to its output ports. Connections always connect an output port to an input port and pass values from the input port to the output port. Beyond being a computational scheme to ensure distributed computation, COAST defines the core properties of key cognitive modeling abstraction: Long Term Memory systems (LTM), Working Memory systems (WM), and the related Schemas and Schema Instances. In doing so COAST offers a way to model processes using the type of systems hypothesized by ST while also offering the option to build new systems types when necessary.

### A.2 Model as System-of-Systems (SoS)

A model is defined as a System-of-Systems (SoS). Each model consists of set of systems and their connections forming a distributed functional system incrementally processing inputs received through input ports  $ports_{in}$  and generating outputs posted in output ports  $ports_{out}$ .

At each time step, a model updates its state in the following way:

Figure A.1 shows a simple example of a System-of-Systems composed of two sub-systems and a single connection.

When the model is updated at time  $t$  port  $p_1$  receives its value from the input  $p_1(t) = input(t)$ . The state of both procedural schemas are updated  $(p_2(t+1), S_1(t+1)) = \mathcal{F}_1(p_1(t), \vec{P}_1, S_1(t))$  and  $(p_4(t+1), S_2(t+1)) = \mathcal{F}_2(p_3(t), \vec{P}_2, S_2(t))$ . Data is passed through the connection  $p_3(t+1) = p_2(t)$ . Finally,  $output(t) = p_4(t)$ .

A specific activity value is defined for each sub-system separately from its state. An important aspect of schema theory consists in dealing with not only the definition of static network of sub-systems but with the

---

<sup>1</sup>See among other the work of (Mange and Tomassini, 1998; Sanchez et al., 1997; Sipper et al., 1997) for a discussion of the role of those three principles in designing bio-inspired computing machines

---

**Algorithm 1** Model state update

---

```

Step1: Check inputs.
for all  $p_i \in ports_{in}$  do
    Update  $p_i.value$  based on received inputs.
Step2: Update sub-systems' states
for all  $sub-system S_i$  do
     $S_i.update()$ 
Step3: Pass message through connections between sub-systems.
for all  $C_i \in connections$  do
     $C_i.update()$ 
Step4: Store or display output values
for all  $p_i \in ports_{out}$  do
    return  $p_i.value$ 

```

---

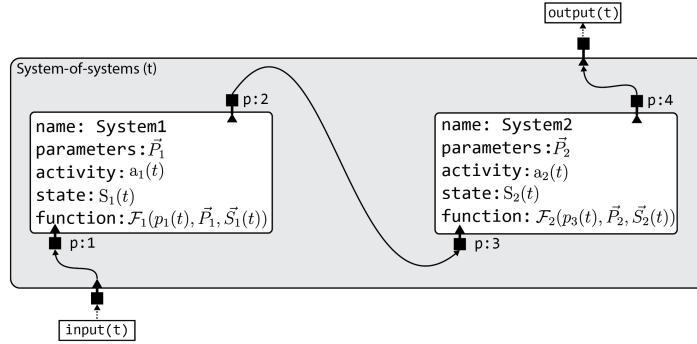


Figure A.1: Schema theory simple System-of-System (SoS) informal example. The SoS here is composed of two sub-systems: System1 and System2. Each have respectively one input and one output port (p:1 & p:2 for System1, p:3 & p:4 for System2). The whole SoS has receives inputs  $input(t)$  that are passed on to System1 and generates outputs  $output(t)$  received from System2. The SoS contains only one connection between the two sub-systems linking the output port of System1 to the input port of System2: the sub-systems are serially connected, but they are processing information asynchronously. Each sub-system  $i$  defines a set of parameters  $\vec{P}_i$ , a state  $S_i(t)$ , an activity level reflecting its relevance in the overall SoS process  $a_i(t)$ , and a function  $\mathcal{F}$  that maps the state, the parameters and the inputs to the next state.

possibility to design systems that flexibly respond to the task at hand, and do so by usually incorporating multiple sub-systems that overlap in their function and can therefore either compete or cooperate. In this context activity levels serve a specific purpose as indicators of the relevance of a given sub-system to the current process.

## A.3 Long Term Memory System

### A.3.1 Schema

Schemas formalize functional entities tied to and existing within a long term memory system (LTM)(see below). A schema carries content that represents an associated piece of knowledge (declarative or procedural). Different sub-classes of can be defined to account for qualitatively different types of knowledge and knowledge representation (e.g. perceptual knowledge, conceptual knowledge, ...).

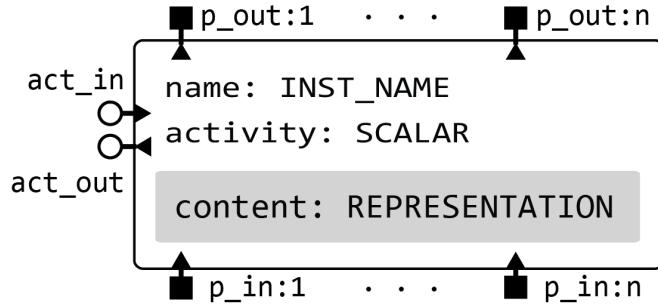


Figure A.2: Generic view of a schema instance. The content of each schema instance defines a hypothesis that can take the form of knowledge or processes (declarative or procedural content). At each time, the activity level of the instance reflects the relevance of its hypothesis to the current goal. Input and output ports allow the instance to receive and send information to the other instances it cooperates with. Activity ports are used to as input or output to cooperation-competition links (C2 links) formed with cooperating and competing instances (respectively). At each time-step, *act\_in* gathers the weighted inputs received from all the competing and cooperating instances which will be used to update the instance's activity level. *Act\_out* broadcasts this activity level.

### A.3.2 Long Term Memory as Schema network

A long term memory (LTM) system has for state a network of schemas. Taken together, the schema network of a given LTM formalizes the knowledge over a given domain. The LTM system can update the state and select schemas.  $LTM_{state} = (S, C)$  where  $S$  is a set of schemas all defining a same type of knowledge representations, and  $C \in S \times S$ , modeling connections between schemas.

### A.3.3 Schema instance & Schema instantiation

A LTM is always linked to a Working Memory (WM) in which the knowledge stored in the LTM is put to use. Once a schema is deemed relevant to current state of the computation, it is invoked in WM in the form a schema instance. Each schema instance can be considered to represent a hypothesis offering a partial solution to the problem the WM attempts to solve. Each schema instance is always associated with an activation value that represents, at each time step, the degree of confidence associated with this hypothesis.

## A.4 Working Memory & Cooperative Computation

### A.4.1 Overview

Working memories are systems that process relevant knowledge information (schema instances) retrieved from the associated LTM. Cooperative computation (C2) fuels this process. Instances compete and cooperate, boosting and inhibiting their respective activation values. At time step, cooperating instances form assemblages, each corresponding to a potential composition of instances that would provide a processing solution at this time. The precise process through which instance form cooperation link (*coop\_link*) or competition link (*comp\_link*), is specific to each WM sub-system but the ST design philosophy is that instances that correspond to hypotheses or process that support each-other create *coop\_links* while those that correspond to contradictory hypotheses or processes form *comp\_links*. At each time the whole set of *coop\_links* and *comp\_links* (C2links), and instances form a C2 network that governs the dynamics of the activation value associated with each active instance.

The Figure A.3 below provides an example C2 state of a WM. The example represents a simplified version of the VISIONS processes of object recognition.

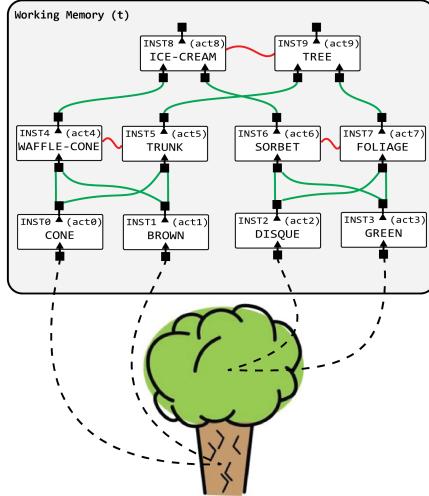


Figure A.3: Example of C2 of construction instances in WM adapted from VISIONS. (Bottom) An ambiguous visual object. (Top) A view of the state of the WM at time  $t$ . Green: cooperation links. Red: Competition links. At the bottom, perceptual schema instance present hypotheses regarding low-level perceptual features of image regions (color, shape). Those features can collaborate and serve as basis for object part level instances to propose hypotheses regarding the perceptual identity of those regions. When those instances carry incompatible hypotheses (SORBET vs. FOLIAGE), they enter in competition. Finally, entity level instances proposes hypotheses regarding the identity of the visual object based on the collaborative input they receive from part level instances. Here again, the two instances are in competition (TREE vs. ICE-CREAM). The process was described here bottom-up but takes place simultaneously in a top-down fashion with higher level instances reinforcing lower level instances with which they cooperate.

#### A.4.2 C2\_links

Functional links represents the functional connections that are established between instances within a working memory as they enter in cooperative computation. As opposed to *connect* which are fixed in the model and can in theory be mapped onto biological connections within the nervous system, *C2\_links* can be dynamically created and removed between schema instances within a working memory, dynamically shaping the competition-cooperation network. Functional links are of two types: cooperation links (*coop\_link*) and competition links (*comp\_link*). *C2\_links* are defined as directed connections:

$$C2\_link : (instance\_from, instance\_to, weight, asymmetry\_coef, f : (weight, t, P) \rightarrow weight) \quad (A.1)$$

With  $weight \in \mathbb{R}$  (positive weight for cooperation links, negative weight for competition links).

$asymmetry\_coef \in [0, 1]$  defines how bidirectional the connection is (0: fully bidirectional, 1: only from  $inst\_from$  to  $inst\_to$ ). Finally, the function  $f : (X, t) \rightarrow weight$  define (if necessary) the temporal dynamics of the *C2\_link* weight as a function the *weight* itself, time  $t$  and parameters  $P$ , e.g. the weights can be endowed with a dynamical properties that result in “dig-in” effects - older links become harder to undo - (see for example (Tabor and Hutchins, 2004; Vosse and Kempen, 2000) for approaches that play with the temporal dynamics of *C2\_links*.) In the current work, the models’ *C2\_links* will be taken to have time-independent weights. But future work will need to address this issue and compare the results with those establish with the two models cited above.

Cooperation between two instances goes hand in hand with establishing connection between one their ports (an output and an input port), forming an assemblage. For this reason, a cooperation link is always associated with *connect* data, linking an output port of the child instance to an input port of the parent instance. For this reason, a set of *coop\_links* form a network of output-input connections between instances in working memory, creating functional assemblages.

### A.4.3 WM State

The state of a given WM system consists of the currently relevant schema instances as well as their cooperation and competition function links through which they enter in cooperative computation (the C2\_network).  $WM(t) = (Insts(t), l_{coop}(t), l_{comp}(t))$ .

### A.4.4 WM processes

Schema instances invoked in WM enter in cooperative computation. Forming cooperation and competition links, they create a network that defines the state of the WM (see previous paragraph). Being an abstract class, the specific processes that pilot the establishment of cooperation and competition links between instance are to be specified before a given WM can be used in a model. The implementation of the state dynamics as described above is shared by all WM. At each time step  $dt$  the weighted activation values are passed through the network of cooperation and competition links. The activation value of each instance is then updated based on the input it received (see above). The two parameters  $P_{coop}$  and  $P_{comp}$  can be used to govern the probability that a two instances in cooperation (or in competition respectively) will cooperate for this time step. Two instances  $I_1$  and  $I_2$  linked by a cooperation link, will cooperate with probability  $P_{coop}$  at each time step. This can be used to limit the portion of the network's activity that is updated at each time step.

The pruning process removes instances from WM if their activity falls below the pruning threshold  $\theta_{prune}$ .

Because of the processes of invocation of new instances and pruning of old ones, the topology of the C2 network is not fixed.

### A.4.5 Cooperative Computation Dynamics

For a schema instance  $i$ , active in a WM as part of C2 network, its activity  $Act_i^t$  is updated following a leaky integrator equation:

$$Act_i^{t+1} = \alpha Act_i^t + (1 - \alpha)\sigma(Inst_i^t + noise^t) \quad (\text{A.2})$$

- $\alpha$  defines the characteristic time of the system  $\alpha = (1 - \tau^{-1})$

- $\sigma$  the logistic function:

$$\sigma(x) = \frac{1}{(1 + \exp(-k(x - x_0)))} \quad (\text{A.3})$$

- Gaussian noise:

$$noise^t \sim \mathcal{N}(0, std) \quad (\text{A.4})$$

$Inst_i^t$  is defined as:

$$Inst_i^t = w_I \left\{ \sum_{k \in comp(i,k)} w_{comp} \cdot Act_k^t + \sum_{j \in coop(i,j)} w_{coop} \cdot Act_j^t \right\} + w_{ext} \sum_{e \in ext(i)} w_e \cdot Ext_{(e,i)}^t \quad (\text{A.5})$$

$Ext_{(e,i)}^t$  represents activation that an instance  $i$  receives from outside the working memory by subsystem  $e$ .

Here, the competition, cooperation, and external weights are taken to be the same for all instances within a WM and time-independent<sup>2</sup>.

$w_I, w_{ext}$  are tied to the WM,  $w_I/w_{ext}$  balances the strength of internal and external activation inputs.  $\{w_e, e \in ext(i)\}$  depends on the connections established across WM by the individual instances and the strength of the functional connectivity between WM subsystems, as defined by the the system-of-system architecture they belong to.

The parameters of the logistic function  $\sigma$  are chosen so that, in addition to  $\sigma(\infty) = 1$  and  $\sigma(-\infty) = 0$ ,  $\sigma(0) = Act_{rest}$  the activity in the absence of input. The remaining degree of freedom can be used to set

---

<sup>2</sup>This condition can be relaxed and is not a strong requirement of Schema Theory.

$\sigma(x_0)' = \sigma'_0$  in order to define the steepness of the logistic function. In this case the dynamics of the leaky integrator is defined by the parameters

$$(\alpha, A_{rest}, \sigma'_0, w_{coop}, w_{comp}, w_I, w_{ext}, \{w_e\}, noise_{std})$$

. In addition,  $\theta_{prune}$  defines the pruning threshold. A constructions whose activation values falls below  $\theta_{prune}$  is pruned out of working memory. Finally a confidence threshold  $\theta_{conf}$  defines the activation level that has to be reached by a instance (or assemblage) to be considered as a possible solution to the computational problem at hand. Each WM system has its own set of parameters.

#### A.4.6 Assemblage

Each schema instance active in working memory represents a partial solution to the computation the WM system tries to dynamically and incrementally generate. Schema instance assemblages correspond to sets of collaborating schema instances whose combined data/function results in such a solution. An assemblage is therefore generally defined as a set of instances plus a set of cooperation links. In addition an assemblage, just like any schema, is assigned an activation value which is derived from the activation values of the instances that it contains. It is also assign a score, which can differ from the activation value and can incorporate task dependent scoring criteria (e.g. number of instances, . . .). If the state of a WM can be seen as a set of competition and cooperation links between instance, it can also be seen as a superposition of assemblages, each ultimately corresponding to a solution to the computational problem the WM is trying to solve. Said differently, the C2 dynamics in WM between instances implements a dynamic search in the space of assemblages.

$$\text{Assemblage} = (\text{Insts}, \text{Coop\_Links}, \text{act}, \text{score})$$

Since each coop\_link is associated with a connection from an output port to an input port, an assemblage can also be defined as a set of instances and the connections they established between their ports. It is therefore a schema instance network, and by way of composition, a schema instance (albeit a possibly transient one). The system allows for an assemblage to become entrenched (or simply summarized for modeling simplicity) into a stable schema (instance) directly stored in LTM (possibly alongside the very schemas that it is composed of), opening avenue for modeling learning procedure through principles such as template memorization, fragmentation, & generalization (Arbib et al., 1987), adoption, generalization, & consolidation (Spranger and Steels, 2015; Tomasello, 2009), or, for a diagnostic and repair approach to learning, consolidation, generalization and re-specialization (Van Eecke and Beuls, 2017).

## Appendix B

# SALVIA Production Simulations

### B.1 Parameter Space and Default Values

Unless specified, the parameters used for the model in the following simulations are those indicated in the tables B.1, B.1, B.1, B.1.

**Working memories** Each working memory is parametrized by both dynamic and C2 parameters. Dynamic parameters control the state temporal characteristics while the C2 parameter more specifically define the competition-cooperation process. The model contains 4 working memories each involving 7 dynamic parameters. Since only the Grammatical WM includes a full fledged C2 in processing, its behavior is defined by 6 C2 parameters while all the other WMs only require a single C2 parameter (the pruning threshold).

The working memories therefore define, in the absence of any other assumptions, a 41 dimensional parameter space. Such a large space cannot be exhaustively analyzed and further assumptions on which relevant values will be considered fixed as well as clearer analysis of the relations that exist between the dimensions will be needed to analyze the model's behavior.

GrammaticalWM		
Types	Params	Values
Dynamic	$\tau$	30.0
	$act_{inf}$	0.0
	$L$	1.0
	$k$	10.0
	$x_0$	0.5
	$noise_m$	0.0
	$noise_{std}$	0.2
C2	$w_{coop}$	1.0
	$w_{comp}$	-4.0
	$\theta_{confidence}$	0.7
	$\theta_{prune}$	0.01
	$r_{deact}$	0.8
	$w_{deact}$	0.0

Table B.1: Grammatical WM parameters

SemanticWM		
Types	Params	Values
Dynamic	$\tau$	1000.0
	$act_{inf}$	0.0
	$L$	1.0
	$k$	10.0
	$x_0$	0.5
	$noise_m$	0.0
	$noise_{std}$	0.2
C2	$\theta_{prune}$	0.01

Table B.2: Semantic WM parameters

PhonologicalWM		
Types	Params	Values
Dynamic	$\tau$	100.0
	$act_{inf}$	0.0
	$L$	1.0
	$k$	10.0
	$x_0$	0.5
	$noise_m$	0.0
	$noise_{std}$	0.2
C2	$\theta_{prune}$	0.01

Table B.3: Phonological WM parameters

VisualWM		
Types	Params	Values
Dynamic	$\tau$	300.0
	$act_{inf}$	0.0
	$L$	1.0
	$k$	10.0
	$x_0$	0.5
	$noise_m$	0.0
	$noise_{std}$	0.2
C2	$\theta_{prune}$	0.01

Table B.4: Visual WM parameters

**Control** The Control schema defines 2 task related parameters, reflecting the characteristic of the task itself: time pressure and time at which the system needs to start producing a description. It also defines 4 style parameters (although, their sum being necessarily equal to 1, they only form a 3 dimensional parameter space). (see tab. B.1)

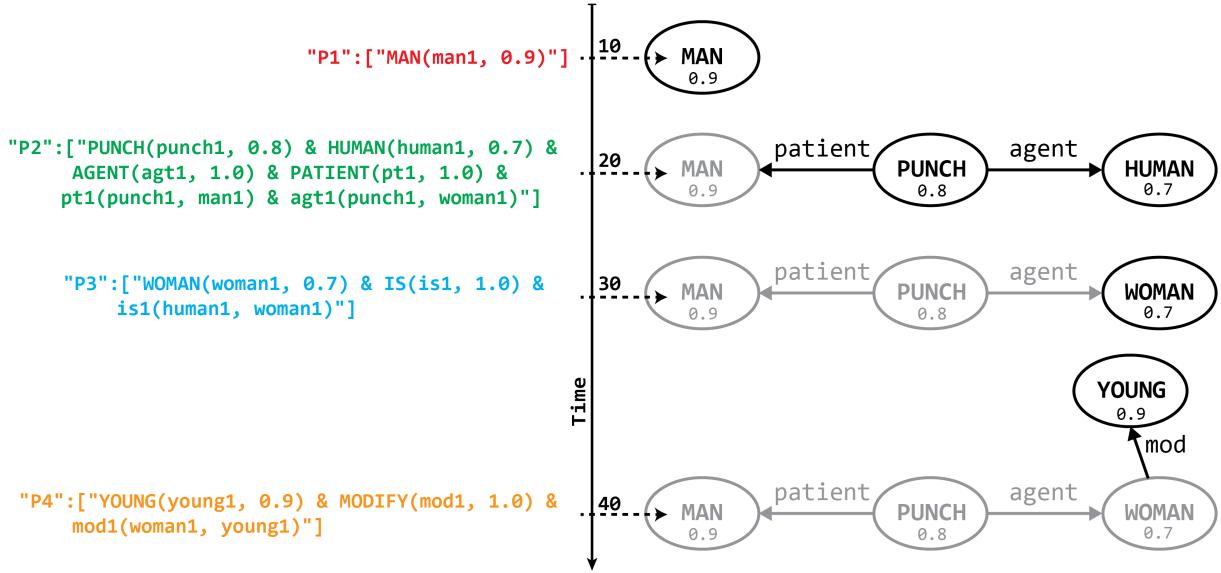


Figure B.1: A SemRep incremental input. This type of input bypasses the scene recognition system and the visual WM, directly providing to the system the incremental semantic content it should process as a basis to generate utterances (simulating only the incrementality inherent to the gathering of semantic content during visual scene parsing). (Left): The inputs are defined as simple propositions, each defining either concepts or relations. A value stipulate the rate at which the propositions should be interpreted and used to update the state of the SemanticWM. (Right) State of the SemanticWM updated each time new semantic content is received as input. The SemRep is built incrementally (at each time step, novel semantic content is shown is highlighted).

Control		
Types	Params	Values
Task	<i>time_pressure</i>	500.0
	<i>start_produce</i>	500.0
Style	<i>w<sub>act</sub></i>	0.7
	<i>w<sub>sem</sub></i>	0.3
	<i>w<sub>form</sub></i>	0.0
	<i>w<sub>cont</sub></i>	0.0

Table B.5: Control parameters

**Schema System** The only parameter carried by the schema system is the time step which can be set to 1 without loss of generality.

## B.2 SemRep Incremental Input

Figure B.1 presents a SemRep incremental input. The detail regarding how the incremental semantic input are defined is provided in Appendix C, Sec. C.1.5.

### B.3 Complex SemRep Incremental Input

t	SEMANTIC INPUT
100	EVENT(evt1,F) & TRANS_ACT(a1, F) & IS(is1) & is1(evt1,a1)
172	KICK(kick) & IS(is2) & is2(a1, kick) & ENTITY(e1,F) & ENTITY(e2,F) & AGENT(agt1) & PATIENT(pt1) & agt1(a1, e1) & pt1(a1, e2)
270	WOMAN(woman) & IS(is3) & is3(e1, woman)
343	BALL(ball) & IS(is4) & is4(e2, ball)
424	EVENT(evt3, F) & TRANS_ACT(a3, F) & IS(is8) & is8(evt3, a3) & MODIFY(mod1) & mod1(e2, evt3)
508	CHASE(chase) & IS(is9) & is9(a3, chase) & ENTITY(e4,F) & AGENT(agt3) & PATIENT(pt3) & agt3(a3, e4) & pt3(a3, e2)
594	DOG(dog) & IS(is10) & is10(e4, dog)
672	EVENT(evt2,F) & CONCURRENT(concurrent) & concurrent(evt1, evt2)
725	INTRANS_ACT(a2,F) & IS(is5) & is5(evt2, a2)
800	BOY(boy) & IS(is7) & is7(e3, boy) & LAUGH(laugh) & IS(is6) & is6(a2, laugh) & AGENT(agt2) & ENTITY(e3, F) & agt2(a2, e3)

Table B.6: Semantic input. (Right column) Concept schema instances sets to instantiate. (Left column) time at which this instantiation should take place (time at which it is assumed the semantic information will be used to update the message - SemRep).

## B.4 Simulations

In the following, the state of the Linguistic WM (Grammatical WM and Semantic WM) is shown for each simulation at different times chosen as they highlight an interesting processing step. In all the figures below. Grammatical WM: Holds active construction schema instances. Semantic WM: Holds active concept schema instances forming the semantic representation (SemRep). Dashed lines: Cross working memory links between a construction instance and the SemRep subgraph that triggered its invocation. Activation flows from the Semantic WM to the Grammatical WM (for clarity, links are only shown between construction instances and concept nodes, not edges). In Grammatical WM, green links indicate cooperation, red links competition.

### B.4.1 Simulation 1

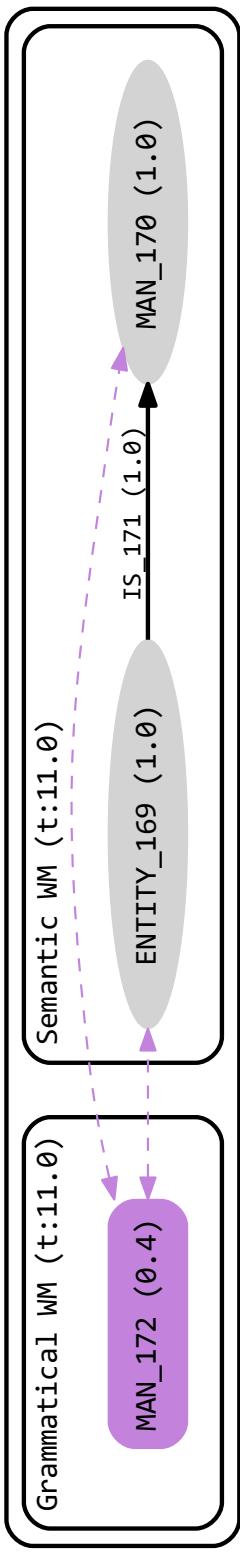


Figure B.2: t=11; MAN semantic information received which has triggered the invocation of the MAN construction instance. The system starts by identifying a single entity (narrow focus, detail first)

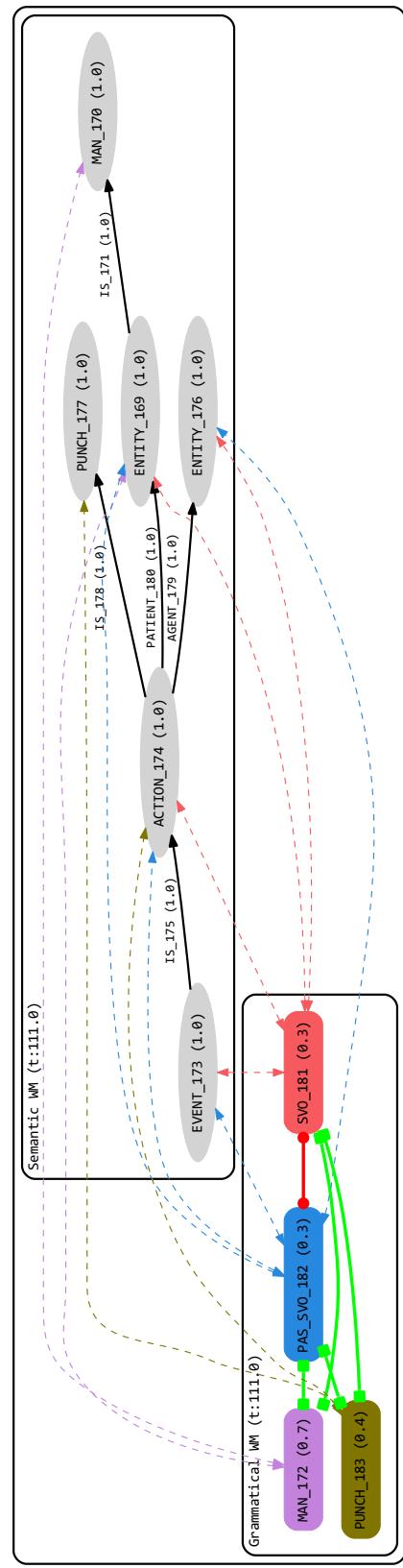


Figure B.3: t=111.0; Detect PUNCH transitive action (switch to a larger focus, event frame). Man is detected to be the PATIENT of the transitive frame. The AGENT is simply construed as ENTITY (kept underspecified). In Grammatical WM: The PUNCH construction has been invoked. The active voice (SVO) and passive voice (PAS\_SVO) argument structure construction instance compete.

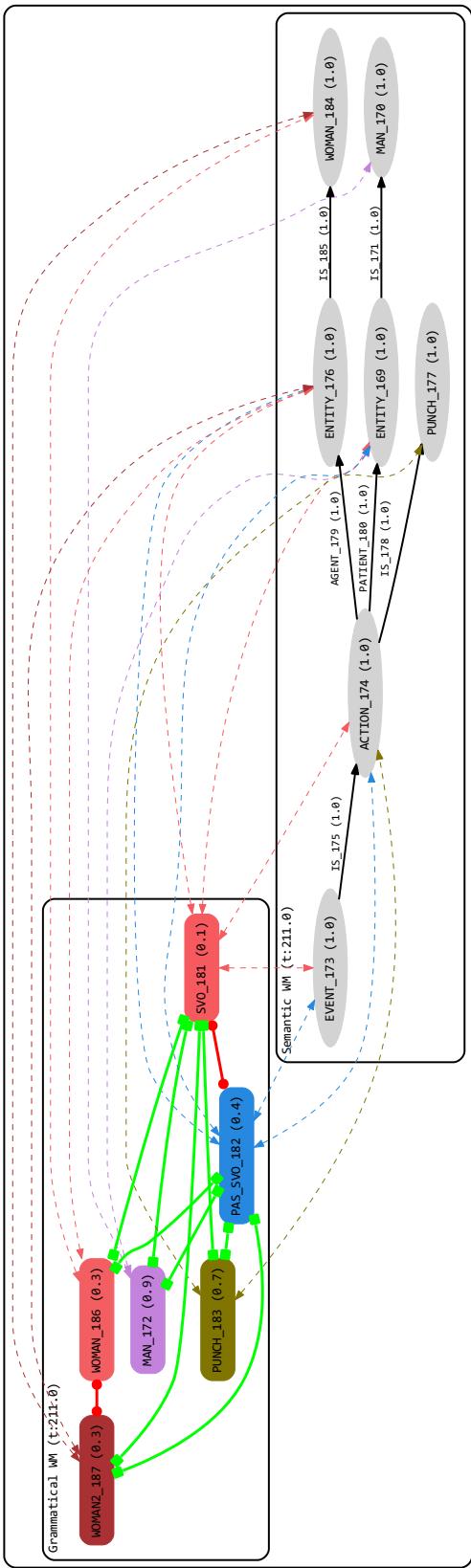


Figure B.4: t=211.0; The AGENT is construed as being a WOMAN. The two WOMAN lexical construction instances (mapping the meaning onto “woman” and “lady”) compete.

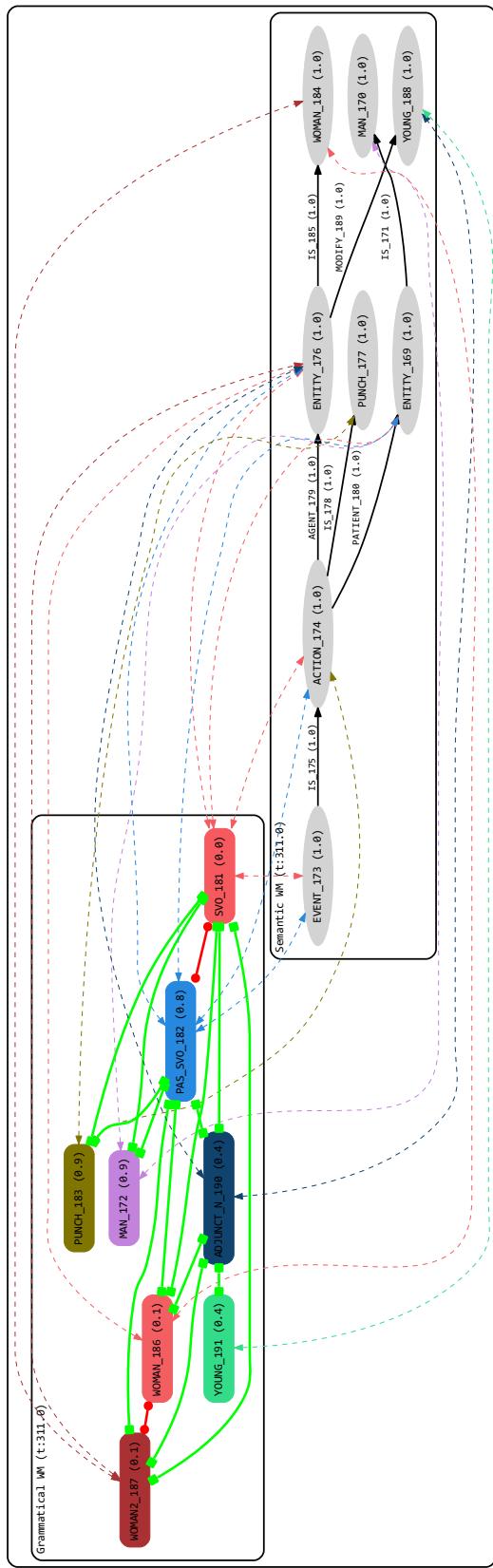


Figure B.5:  $t=311.0$ ; WOMAN is semantically specified as YOUNG. This results in the invocation of both the YOUNG and ADJUNCT construction instances. Passive voice is winning.

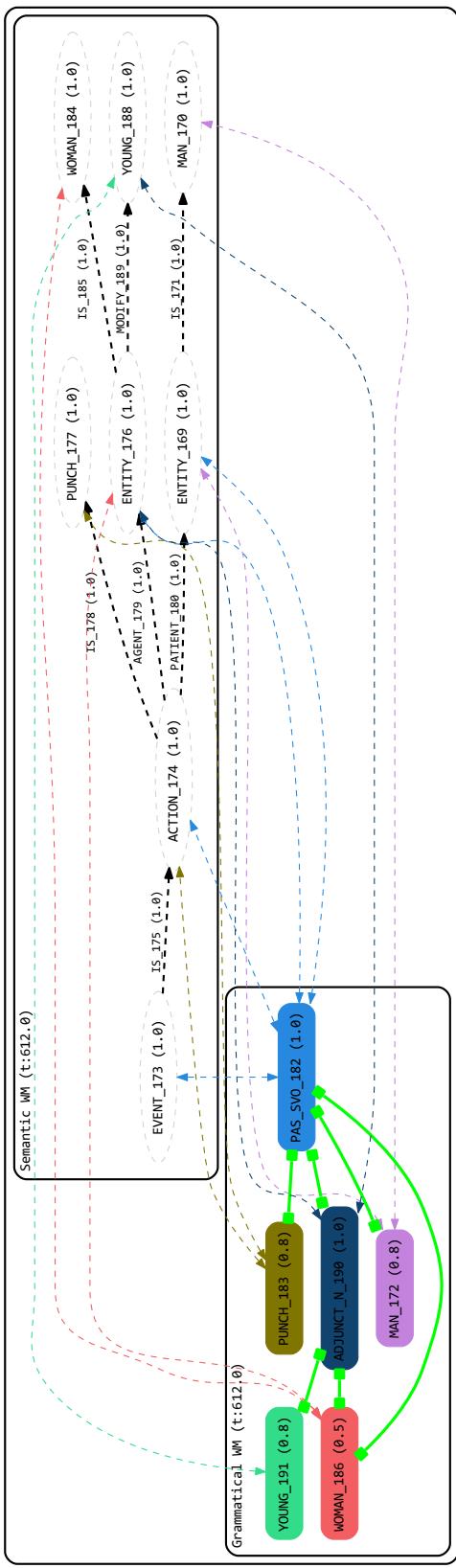


Figure B.6:  $t=612.0$ ; Active voice lost, passive voice selected. In addition only one WOMAN construction remains. This is the final construction instance assemblage from which the linguistic form is generated. It expresses all the semantic content (expressed SemRep noted in dashed lines).

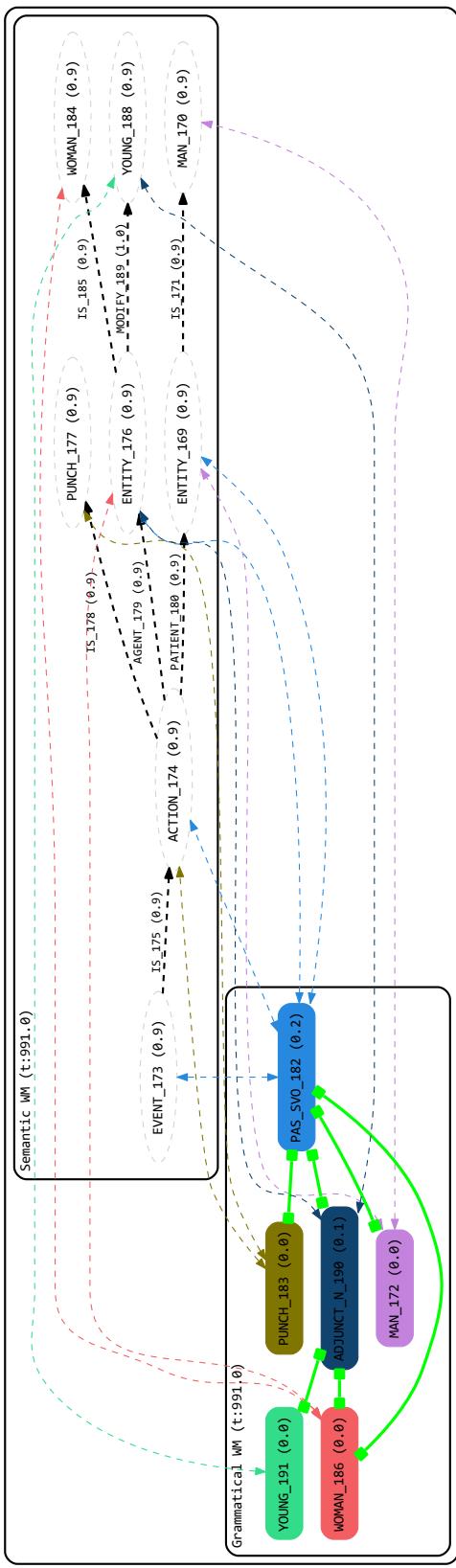


Figure B.7:  $t=991.0$ ; The model outputs "Man is punched by young woman". The instances' activations start to decay.

#### B.4.2 Simulation 2

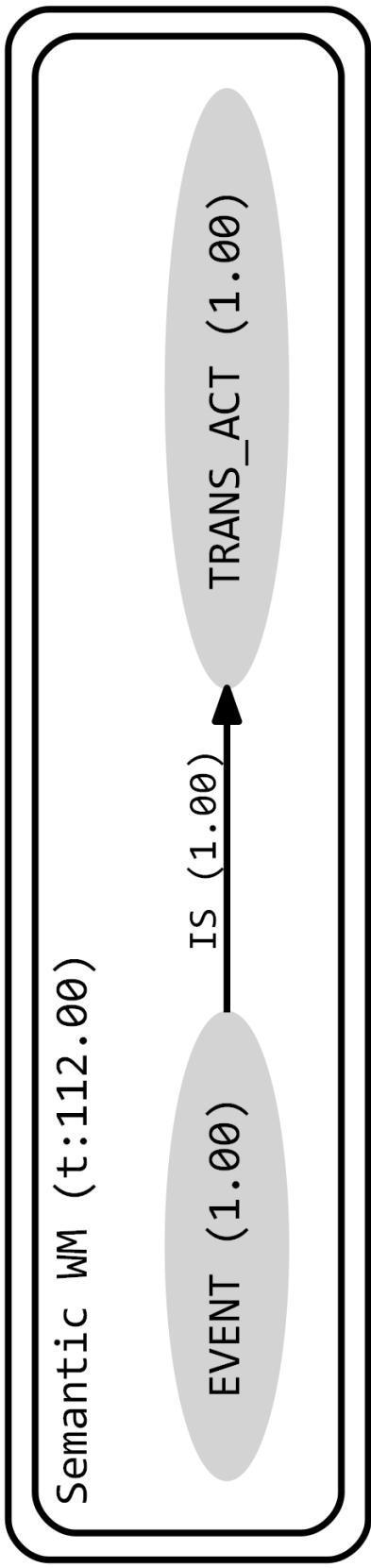


Figure B.8:  $t=112.0$ ; Active event frame retrieved first (large focus). Entities participating in the action have not yet been conceptualized (although they might have already been attended to and partially perceptually processed). The Grammatical WM is empty.

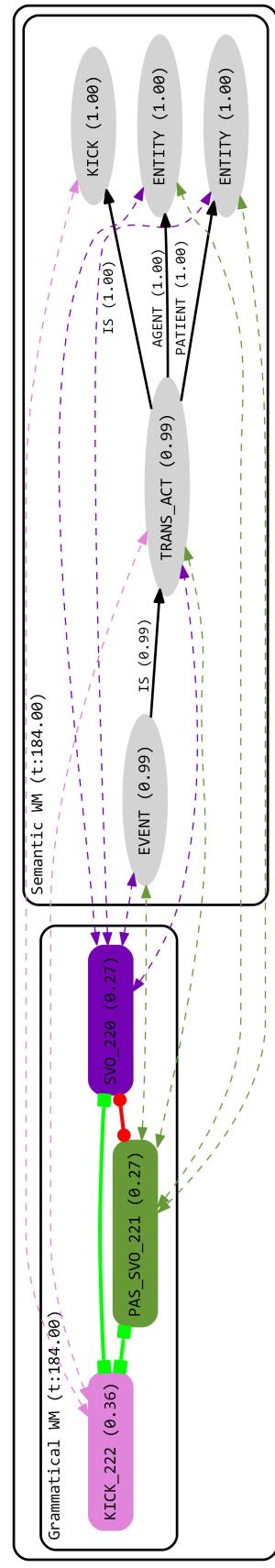


Figure B.9:  $t=184.0$ ; KICK action is perceived and construed as involving AGENT and PATIENT roles. In Grammatical WM, KICK lexical construction instance is invoked. Passive voice (PAS\_SVO) and active voice (SVO) compete for the argument structure (high level grammatical planning).

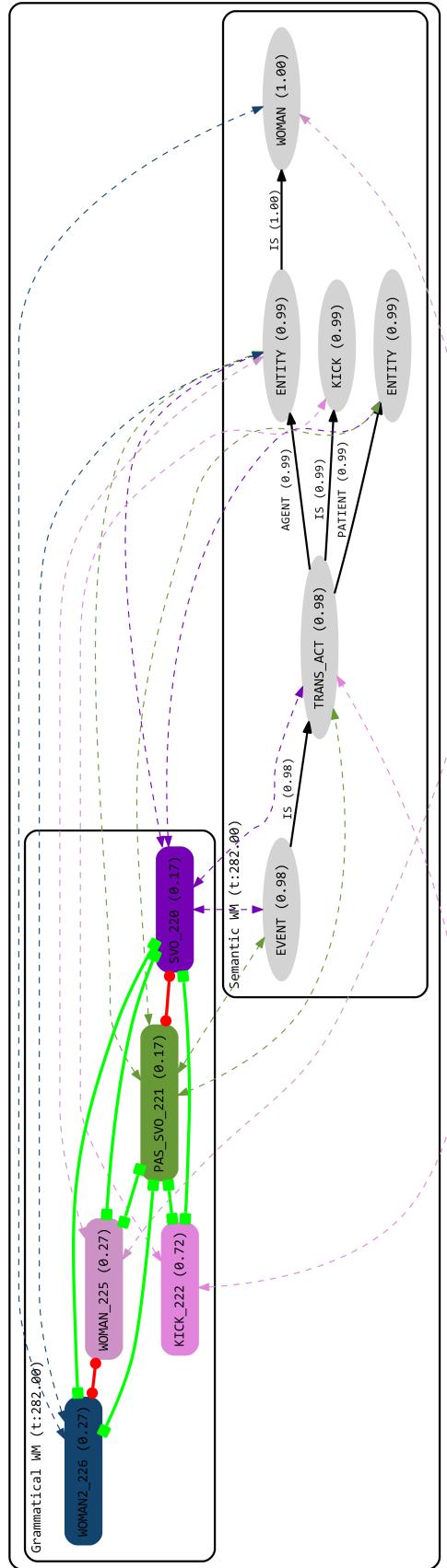


Figure B.10: t=282.0; AGENT construed as WOMAN. Competition between the two possible mapping of WOMAN onto form (“woman” or “lady”), i.e. between two WOMAN construction instances.

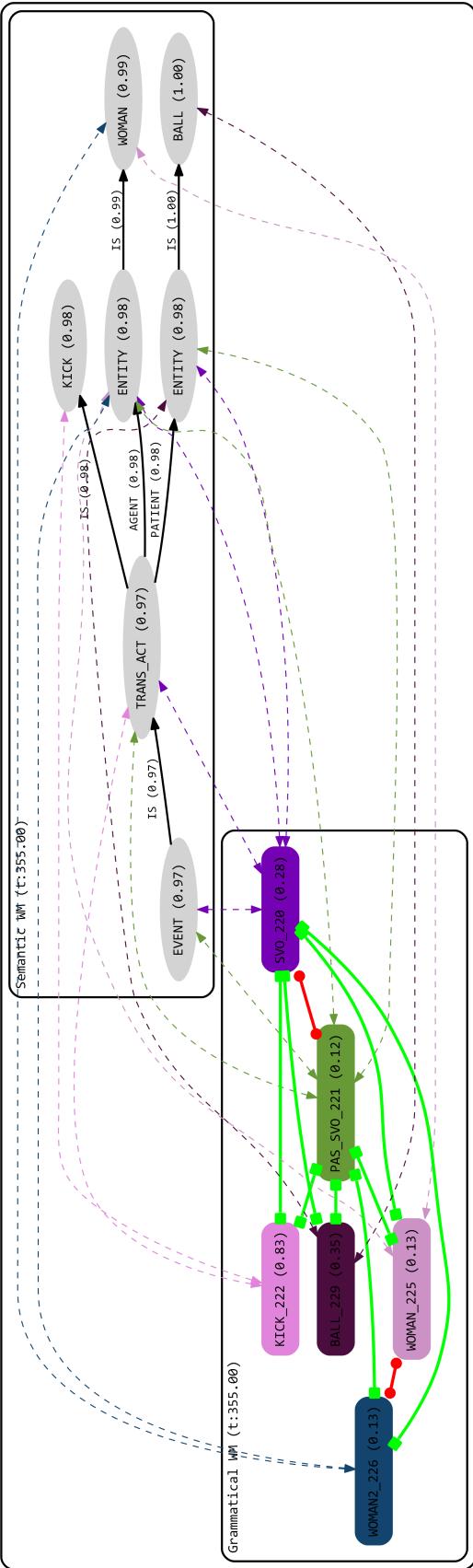


Figure B.11: t=355.0; PATIENT is conceptualized as BALL, and the associated BALL construction instance is invoked.

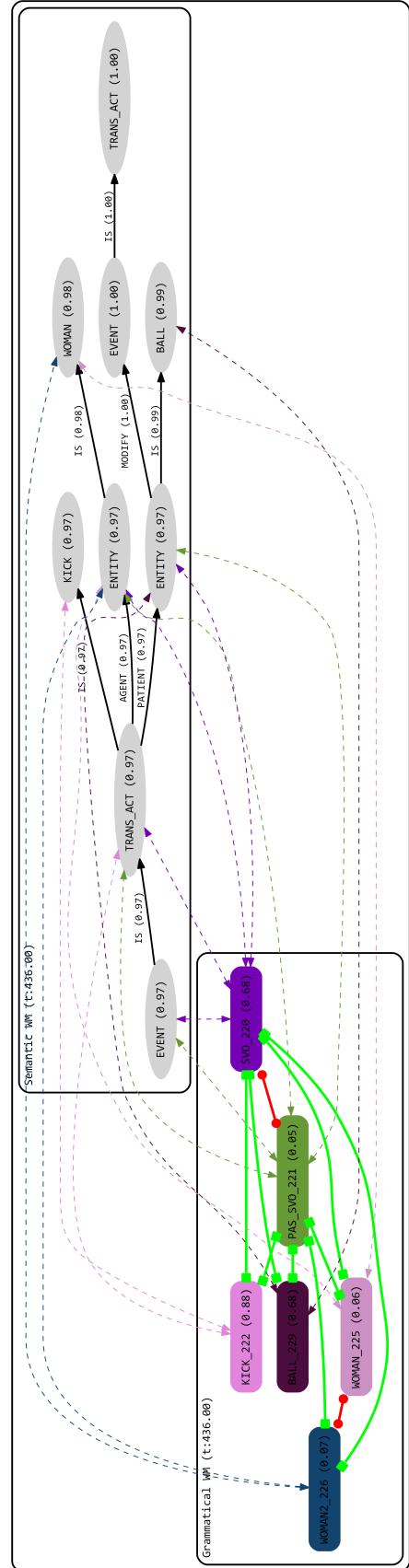


Figure B.12: t=436.0; New event identified and construed as a sub-event of the main event. It is defined as a modifier of the PATIENT ENTITY. Due to the structure of the SemRep, this new event appears as contained in the first event (SemRep forms a partial hierarchy).

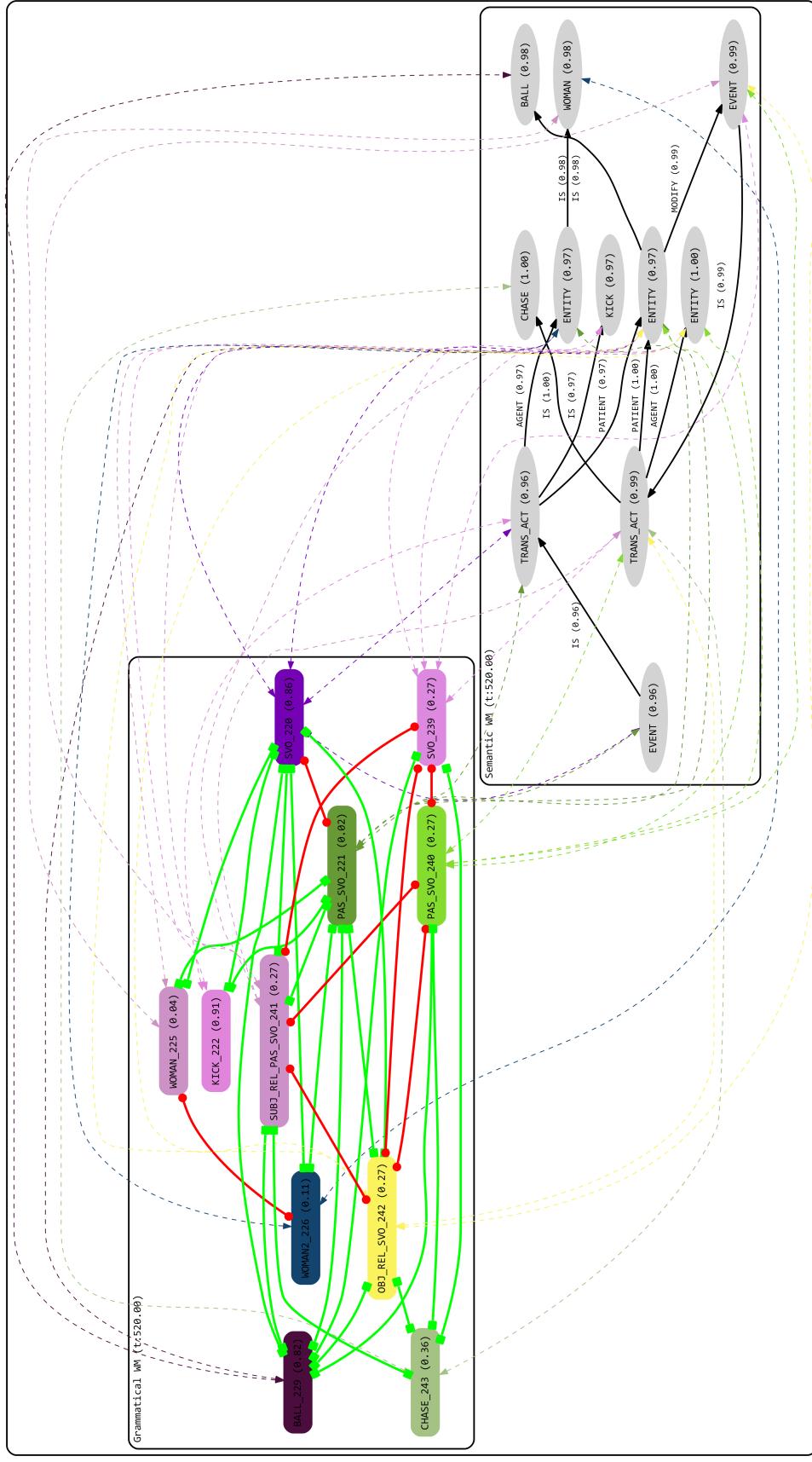


Figure B.13:  $t=520.0$ ; The action of the sub-event is conceptualized as CHASE and BALL is defined as the patient. The new sub-event modifying the patient triggers the retrieval of associated grammatical constructions reflecting the semantic hierarchy in the syntactic hierarchy (subject relative passive voice construction, object relative active voice construction).

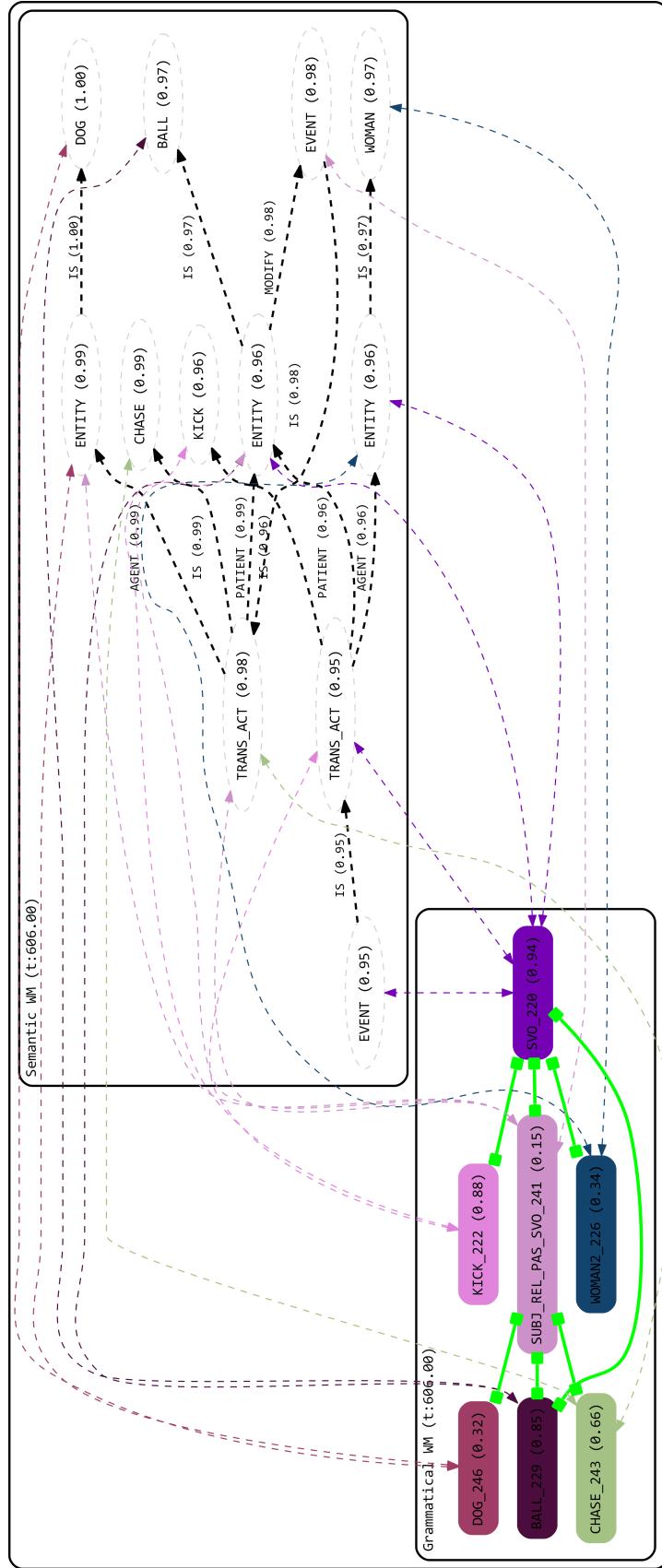


Figure B.14: t=606.0; AGENT of the sub-event is specified as DOG. All competitions are over in Grammatical WM. The final construction instance assemblage expressing the entire SemRep as emerged.

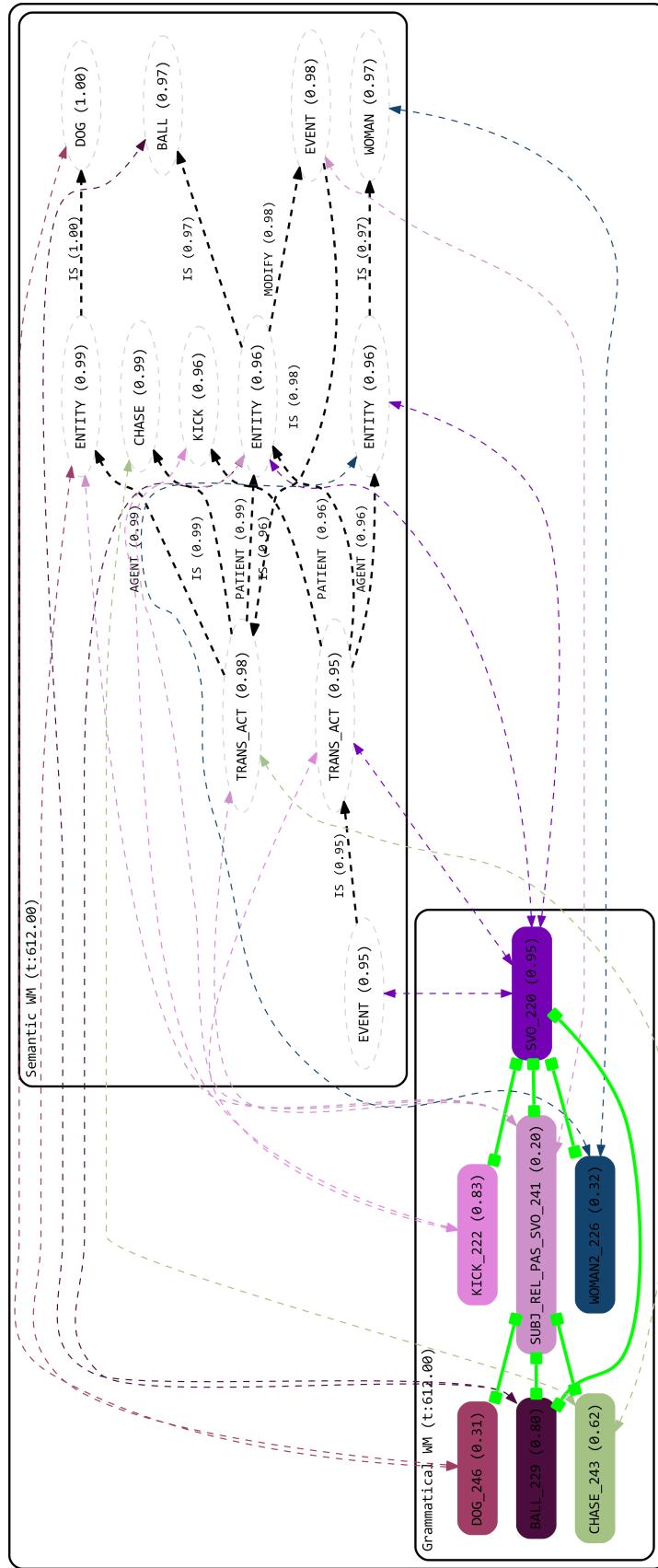


Figure B.15: t=612.0; The model outputs “woman kick ball that is chased by dog”.

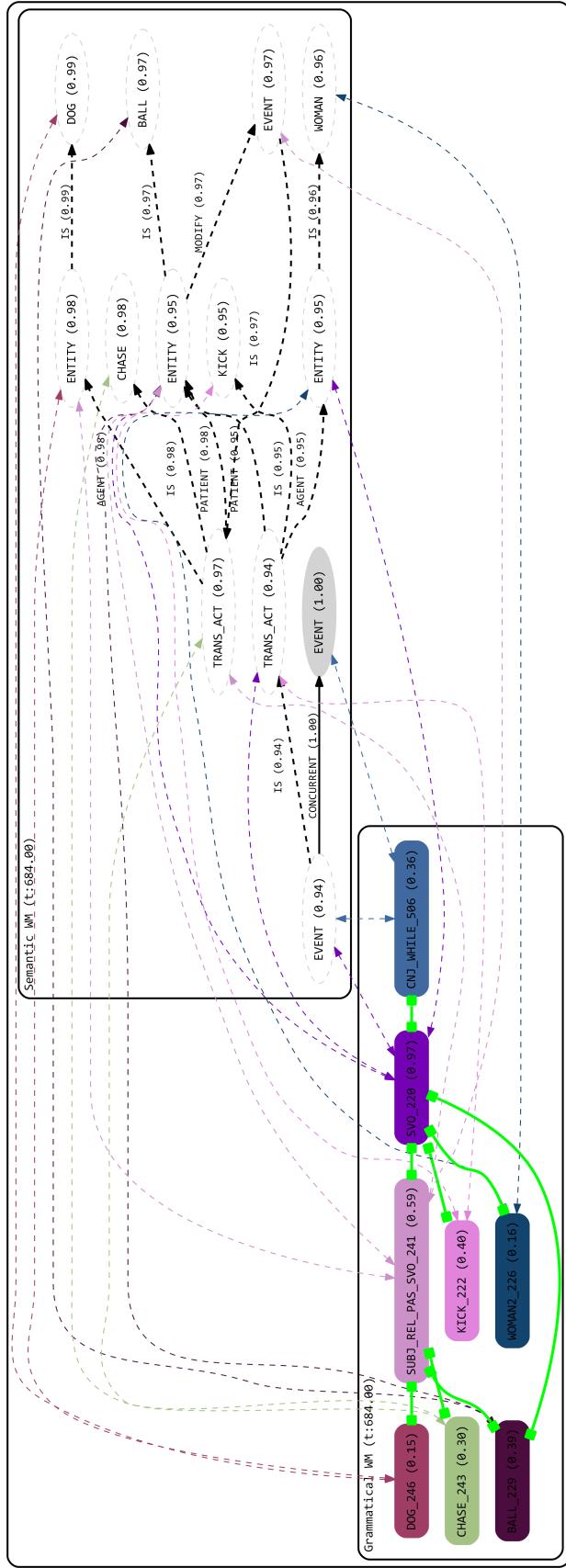


Figure B.16: t=684.0; New event detected and construed as CONCURRENT to the first one. The CNJ WHILE construction is invoked.

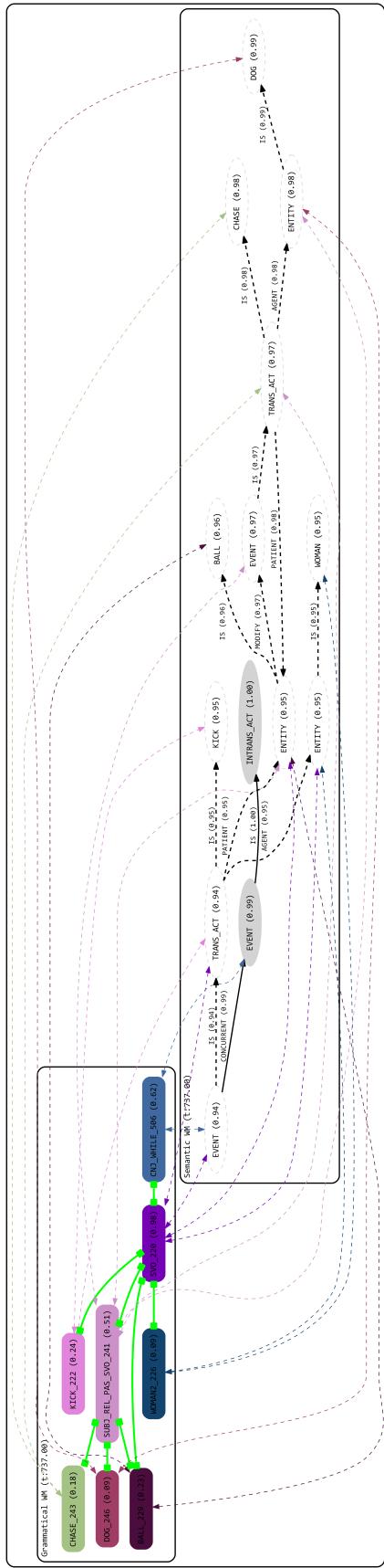


Figure B.17: t=737.0; The new event involves an intransitive action.

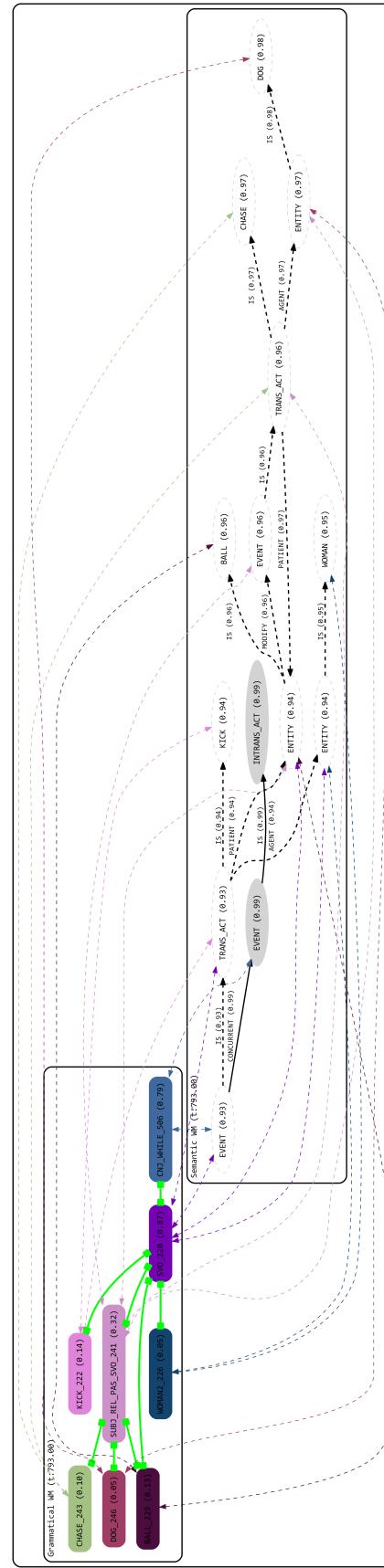


Figure B.18: t=793.0; The model outputs “while”, setting up a smooth continuation for the previous output.

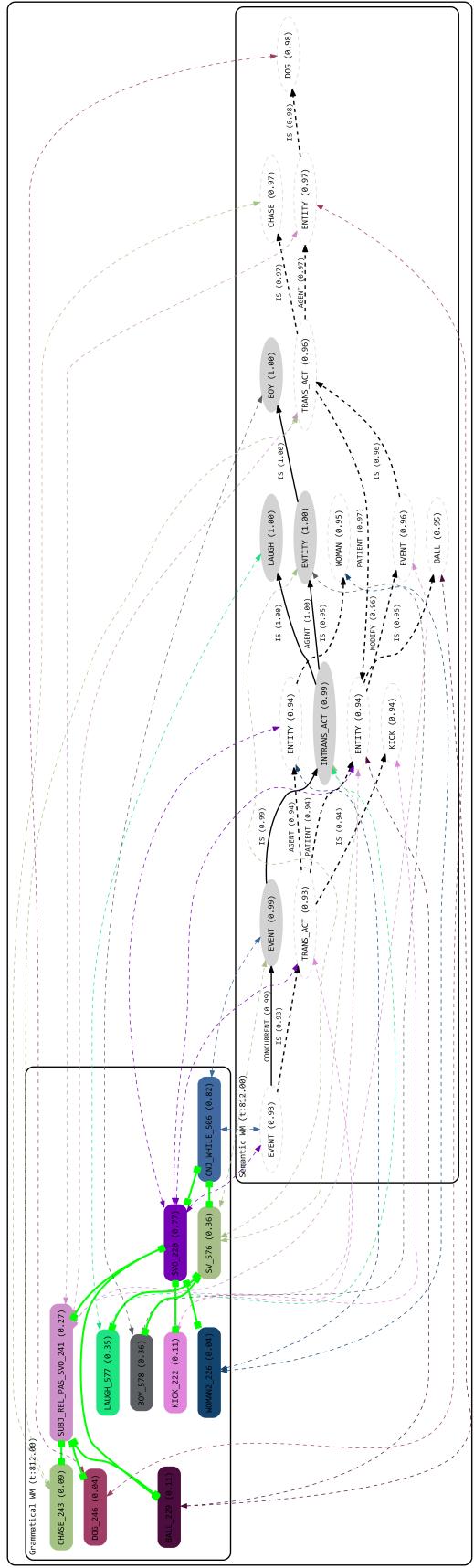


Figure B.19:  $t=812.0$ ; AGENT of the new event is conceptualized as BOY, the action as CRY. Associated lexical construction are invoked plus the intransitive argument structure construction.

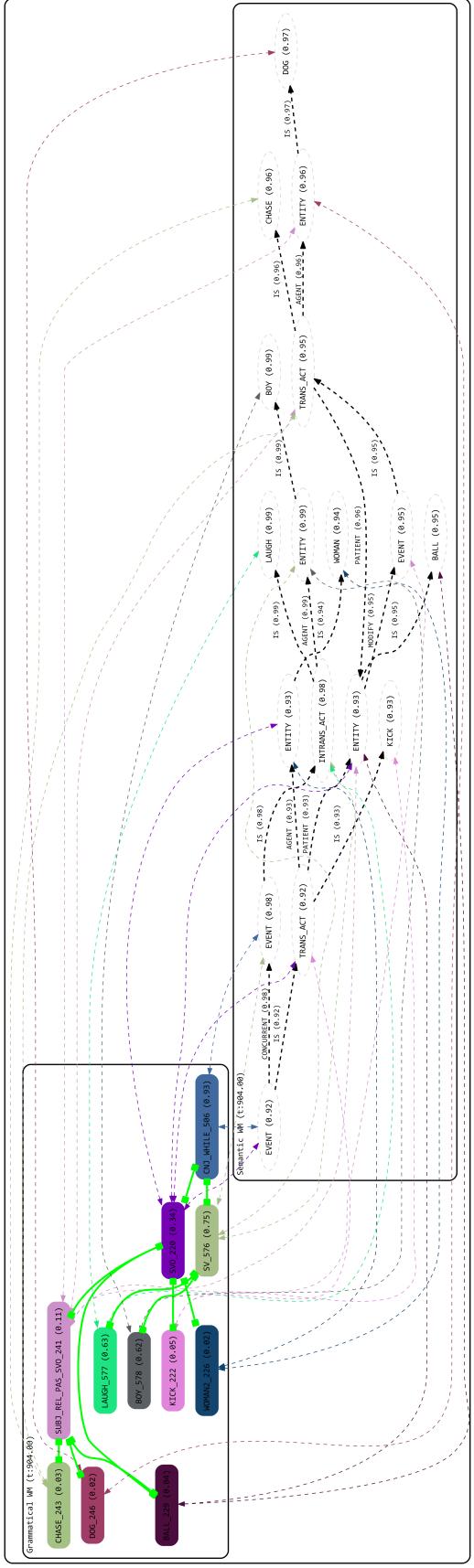


Figure B.20:  $t=904.0$ ; The model outputs “boy cry”, again, providing a smooth continuation for the previous outputs. To summarize, the model overall outputs “[612]woman kick ball that is chased by dog...[793]while...[904]boy cry

### B.4.3 Simulation 3

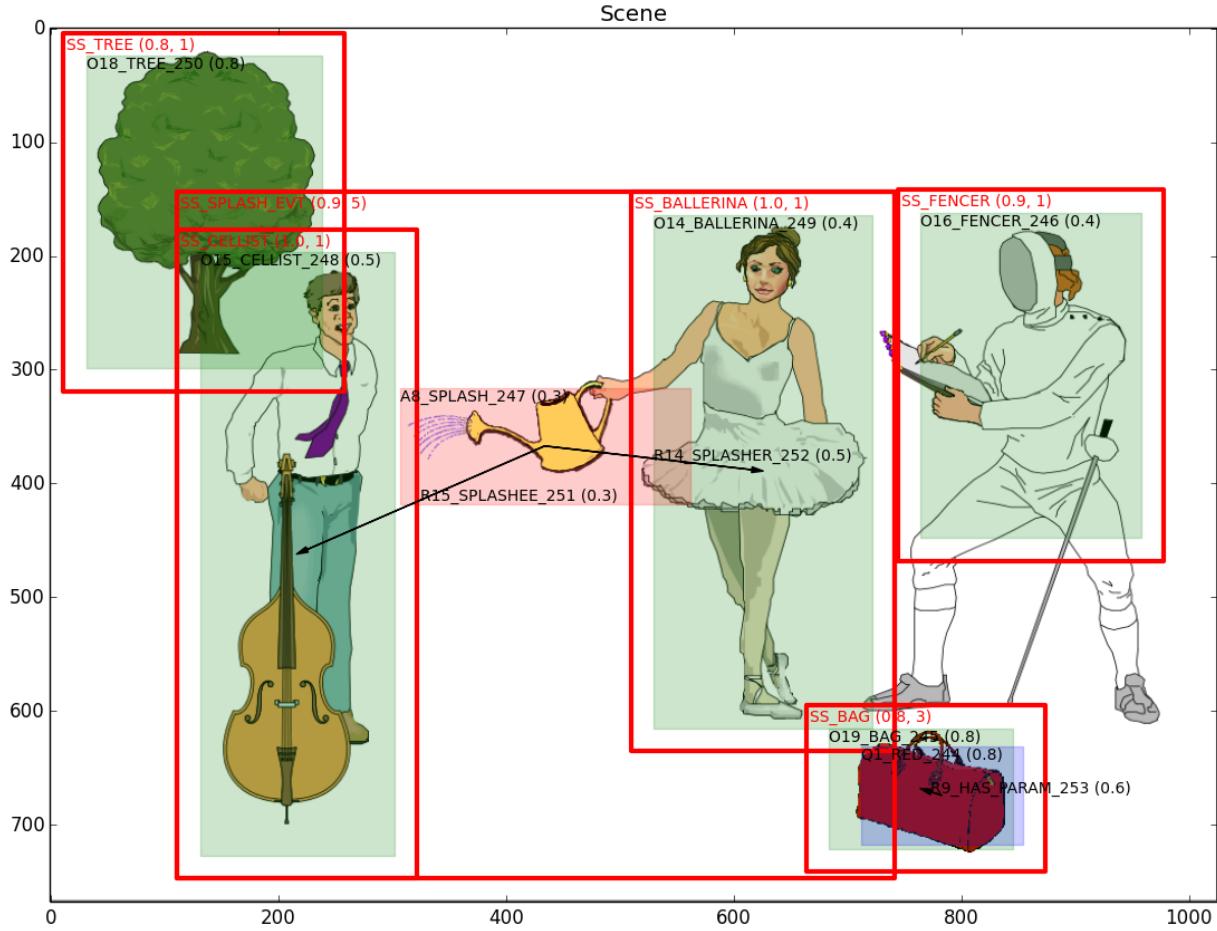


Figure B.21: Input scene. Set of manually defined subscenes (boxes, name with prefix “SS\_”). Each subscene contains a perceptual schema structure (shaded areas and arrows, the latter denote perceptual relation). Here each object is its own subscene. In addition, the subscene that contains the SPLASH perceptual action schema instance, also incorporates in its perceptual structure the BALLERINA and CELLIST subscenes. The perceptual structure defined as input is therefore hierarchical. Each subscene and schema is linked to a spatial region of the visual scene. They also all have a saliency value. In addition, each subscene is given a difficulty value that here is simply defined as the number of perceptual schemas it contains.

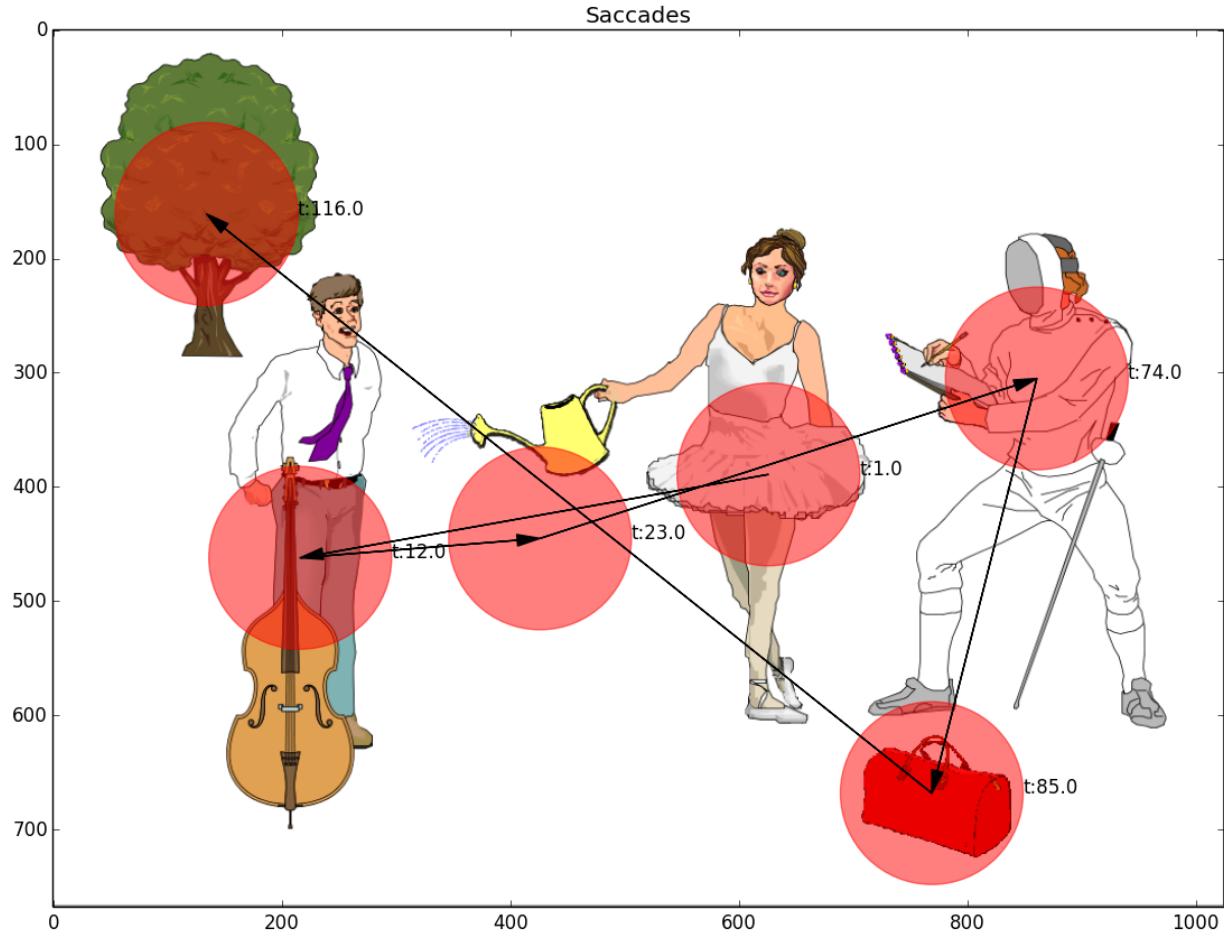


Figure B.22: Simple example of the model’s saccades during language production. Location of gaze is indicated by circles (size of focus window is not shown). Time of gaze start is indicated next to each circle. Arrows mark saccades. At each gaze position, the visual system retrieves the perceptual schema instances pre-associated with the area attended.

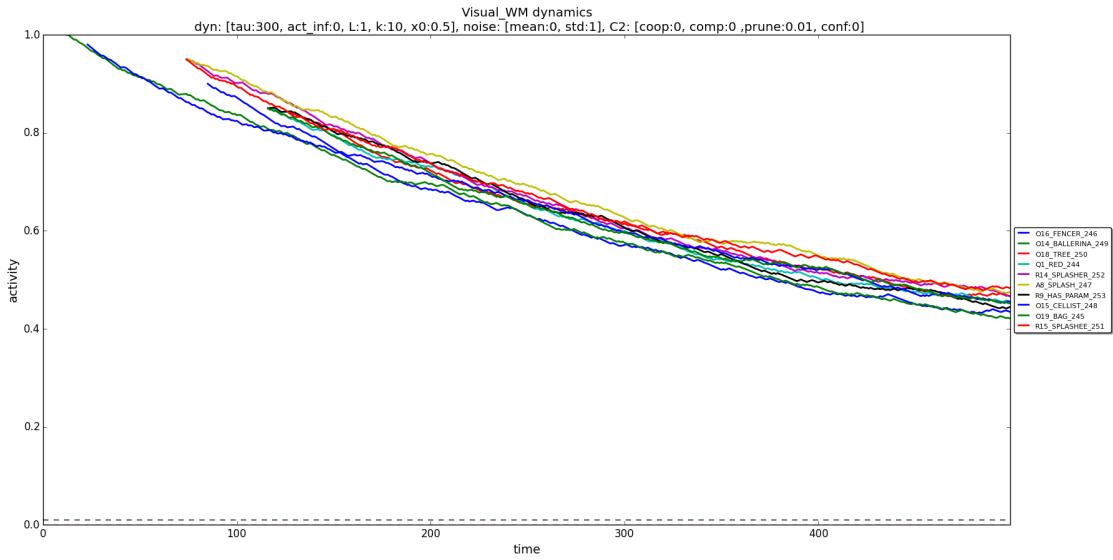


Figure B.23: Percept schema instance dynamics in visual WM. The percept schema instances see their activation values simply slowly decay in Visual WM once they have been invoked. The initial activation value is tied to the saliency level of the visual region they are linked to. Retrieval order follows saliency values. Note that the difference in the activation values due to saliency interact with the decay (competition between saliency and recency of retrieval.)

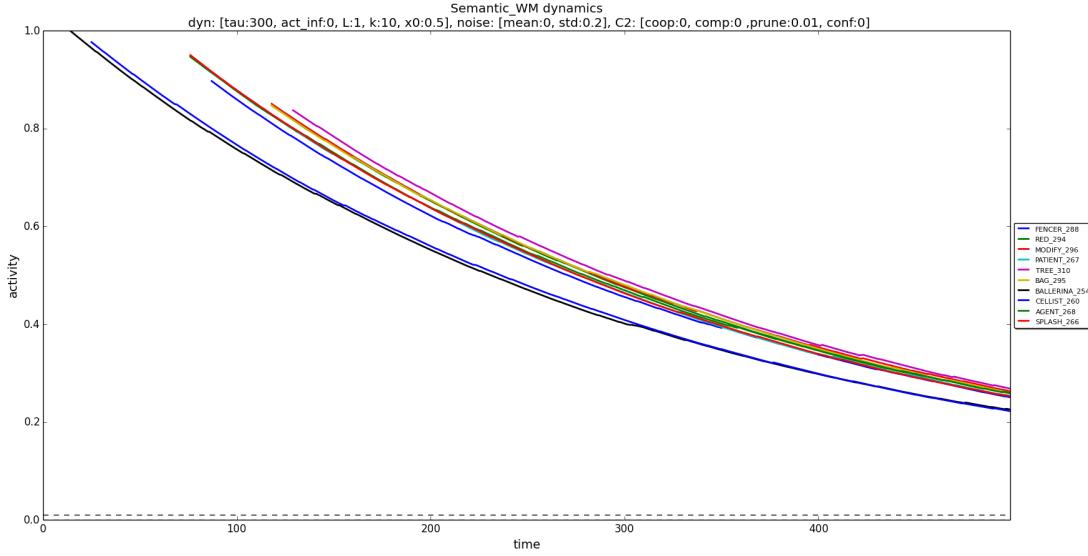


Figure B.24: Concept schema instance dynamics in semantic WM. Due to the simple conceptualization mapping scheme, the Semantic WM concept schema instances' activation pattern follow closely that of the Visual WM. Initial activation value depends on the activity value of the percept schema instance at the time of conceptualization. Here again, relative activation values depend on the interaction between recency of invocation and initial activation value.

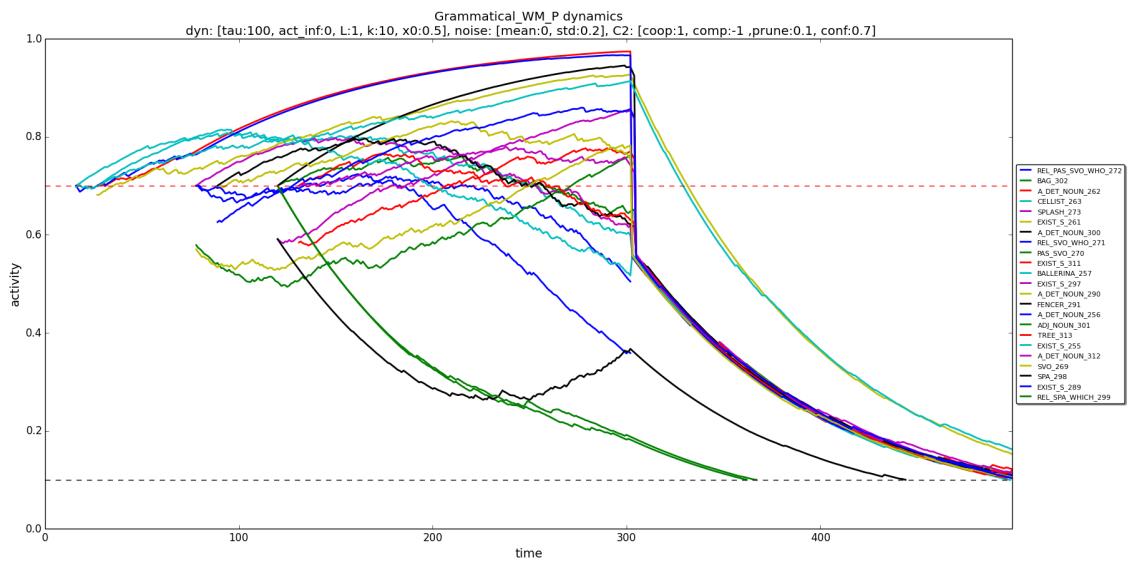


Figure B.25: Construction instances activation in grammatical WM.

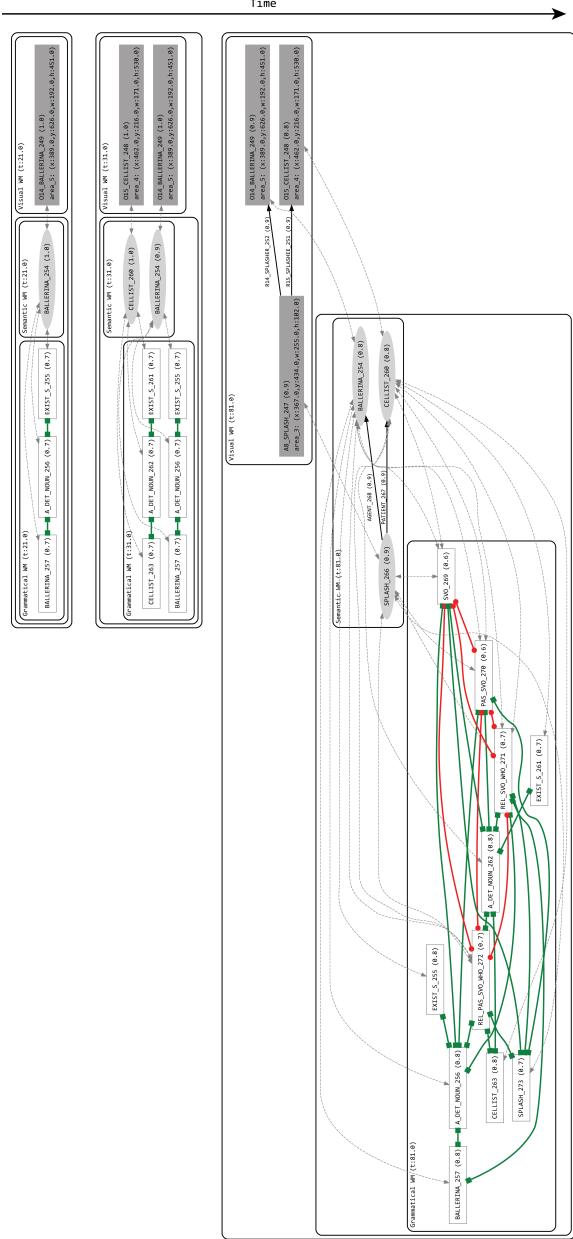


Figure B.26: View of the interactions between visual, semantic and grammatical WM at various time points (arrow of time goes from top to bottom). Grammatical and Semantic WM follow the same convention as before. Here the Visual WM is added. In Visual WM, percept schema instances are active (edge denote relation percept schema instances). Each also carries information about the perceptual retion ( $x, y, w, h$ ) it interprets (and is linked to). Arrows between Visual WM and Semantic WM denotes cross WM links with the flow of activation going from Visual WM to Semantic WM.

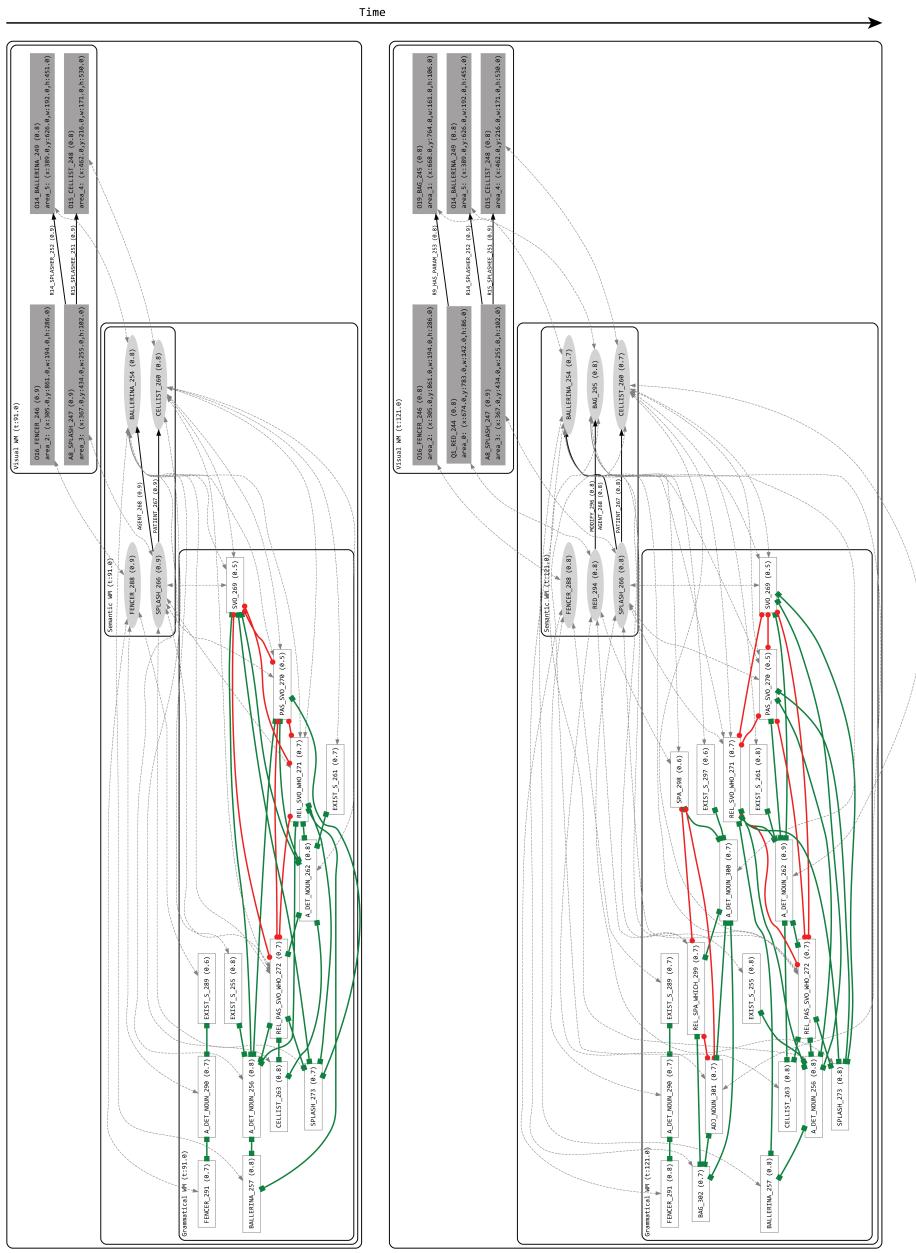


Figure B.27: View of the interactions between visual, semantic and grammatical WM at various time points.(continued)

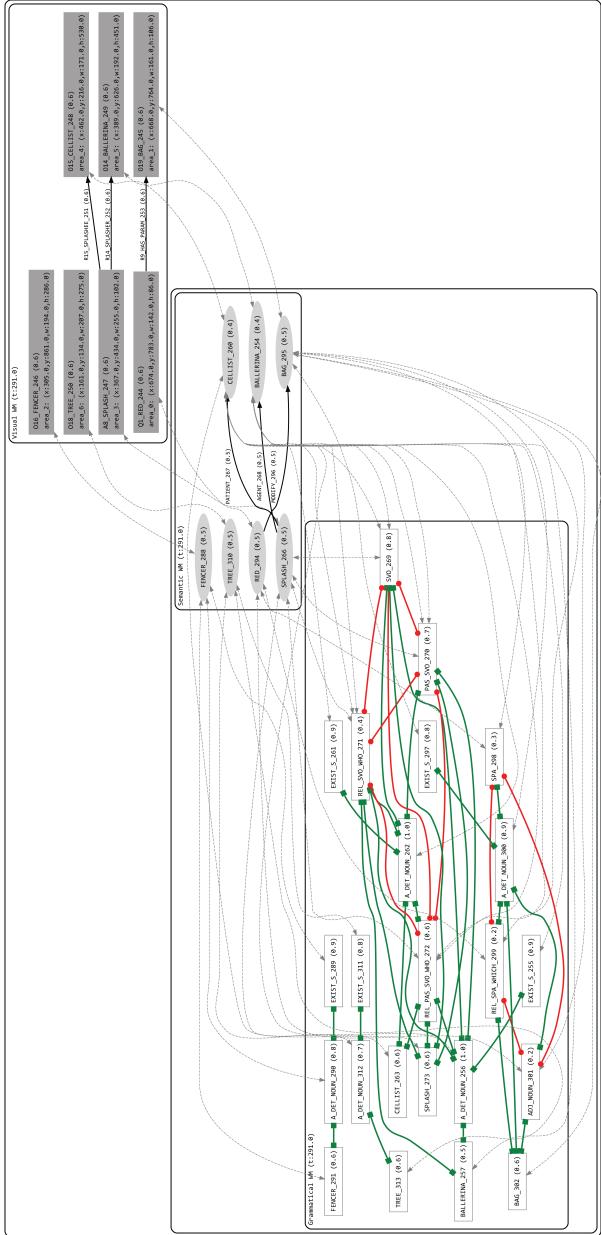


Figure B.28: View of the interactions between visual, semantic, and grammatical WM at various time points.(continued)

#### B.4.4 Simulation 4

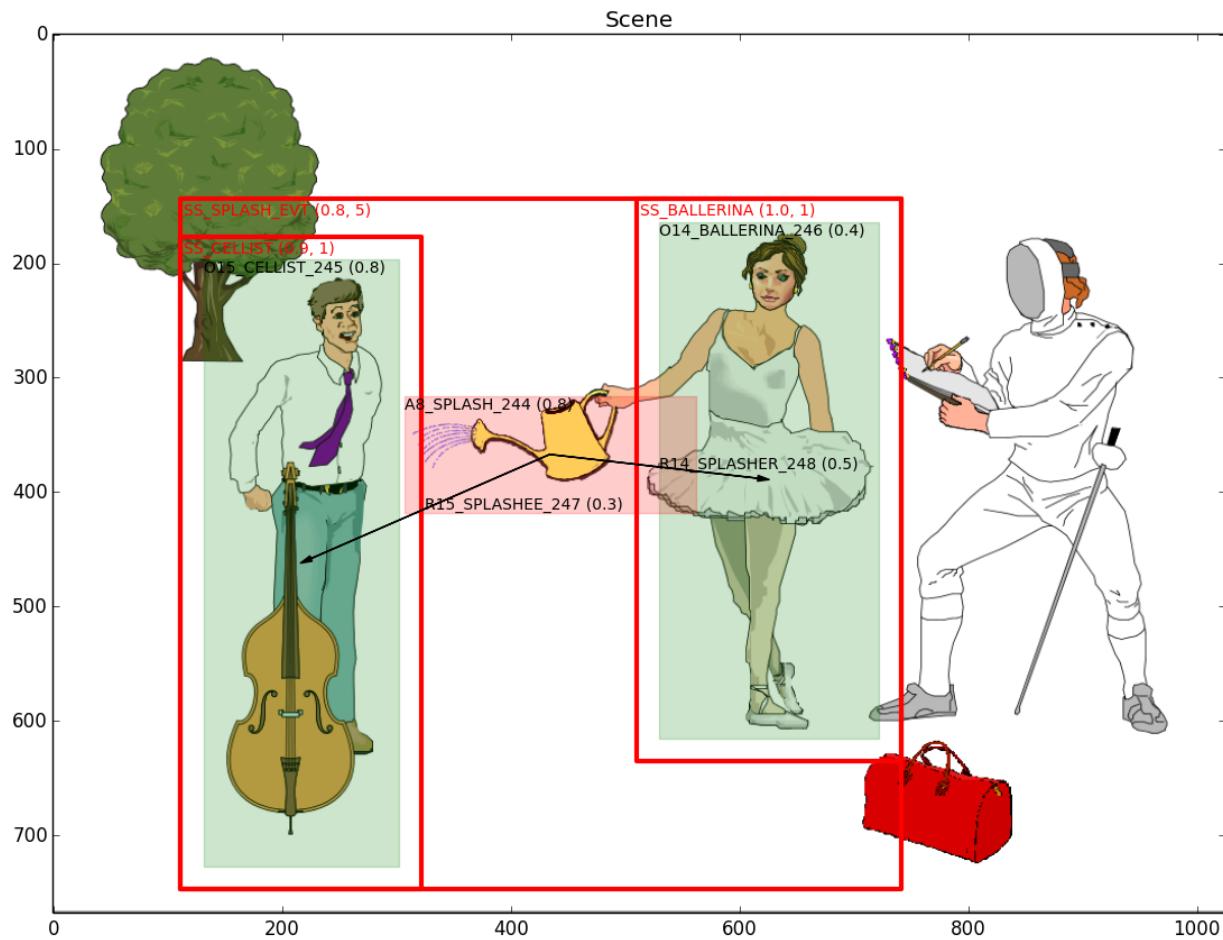


Figure B.29: Input scene. In this case the SS\_BALLERINA subscene containing the agent has saliency 1, while the SS\_CELLIST subscene containing the patient has saliency 0.8.

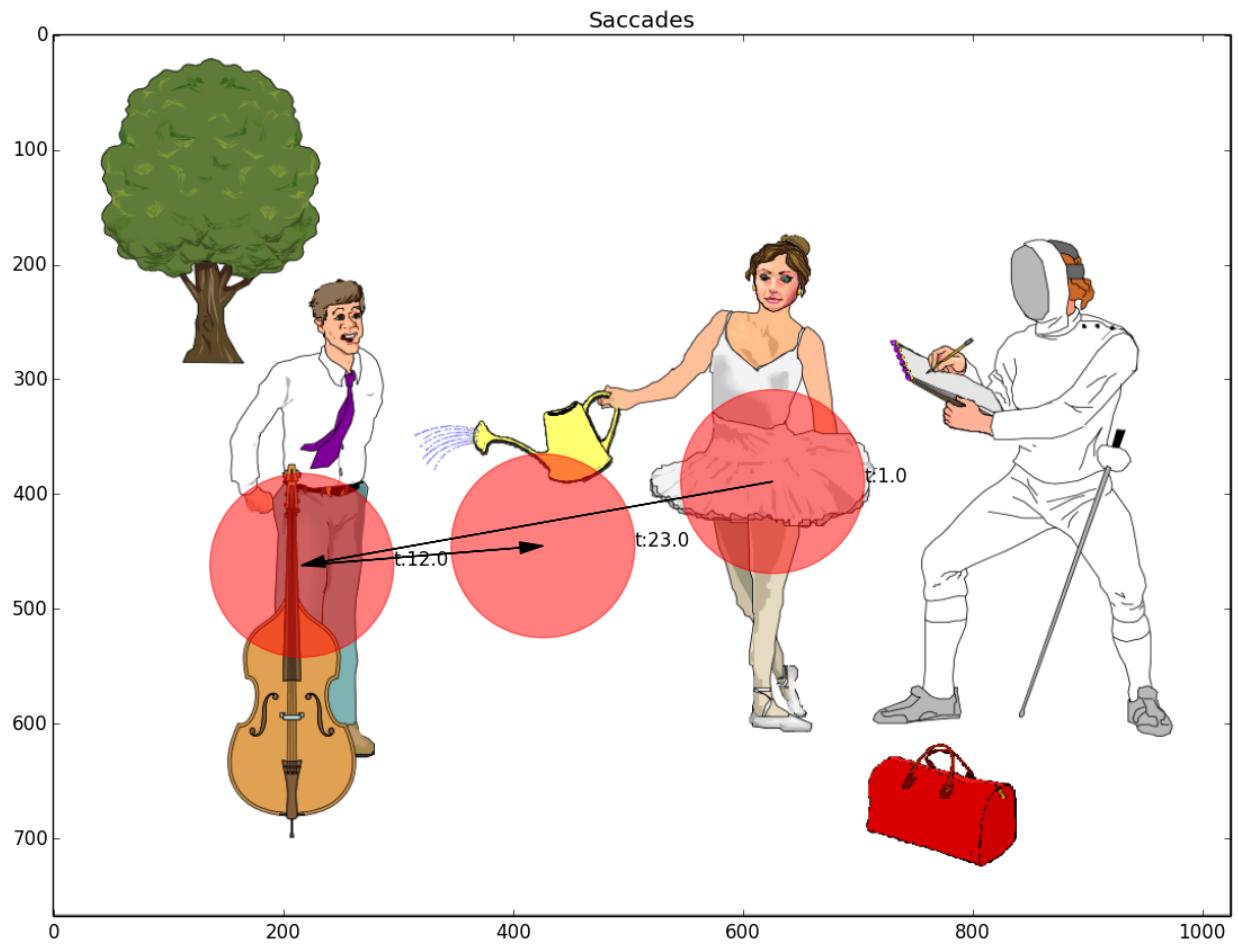


Figure B.30: Model's saccades. The model inspect the BALLERINA subscene first, then the CELLIST and finally the SPLASH\_EVT subscene.

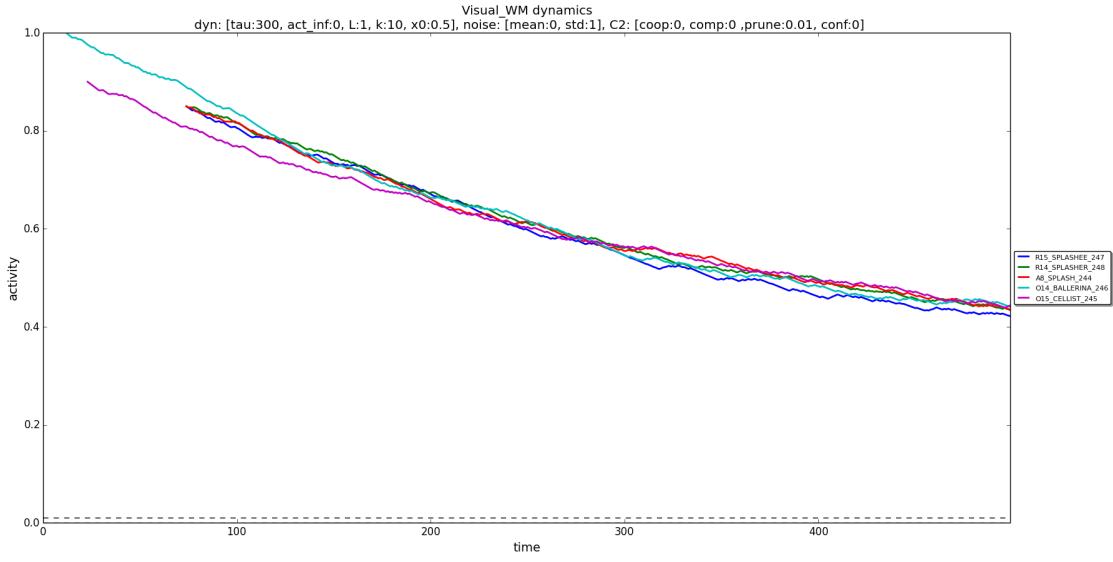


Figure B.31: Percept schema instance dynamics in visual WM. The initial activation of the schema instances reflect the saliency of the subscene they belong to. O14\_BALLERINA is the first percept schema instance to enter visual WM followed by O15\_CELLIST with the later receiving a lower activation value. Note that because of the leaky integrator nature of the dynamics, the difference in activation level between agent and patient percept instances diminishes with time.

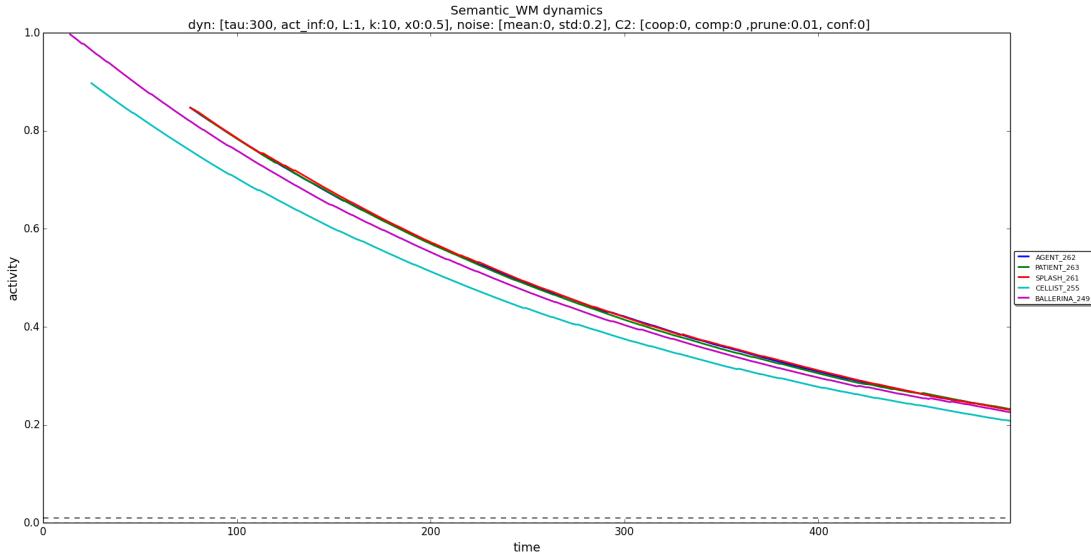


Figure B.32: Concept schema instance dynamics in semantic WM. Following the timing in visual WM, the BALLERINA concept is instantiated first followed by the CELLIST concept. Their initial activation values reflect the activation values of the percept schemas they conceptualize at the time of instantiation. The difference in activation level between the BALLERINA and the CELLIST concept instance implements the information structure level of the SemRep with a higher activation value corresponding to a higher "focus" value.

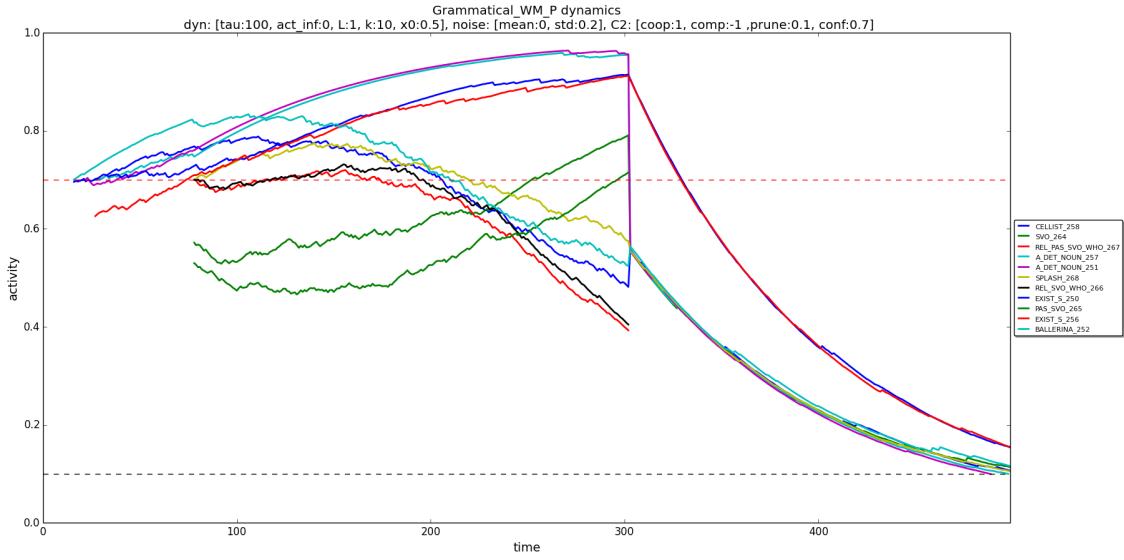


Figure B.33: Construction instances activation in grammatical WM. The main element to look at here are the two green line at the bottom of the graph. The top one represents the SVO cxn instance while the bottom one represents the PAS\_SVO cxn instance. In the grammar used for this experiment, both constructions have the same preference value and should therefore start with a similar activation level. The difference in initial activation value results from the fact that the SemFrame of SVO matches the SemRep better since it favors an Agent focus. This initial boost results in SVO instances staying above PAS\_SVO during processing until production is triggered at  $t = 300$ . Consequently, the model will produce an utterance structured by the SVO cxn and therefore in active voice.

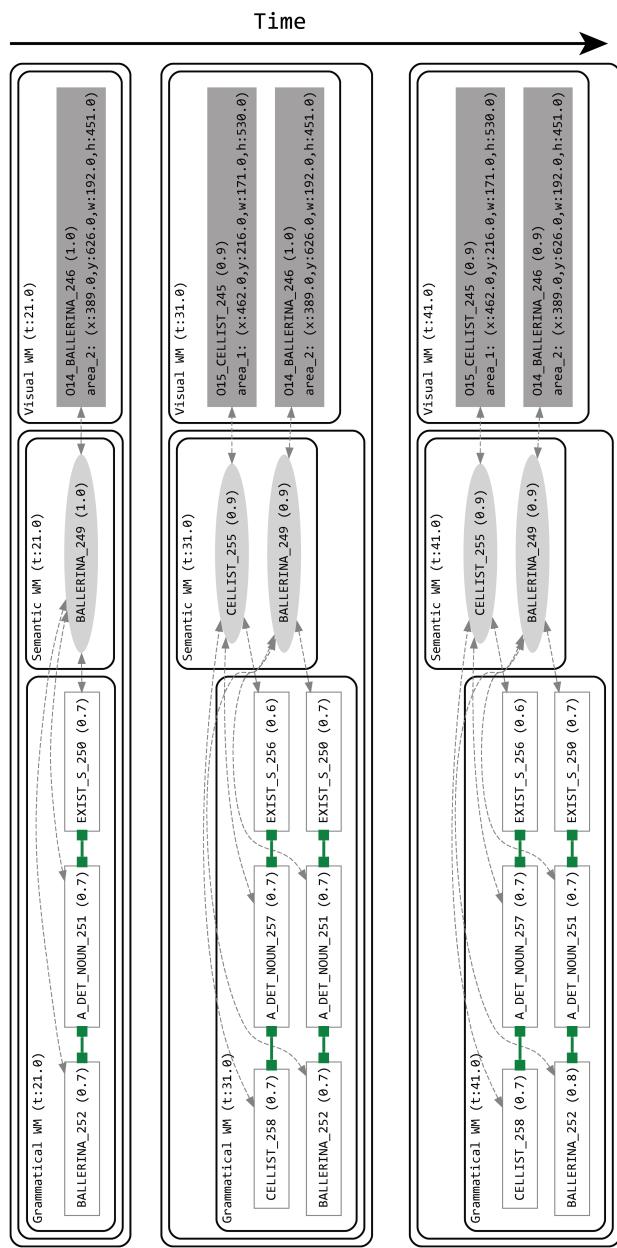


Figure B.34: View of the interactions between visual, semantic and grammatical WM at various time points.

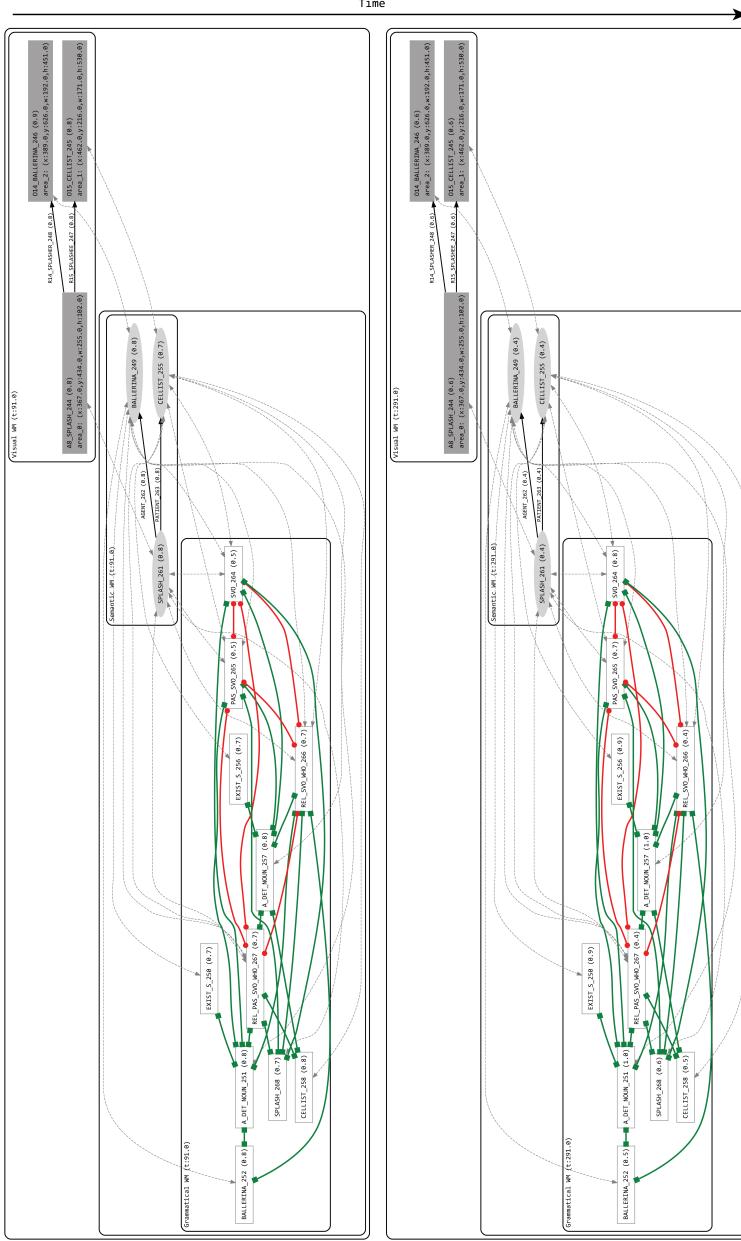


Figure B.35: View of the interactions between visual semantic and grammatical WM at various time points. (continued). The final output of the model is the utterance "a ballerina splash a cellist".

### B.4.5 Simulation 5

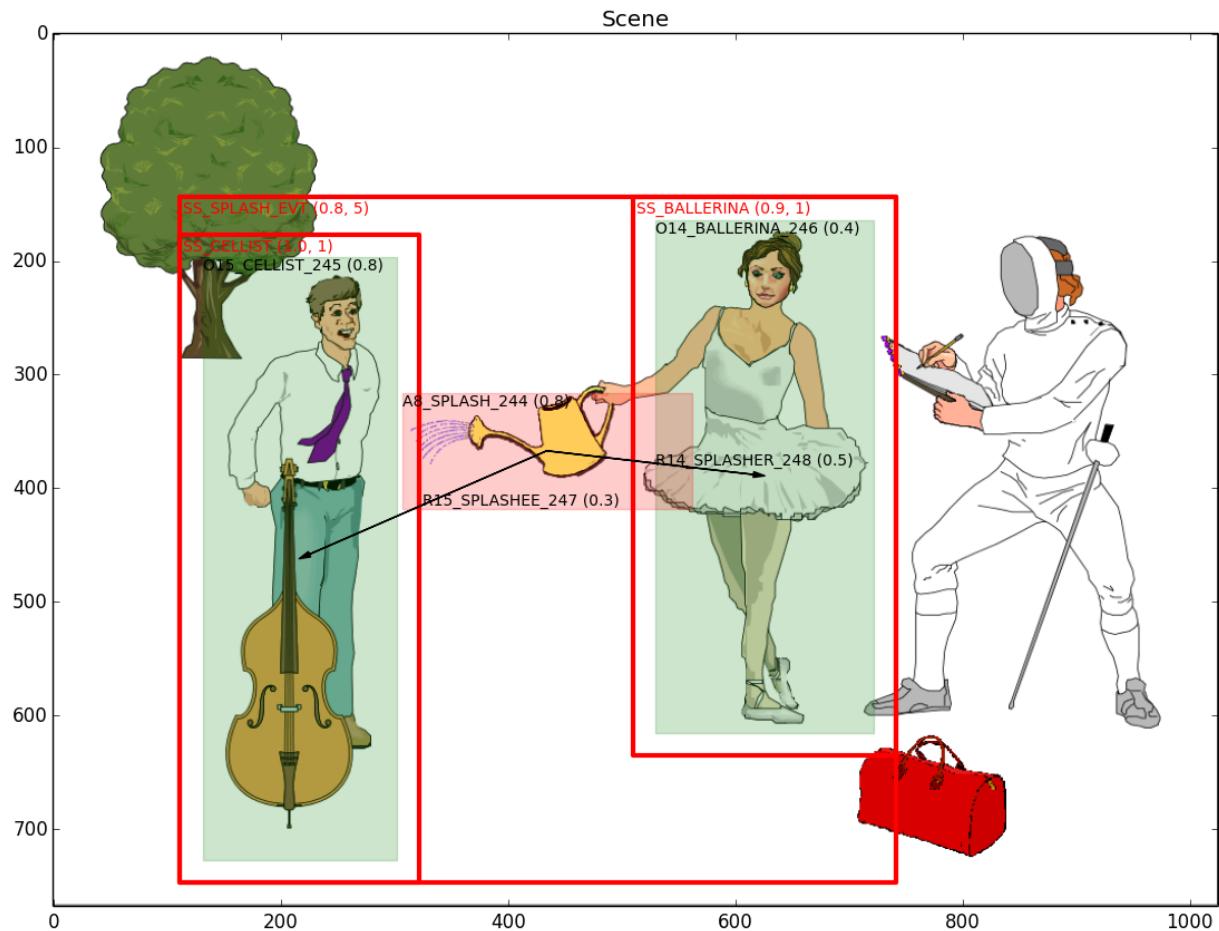


Figure B.36: Input scene. In this case the SS\_BALLERINA subscene containing the agent has saliency 0.9, while the SS\_CELLIST subscene containing the patient has saliency 1.

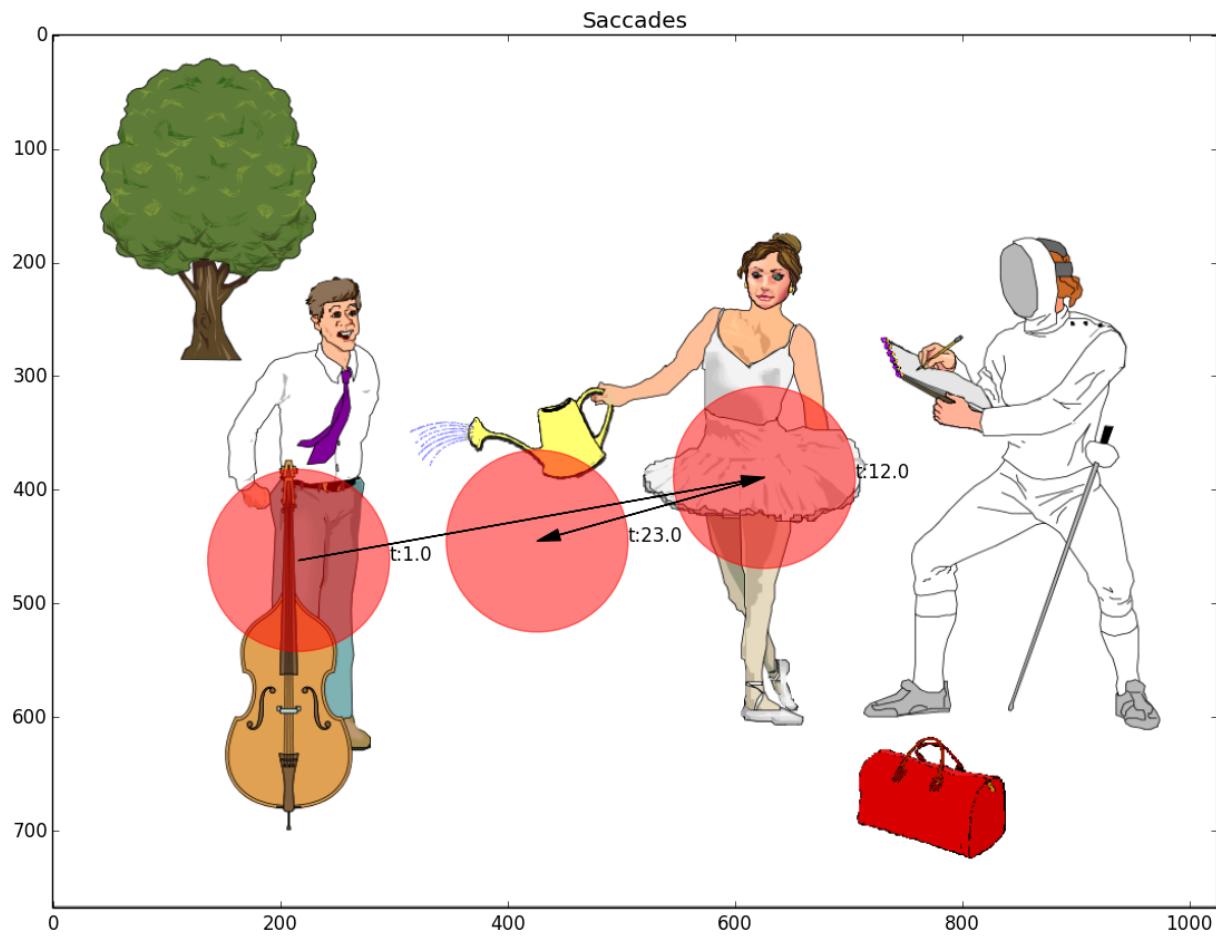


Figure B.37: Model's saccades. The model inspect the CELLIST subscene first, then the BALLERINA and finally the SPLASH\_EVT subscene.

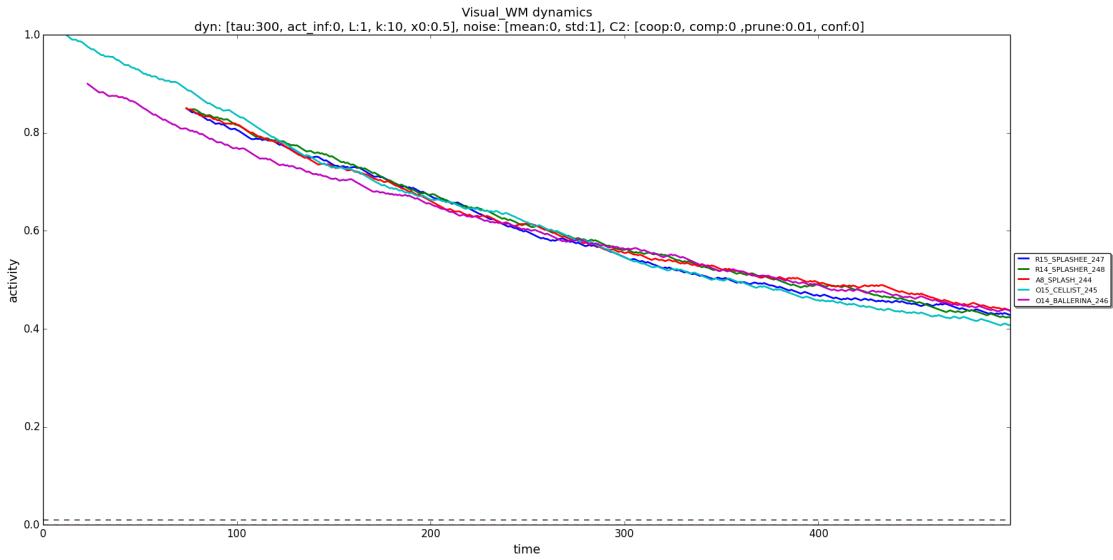


Figure B.38: Percept schema instance dynamics in visual WM. The initial activation of the schema instances reflect the saliency of the subscene they belong to. O15\_CELLIST is the first percept schema instance to enter visual WM followed by O14\_BALLERINA with the later receiving a lower activation value. Note that because of the leaky integrator nature of the dynamics, the difference in activation level between agent and patient percept instances diminishes with time.

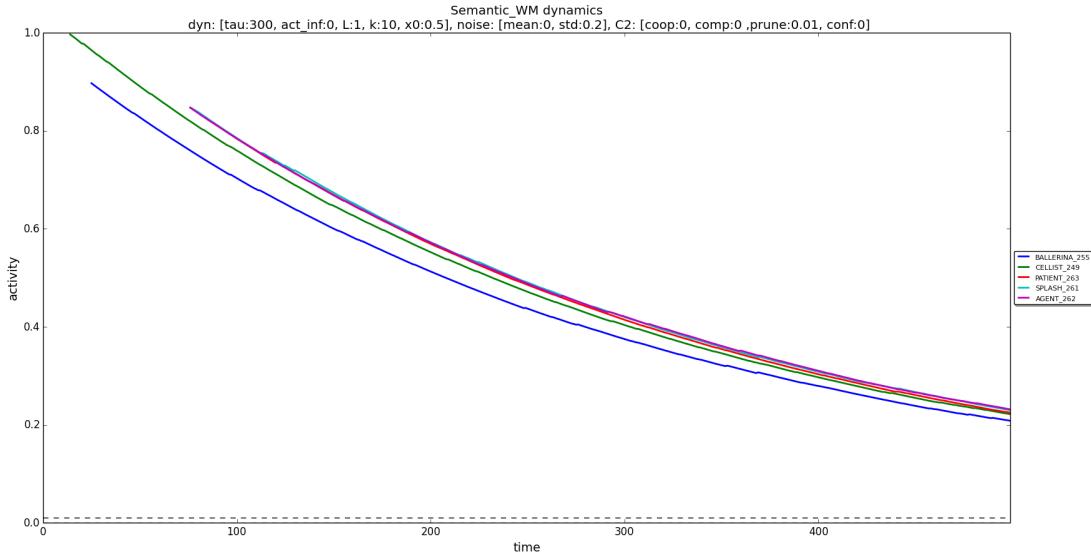


Figure B.39: Concept schema instance dynamics in semantic WM. Following the timing in visual WM, the CELLISTA concept is instantiated first followed by the BALLERINA concept. Their initial activation values reflect the activation values of the percept schemas they conceptualize at the time of instantiation. The difference in activation level between the CELLISTA and the BALLERINA concept instance implements the information structure level of the SemRep with a higher activation value corresponding to a higher "focus" value.

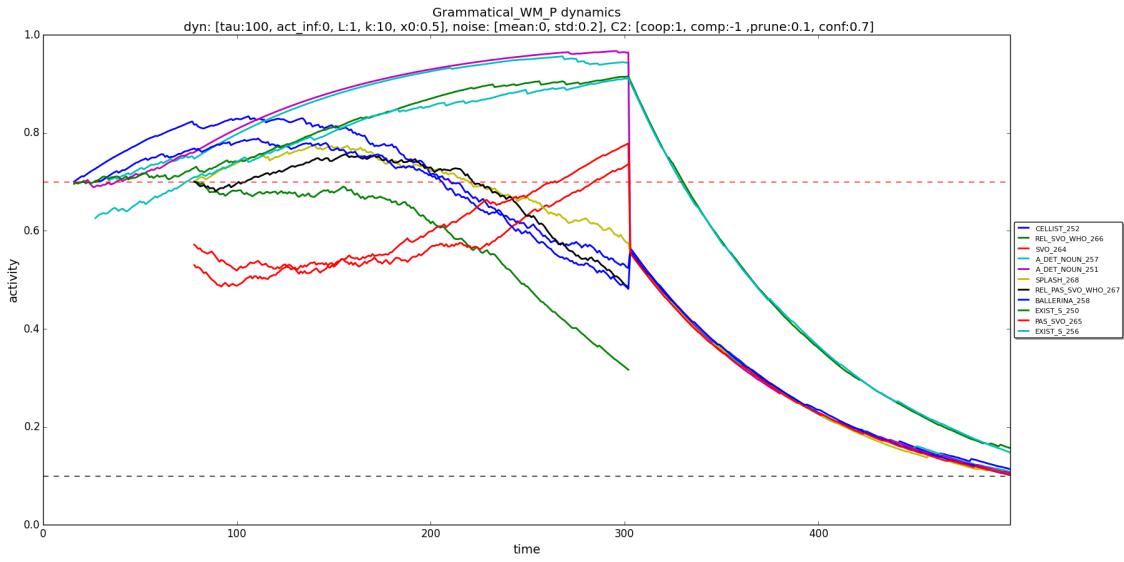


Figure B.40: Construction instances activation in grammatical WM. The main element to look at here are the two green line at the bottom of the graph. The top one represents the PAS.SVO cxn instance while the bottom one represents the SVO cxn instance. In the grammar used for this experiment, both constructions have the same preference value and should therefore start with a similar activation level. The difference in initial activation value results from the fact that the SemFrame of PAS.SVO matches the SemRep better since it favors an Patient focus. This initial boost results in PAS.SVO instances staying above SVO during processing until production is triggered at  $t = 300$ . Consequently, the model will produce an utterance structured by the PAS.SVO cxn and therefore in passive voice.

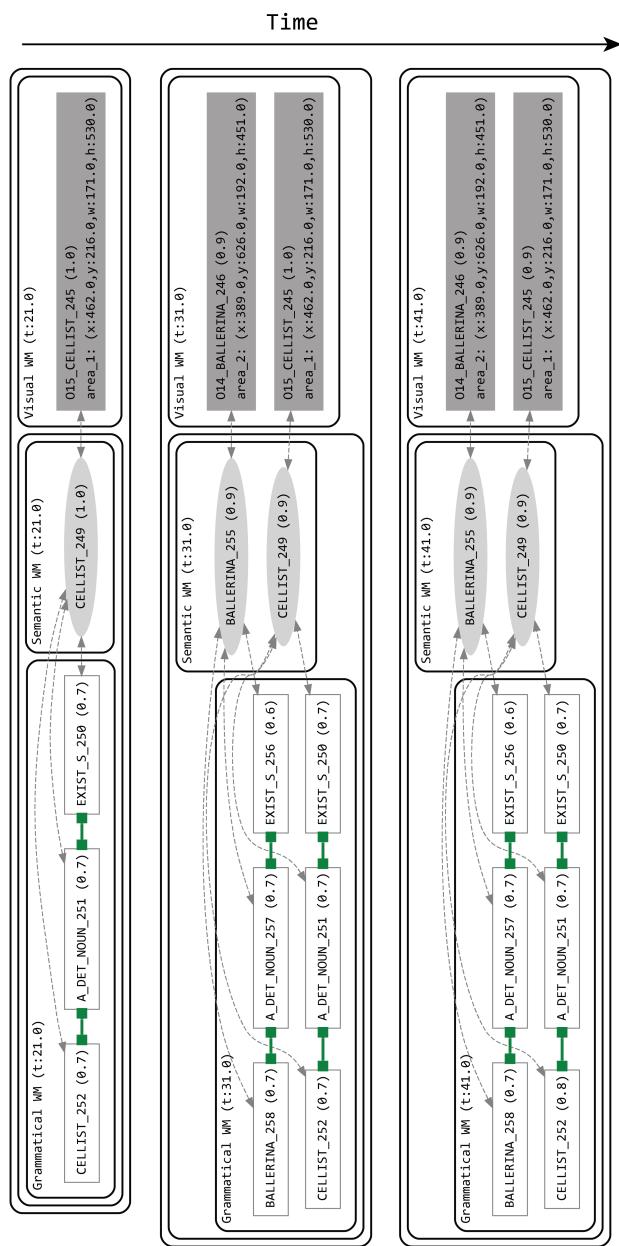


Figure B.41: View of the interactions between visual, semantic and grammatical WM at various time points.

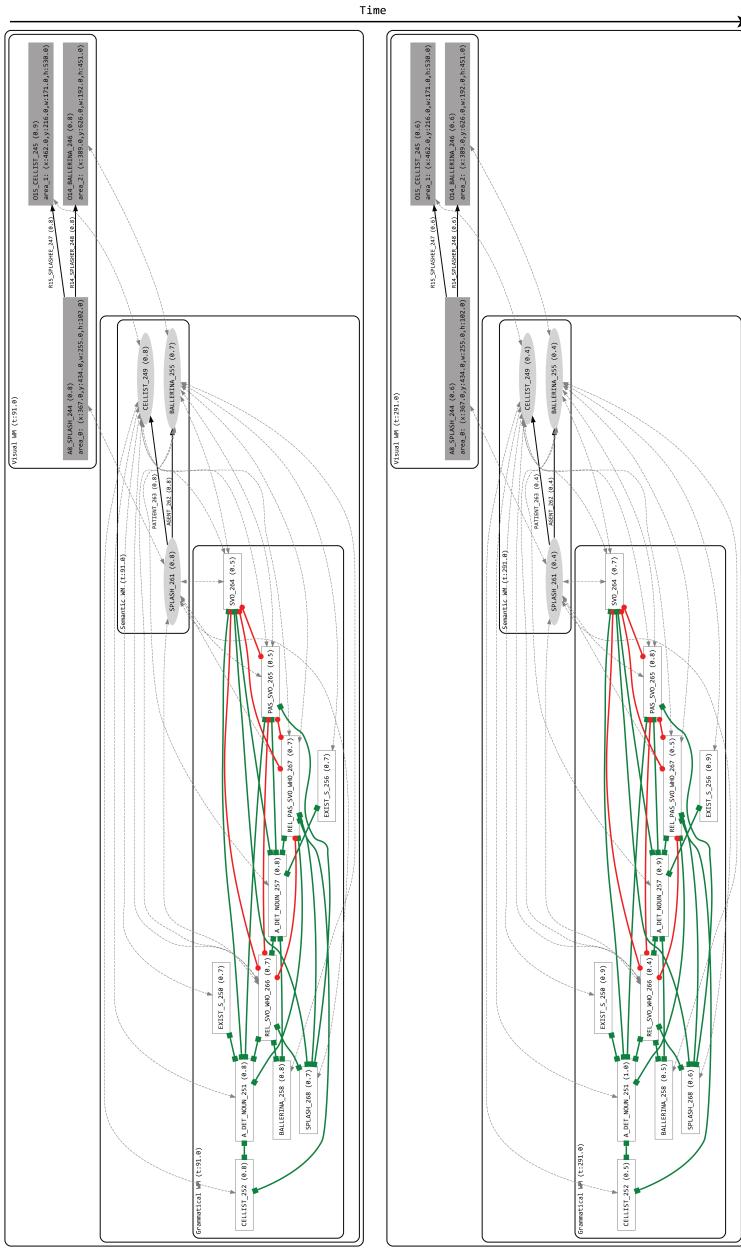


Figure B.42: View of the interactions between visual semantic and grammatical WM at various time points. (continued). The final output of the model is the utterance "a cellist is splash-ed by a ballerina".

## B.5 Scene Builder



Figure B.43: TCG Scene Builder. Web application. A scene image can be selected and loaded (top left). Perceptual schema instances can be added (ADD PERCEPTUAL SCHEMA) by creating a mask over the region hypothesized to be interpreted by the schema instance and then selecting in a drop-down menu the type of percept schema (the selection is loaded based on the model's perceptual knowledge). Saliency and uncertainty (affecting processing time) can be manually defined. For relation percept schema instances, a source and a target instances are selected instead of a region of the visual scene. Percept schema instances can be organized into subscene (BUILD SUBSCENE). Each subscene is a perceptual structure defined as a set of perceptual schemas. (SCENE CONTENT) List of all the percept schema instances defined (left) and of subscenes (right). Each can be expanded, updated, or removed. Upon selecting "Generate Scene Data", the TCG scene builder generates a Json formatted version of the data that can then be used as a model input. Any scene input can be reloaded in the TCG scene builder to be viewed or modified.

# Appendix C

# TCG Production Theory

## C.1 TCG Production Processing Algorithms

In what follows I detail the algorithms supporting the TCG construction processes. I will use the  $X.Y$  dot chaining notation to indicate that  $Y$  is a method or attribute of  $X$ .

For a given construction instance  $cxn\_inst_i$ :

- $\text{covers}_i^{-1}(\text{SemFrame}_i)$  denotes the SemRep subgraphs that this construction instance is expressing (that its SemFrame covers), the inverse simply results from  $\text{covers}_i$  being defined by schema theory convention as a mapping from the element that triggered the instantiation (the SemRep subgraphs) to part of content of the instance that made this instantiation possible (here  $\text{SemFrame}_i$ ).
- $SL_i$  is the symbolic link mapping between nodes in  $\text{SemFrame}_i$  and elements of  $\text{SynForm}_i$  (elements that can either be a Slot or a Phonetic form).
- $\mathcal{D}_{SL_i}$  denotes the set of SemFrame elements that are explicitly symbolized through symbolic links. In the current formalism, any element of the SemFrame that is not in  $\mathcal{D}_{SL_i}$  is considered to be implicitly symbolized by the construction as a whole.

### C.1.1 SemMatch and Construction Invocation

The core computation involved in invoking construction instances in GrammaticalWM rests on finding labeled sub-graph isomorphisms between the current SemRep graph and the SemFrames of the constructions stored in Grammatical LTM. This corresponds to finding the possible variable bindings between the SemRep as the given data structure carrying the semantic information at each time, and the constructions that carry the possible linkages of parts of this data structure onto linguistic form.

It is worth noting that the *is-a* relation used to ensure the concept matching implies here a categorical match. However, other relations including distances can be implemented, in which case the model returns, each possible sub-graph isomorphism associated along with a value reflecting the quality of label (concepts) matching based on the implemented metric.

The processes governing the instantiation of constructions is a direct application of the *SemMatch* algorithm

The *instantiate\_cxn* algorithm (alg. 3) here omits a few operations. The initial activation of a construction instance can be modulated by the relation between the information structure it stipulates (defined in terms of a Focus feature carried by the SemFrame) and the pattern of activation its matches in the SemRep subgraph. Initial activation is raised if the node under Focus is the more activated than the other nodes in the subgraph. In addition, when a SemRep element (node or edge) results in the instantiation of a construction, it is marked as Old. *instantiate\_cxn* therefore, only considered SemRep subgraphs that contain at least one New SemRep element.

Search for subgraph isomorphisms is known to be quite a costly operation. However, the relative small size of the graphs the model is dealing with ensures that the problem remains tractable (Typically, at each time a SemRep contains less than a dozen nodes and SemFrames are usually limited to a few nodes.).

---

**Algorithm 2** *SemMatch*( $S, cxn\_schema$ )

---

**Require:** SemRep  $S$  and a construction schema  $cxn\_schema \in \text{GrammaticalLTM}$ .

```
 $SemFrame \leftarrow cxn\_schema.cxn.SemFrame$ 
 $F \leftarrow \emptyset$ 
for all node induced subgraphs  $s_i$  of  $S$  where at least one node is new do
    #(Check that graph topology matches)
     $f_i \leftarrow find.isomorphism(s, SemFrame)$ 
    if  $f_i$  exists then
        #(Check that the semantic features associated with each node and edge match)
        for all nodes in  $SemFrame$  do
            check  $is\_a(node.concept, f_i(node).concept)$ 
        for all edges in  $SemFrame$  do
            check  $is\_a(edge.concept, f_i(edge).concept)$ 
        if  $f_i$  exists and semantic features match then
             $F \leftarrow F \cup \{f_i\}$ 
return  $F$ 
```

---

---

**Algorithm 3** *instantiate\_cxn*( $S, cxn\_schema$ )

---

**Require:** A semantic representation  $S$  and a  $cxn\_schema \in \text{GrammaticalLTM}$ .

```
 $Insts \leftarrow \emptyset$ 
 $F \leftarrow \text{SemMatch}(S, cxn\_schema)$ 
 $a_0 \leftarrow cxn\_schema.a_0$ 
for all  $f_i \in F$  do
     $id_i \leftarrow$  a unique id
     $a_i \leftarrow a_0$ 
     $covers_i \leftarrow f_i$ 
     $trace_i \leftarrow \{cxn\_schema, f_i(cxn\_schema.SemFrame)\}$ 
     $cxn \leftarrow cxn\_schema$ 
     $cxn\_inst_i \leftarrow (id_i, cxn, a_i, covers, trace)$ 
     $Insts \leftarrow INST \cup \{cxn\_inst_i\}$ 
return  $Insts$ 
```

---

### C.1.2 Match

---

**Algorithm 4** *match( $inst_1, inst_2$ )*

---

**Require:**  $inst_1, inst_2$  two distinct construction instances in grammaticalWM.

```

 $match\_cat \leftarrow 0$ 
 $links \leftarrow \emptyset$ 
 $overlap \leftarrow covers_1^{-1}(SemFrame_1) \cap covers_2^{-1}(SemFrame_2)$ 
#(Check if relation exists)
if  $overlap = \emptyset$  then
     $match\_cat \leftarrow 0$  (NO RELATION)
    return ( $match\_cat, links$ )
#(Check for competition)
if  $overlap \cap Edges \neq \emptyset$  then
     $match\_cat \leftarrow -1$  (MISMATCH)
    return ( $match\_cat, links$ )
else
    for all  $node \in overlap$  do
         $competition \leftarrow comp\_link(inst_1, inst_2, node)$ 
        if  $competition$  then
             $match\_cat \leftarrow -1$ 
            return ( $match\_cat, links$ )
#(Check for cooperation)
for all  $node \in overlap$  do
     $link \leftarrow coop\_link(inst_1, inst_2, node)$ 
    if  $link \neq \text{None}$  then
         $links \leftarrow links \cup \{link\}$ 
     $link \leftarrow coop\_link(inst_2, inst_1, node)$ 
    if  $link \neq \text{None}$  then
         $links \leftarrow links \cup \{link\}$ 
    if  $links \neq \emptyset$  then
         $match\_cat \leftarrow 1$  #(MATCH)
    else
         $match\_cat \leftarrow -1$  #(MISMATCH)
return ( $match\_cat, links$ )

```

---

In the *comp\_link* algorithm (alg 5), *cond1* (and symmetrically *cond2*, stipulate whether or not a given element of the SemFrame  $s_1$  is symbolized by the construction. If  $(s_1 \notin \mathcal{D}_{SL_1})$ , then  $s_1$  is by convention considered to be implicitly symbolized by the construction as a whole. If  $(SL_1(s_1))$  is a phon, then  $s_i$  is explicitly symbolized by the construction through a symbolic mapping onto a lexical form.

Competition therefore occurs each time two constructions attempt to simultaneously symbolize a similar element of the SemRep.

The *coop\_link* algorithm (alg 6 takes a pair of construction instances and a SemRep node on which they overlap and checks whether or not they can collaborate using this node as a “pivot”.

Constructions are invoked through *SemMatch*, and each time a new construction is instantiated in Grammatical WM, it matched against all the construction instances already active. If it overlaps on its coverage of the SemRep with existing construction instances, the construction instance network of cooperation and competition links is updated, with all the links returned by *match* being added to the C2\_network. The C2\_network that forms the basis of the C2 dynamics is therefore incrementally adapted to the state of the SemanticWM.

---

**Algorithm 5** *comp\_link(inst<sub>1</sub>, inst<sub>2</sub>, s)*

---

**Require:** Two construction instances *inst<sub>1</sub>*, *inst<sub>2</sub>* and a SemRep node *s* on which the two constructions overlap.

```

competition  $\leftarrow$  False
#(Find the respective SemFrame nodes that map onto the common SemRep node.)
s1  $\leftarrow$  inst1.covers(s)
s2  $\leftarrow$  inst2.covers(s)
#(Check whether the construction formalize the semantic node.)
cond1  $\leftarrow$  (s1  $\notin$   $\mathcal{D}_{SL_1}$ )  $\parallel$  ( $SL_1(s_1)$  is a phon)
cond2  $\leftarrow$  (s2  $\notin$   $\mathcal{D}_{SL_2}$ )  $\parallel$  ( $SL_2(s_2)$  is a phon)
if cond1 & cond2 then
    competition  $\leftarrow$  True
return competition

```

---



---

**Algorithm 6** *coop\_link(inst<sub>P</sub>, inst<sub>C</sub>, s)*

---

**Require:** Two construction instances *inst<sub>P</sub>*(parent), *inst<sub>C</sub>*(child) and a SemRep node *s* on which the two constructions overlap.

```

#(Find the respective SemFrame nodes that map onto the common SemRep node.)
sP  $\leftarrow$  instP.covers(s)
sC  $\leftarrow$  instC.covers(s)
#(Type constraints – obligagory)
syn1  $\leftarrow$   $SL_P(s_P)$  is a slot (ie is defined and is not a phonological form)
sem1  $\leftarrow$  sC is a Head node
if syn1 = True & sem1 = True then
    #(Metric constraint – qualitative)
    slotP  $\leftarrow$   $SL_P(s_P)$ 
    syn2  $\leftarrow$  instC.cxn.class  $\in$  slotP.cxn_classes
    sem2  $\leftarrow$  sC.concept  $\subset$  sP.concept
    link  $\leftarrow$  (instC, instC:output, instP, instP:slotP) #(inst:x indicates ports)
    if syn2 = True & sem2 = True then
        match_qual  $\leftarrow$  1
    else
        match_qual  $\leftarrow$  0
    return (match_qual, link)
return None

```

---

### C.1.3 Cooperative Computation Dynamics

For a schema instance  $i$ , active in a WM as part of C2 network, its activity  $Act_i^t$  is updated following a leaky integrator equation:

$$Act_i^{t+1} = \alpha Act_i^t + (1 - \alpha)\sigma(Act_i^t + noise^t) \quad (\text{C.1})$$

with  $\alpha$  defining the characteristic time of the WM system  $\alpha = (1 - \tau^{-1})$ ,  $\sigma$  the logistic function, and with a Gaussian noise  $noise^t \sim \mathcal{N}(0, noise_{std})$

$Input_i^t$  is defined as:

$$Input_i^t = w_I \left\{ \sum_{k \in comp(i,k)} w_{comp} \cdot Act_k^t + \sum_{j \in coop(i,j)} w_{coop} \cdot Act_j^t \right\} + \sum_{e \in ext(i)} w_e \cdot Ext_{(e,i)}^t \quad (\text{C.2})$$

$Ext_{(e,i)}^t$  represents activation that an instance  $i$  receives from outside the working memory by subsystem  $e$ .

Here, the competition, cooperation, and external weights are taken to be the same for all instances within a WM <sup>1</sup>.

$w_{ext}$  balances the strength of internal and external activation inputs.  $w_{comp}$  balances the strength of competition and cooperation. The parameters of the logistic function  $\sigma$  are chosen so that, in addition to  $\sigma(\infty) = 1$  and  $\sigma(-\infty) = 0$ ,  $\sigma(0) = Act_{rest}$  the activity in the absence of input. The remaining degree of freedom can be used to set  $\sigma(x_0)' = \sigma_0$  in order to define the steepness of the logistic function. In this case the dynamics of the leaky integrator is defined by the parameters  $(\alpha, A_{rest}, \sigma'_0, w_{comp}, \{w_e\}, noise_{std})$ . In addition,  $\theta_{prune}$  defines the pruning threshold. A constructions whose activation values falls below  $\theta_{prune}$  is pruned out of working memory. Each WM system sets its own set of parameters.

### C.1.4 Construction Schema Instance Assemblage

Construction instances assemblage define a specific class of schema instance assemblage as defined above. Construction assemblages define meaning-form mapping solutions, translating whole or part of the semantic representation into a word sequence. For any given state of the Grammatical WM the set of cooperation links forms a directed graph. Since a construction instance has only a single output port, starting with a construction for which the output port is not connected, it is possible to recursively build the set of all trees that are defined by the network instances that are the children of this top instance. Those trees define the construction instance assemblages. Fig. C.1 provides an informal example of this process.

### C.1.5 Incremental Semantic Representation Format (ISRF)

In order to bypass the visuo-attentional processes, it is possible to directly define the input to the semantic WM using the Incremental Semantic Representation Format (ISRF).

ISRF allows the definition of an incremental semantic input. ISRF syntax contains two primitive formulae:  $F_1$  specifies the assignment of concept instances to variables as:

$$\text{"CONCEPT}_1(var_1, type_1, val_1)"$$

where  $CONCEPT_1$  correspond to an existing concept name in the model's Concept LTM,  $var_1$  is the variable's name to which the concept's instance will be attached, and  $val_1$  is a scalar value in  $[0, 1]$  defining the initial activation value of the concept instance.  $type_1$  can take the value  $F$  to define a *Frame*.<sup>2</sup>

$F_2$  defines the association, for a variable  $var_{rel}$  corresponding to a conceptual relation instance, to their respective origin and target concepts instances associated to the variables  $var_{origin}$  and  $var_{target}$  as:

$$\text{"var}_{rel}(var_{origin}, var_{target})"$$

In ISRF, a proposition  $P$  is defined as a conjunction of formulae of type  $F_1$  or  $F_2$  noted

$$P''_k : [F_{i_1} \& F_{i_2} \& \dots \& F_{i_N}]$$

---

<sup>1</sup>This condition can be relaxed and is not a strong requirement of Schema Theory.

<sup>2</sup>Frames can be used in a way that is similar to most Frame based semantics.

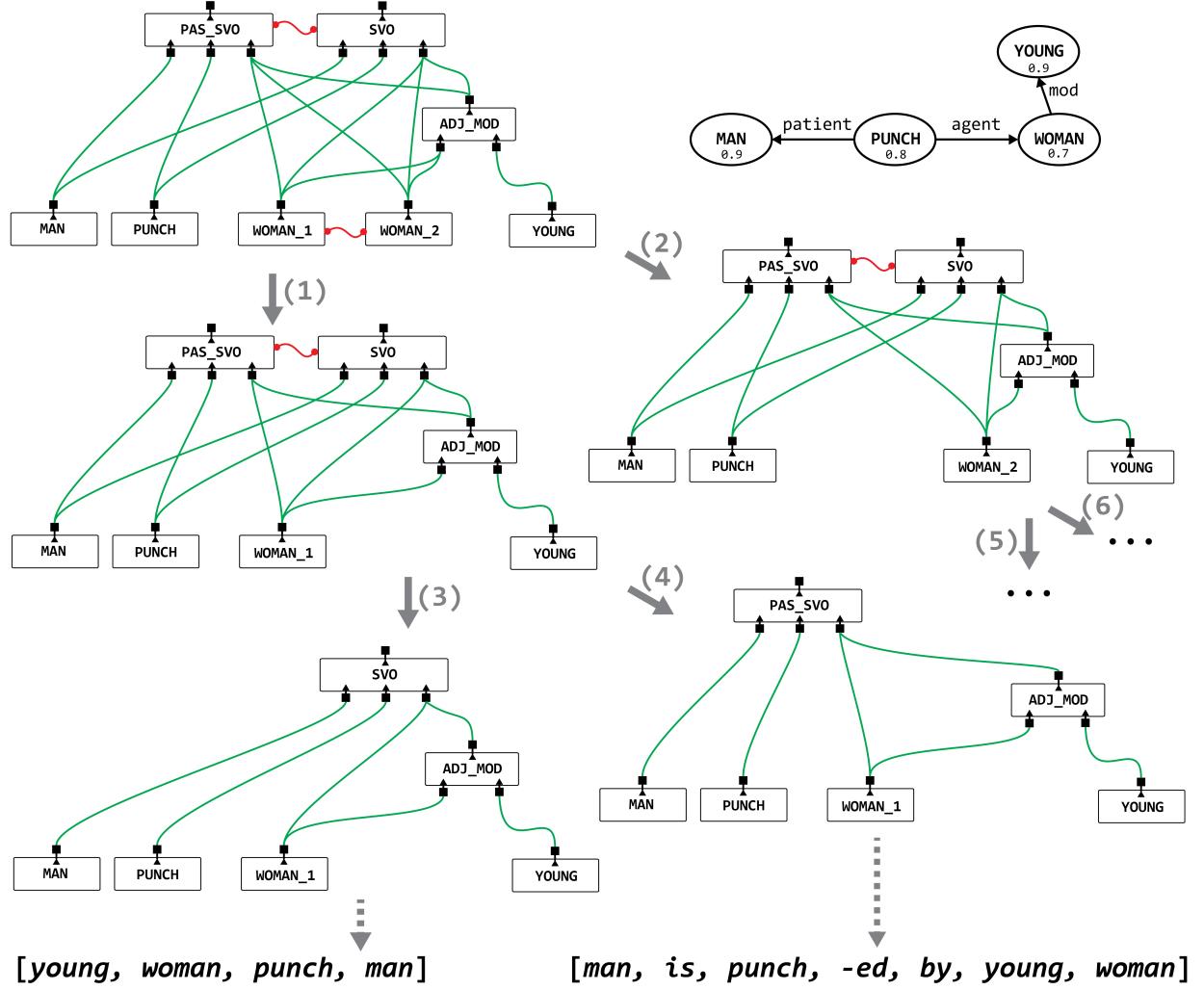


Figure C.1: Examples of construction instance assemblages building from Grammatical WM state. (Top right) The SemRep currently in SemanticWM that the construction instances are attempting to map onto a linguistic form through cooperative computation. (Top left) The state of C2 in Grammatical WM. The situation corresponds to that presented in fig. 2.16. The C2 operations between construction instances has not converged and therefore some competitions remain: competitions between SVO and PAS\_SVO (active vs. passive transitive constructions), competitions between WOMAN\_1 and WOMAN\_2 (two different lexicalizations of the WOMAN concept). In this case, (competing) assemblages are defined as maximal sets of cooperating instances (i.e. not containing any competition links). Assemblages can be constructed based on all the possible sequences of winner choices among the existing competitions. The numbered grey arrows define paths along which such choices are made, ultimately resulting in the set of competing assemblages. (1) and (2) correspond respectively to selecting WOMAN\_1 or WOMAN\_2 as winner. Following (1), only one competition remains between SVO and PAS\_SVO. (3) and (4) correspond respectively to the case of the former and the latter winning the competition. They each lead to a construction instance assemblage (no more competition). (3) maps the SemRep onto the utterance “young woman punch man”, while (4) maps it onto “man is punch-ed by young woman”. The reader can easily derive the similar results of (5) and (6). 4 assemblages can be constructed from the Grammatical WM C2 state and are therefore in competition as meaning-form mappings.

```

"young_woman_punch_man":{
    "sem_rate":10,
    "propositions": {
        "P1":["MAN(man1, 0.9)"],
        "P2":["PUNCH(punch1, 0.8) & HUMAN(human1, 0.7) & AGENT(agt1, 1.0) &
PATIENT(pt1, 1.0) & pt1(punch1, man1) & agt1(punch1, woman1)"],
        "P3":["WOMAN(woman1, 0.7) & IS(is1, 1.0) & is1(human1, woman1)"],
        "P4":["YOUNG(young1, 0.9) & MODIFY(mod1, 1.0) & mod1(woman1, young1)"],
    "sequence": ["P1", "P2", "P3", "P4"]}
}

```

Figure C.2: Example of a (simplified) ISRF input to the Semantic WM. This input stipulates an incremental buildup of the SemRep similar to the one used in the sequence of examples showed in figs 2.15, 2.12,& 2.16. Each proposition stipulates semantic information to be added to the current state of the SemRep. P1 declares that the concept schema MAN should be instantiated with name man1 and activity value 0.9. P2 declares in the same way the instantiation of PUNCH, HUMAN, AGENT, and PATIENT concept schema, the latter two being conceptual relations. It also defines the arguments of those relations. P3 and P4 follow the same pattern, with the addition of the IS relation in P3, used to indicate that a concept schema instance has been updated and became more specific (here HUMAN is specified to be a WOMAN). Sequence defines the order in which the proposition should be interpreted and used as inputs. The rate at which those inputs are passed to Semantic WM is defined by Sem\_Rate defines the rate at which the propositions are triggered as input to the Semantic WM. (ISRF is simplified since, for legibility, it omits EVENT and ENTITY and ACTION frames.)

with  $i \in 1, 2$

Each proposition corresponds to a chunk of semantic information that will be incrementally added to the Semantic WM state. The "*sequence*" define the sequence in which the proposition will be interpreted and used to update the SemanticWM state.

Finally "*sem\_rate*" defines the rate at which the propositions in sequence will be used to incrementally update the SemRep contained in Semantic WM <sup>3</sup>

---

<sup>3</sup>It is also possible to assign a specific time at which each proposition will be interpreted as opposed to using a fixed rate, but for the current work we will only use rate-defined ISRF inputs.

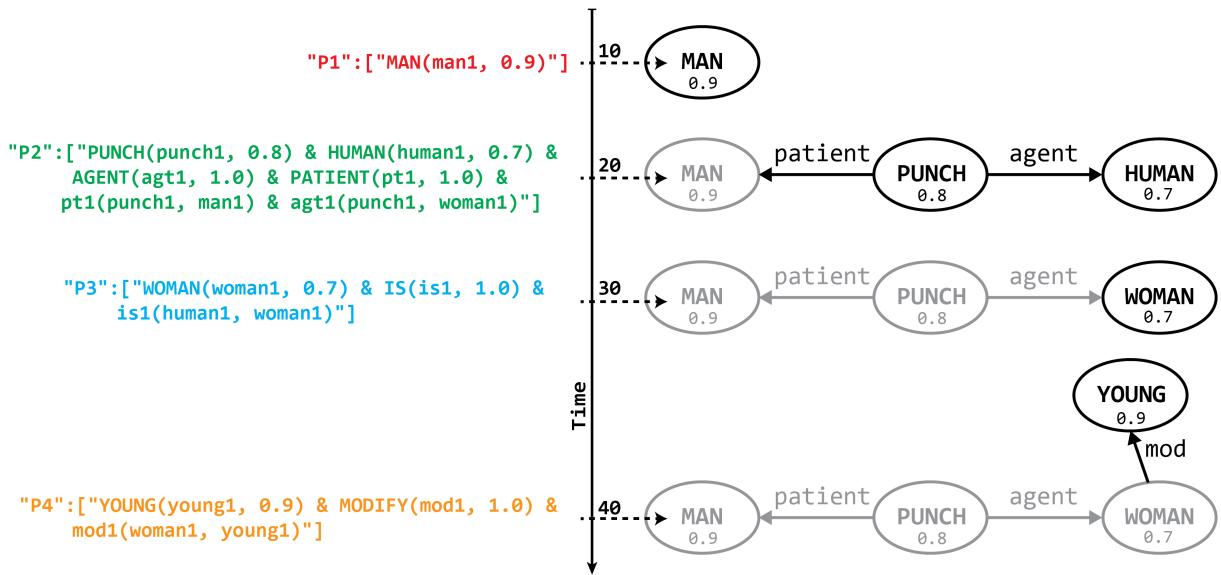


Figure C.3: Example of ISRF proposition inputs to the Semantic WM (left) and the resulting incremental building of the SemRep as the Semantic WM's state (right). The input corresponds the one defined by fig. C.2. The rate of input is set to 10 so a proposition is interpreted and updates the SemRep (state of Semantic WM) every 10 time steps. The propositions are shown on the left. Top-to-bottom defines the time axis. On the right, the state of the SemRep is shown. In black are the SemRep parts that have been updated following the interpretation of the corresponding proposition (links between the two are marked by a dashed arrow). Older part of the SemRep are shown in light grey. Note that for "P3" the "IS" relation is not shown and the concept node HUMAN is simply updated to a MAN concept node. It is assumed here that no decays takes place in Semantic WM ( $\tau_{semWM} = \infty$ ).

# Appendix D

## TCG-SALVIA Comprehension

### D.1 Left-Corner Parser for Context Free Grammars

The following describes the algorithm for left-corner chart parser (LC parser, (Rosenkrantz and Lewis, 1970)) for left-to-right incremental parsing with top-down filtering for context free grammars.

#### D.1.1 Context Free Grammar and Chart Notations

A context free grammar (CFG)  $\mathcal{G}$  is defined as:

$$\mathcal{G} = (V, \Sigma, R, S) \text{ with} \quad (\text{D.1})$$

where  $V$  represents the set of non-terminals,  $\Sigma$  the set of terminals, the projection  $R : V \rightarrow (V \cup \Sigma)^*$  (production rules), and  $S \in V$  the start symbol.

Given a production rule  $p \in R$  with  $p = X \rightarrow Y\delta$ ,  $X$  will be referred to as the ‘mother’ of  $p$ ,  $Y\delta$  as the daughters of  $p$ , and importantly,  $Y$  will be referred to as the left-most ‘daughter’ of  $p$ .

LC algorithm belongs to the class of chart parsing algorithms. A chart  $\mathcal{C}$  is defined as a set of edges. In accordance with the common ‘dot’ notation used for chart parsing, an element of the chart  $\mathcal{C}$  (an edge) will be noted:

$$< A \rightarrow Y.\delta, i, j >$$

Such element refers to the edge of the chart spanning the segment  $[i, j]$  of the chart, associated with the production rule  $A \rightarrow Y\delta$ , and with the dot . marking the boundary between the daughters of the rule that have been consumed (left of the dot) and those that remain to be used (right of the dot).

#### D.1.2 Left-Corner Relation

The left-corner  $LC$  relation in  $V \cup \Sigma \rightarrow \mathcal{P}(V \cup \Sigma)$  can be defined recursively as:

$$X \in LC(A) \Leftrightarrow \begin{cases} X = A \\ \exists p \in R \mid (p = B \rightarrow X\alpha) \wedge (B \in LC(A)) \end{cases} \quad (\text{D.2})$$

(For a given grammar  $\mathcal{G}$ , the  $LC$  relation can be pre-computed)

#### D.1.3 Left-Corner Parser

An incremental left-to-right left-corner parser with top-down filtering can be defined by 5 operations (Moore, 2000).

- op1**  $\forall p \in R$ , if  $S$  mother  $p = S \rightarrow \alpha$  add  $e = < S \rightarrow .\alpha, 0, 0 >$  to chart  $\mathcal{C}$ . (**Top-down driven state initialization**)
- op2**  $\forall (e_1, e_2) \in \mathcal{C}$ , such that  $e_1 = < A \rightarrow \alpha.X\beta, i, k >$  and  $e_2 = < X \rightarrow \gamma., k, j >$ , add  $e_3 = < A \rightarrow \alpha X.\beta, i, j >$  to chart  $\mathcal{C}$ . (**Complete**)
- op3**  $\forall e \in \mathcal{C}$  such that  $e = < A \rightarrow \alpha.a_j\beta, i, j - 1 >$ , if  $a_j \in \Sigma$  is the  $j$ th input, add  $e' = < A \rightarrow \alpha a_j.\beta, i, j >$  to chart  $\mathcal{C}$ . (**Recognize**)
- op4**  $\forall e \in \mathcal{C}$ ,  $e = < X \rightarrow \gamma., k, j >$  completed edge.  $\forall p \in R$ , such  $X$  is the left-most daughter of  $p = B \rightarrow X\delta$ , if  $B \in LC(C)$  (cond1) and  $\exists e' \in \mathcal{C}$  incomplete edge ending at position  $k$  and of the form  $e' = < A \rightarrow \alpha.C\beta, i, k >$  (cond2), then add  $< B \rightarrow X.\delta, k, j >$  to chart  $\mathcal{C}$ . (**Predict** (cond1) with **top-down filtering** (cond2)).
- op5**  $\forall a_j \in \Sigma$  a  $j$ th terminal input,  $\forall p \in R$  with  $a_j$  as left-most daughter, i.e.  $p = B \rightarrow a_j\delta$ , if  $B \in LC(C)$  (cond1) and  $\exists e' \in \mathcal{C}$  incomplete edge ending at position  $j - 1$  of the form  $e' = < A \rightarrow \alpha.C\beta, i, j - 1 >$  (cond2) then add  $< B \rightarrow a_j.\delta, j - 1, j >$  to chart  $\mathcal{C}$ . (**Recognize** and **Predict** (cond1) with **top-down filtering** (cond2)).

Operations (op4) and (op5) are formally equivalent but (5) derives bottom-up predictions (top-down filtered) directly anchored in the linguistic input form, while (4) more generally derives those (filtered) predictions from completed chart edges.

Removing top-down filtering simply consists in removing the top-down driven state initialization (op1), as well as (cond2) in operations (op4) and (op5).

## Appendix E

# SALVIA Comprehension Simulations

All the simulations shown here have been run using a limited grammar in order to keep the computational steps easy to visualize while retaining all the requirements to illustrate the core computational processes.

Articles are shown in the inputs but ignored in processing (no implementation of co-reference resolutions or of quantifiers).

All throughout the simulations, it is worth noting that the number associated with construction instances' names do not matter for the processing, but simply provides them with a unique ID.

### E.1 Simulation 0: Grammatical Route Only

**Input:** “the boy eats the cake in the bedroom”

#### E.1.1 WM states

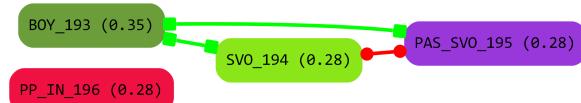


Figure E.1: t=21. Grammatical WM. Input received: “boy...”. The lexical construction instance triggered by BOY\_193 has been invoked. It enters in cooperation (green links) with argument structure constructions instances PAS\_SVO\_195 and SVO\_194 that predict a passive and active voice transitive action respectively. The two are in competition (red links). The PP\_IN\_196 construction instance is also invoked as it could already predict a “in” location complement to the main transitive event. (The predictions are limited due to the limited size of the grammar used in the simulations.)

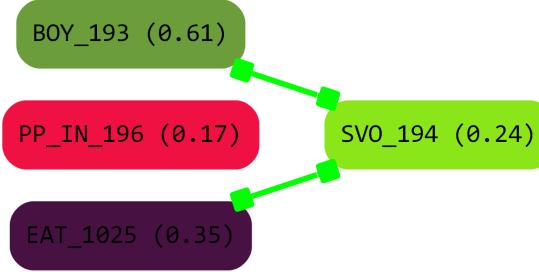


Figure E.2:  $t=71$ . Grammatical WM. Input received “boy eat ...”. EAT\_1025 lexical construction instance is now invoked in Grammatical WM. The SVO\_194 active voice argument structure construction instance has won the competition since the passive voice prediction has been disproved by the linguistic input. The PP\_IN\_196 construction instance’s activation decays since it has not yet found confirmation in the input.

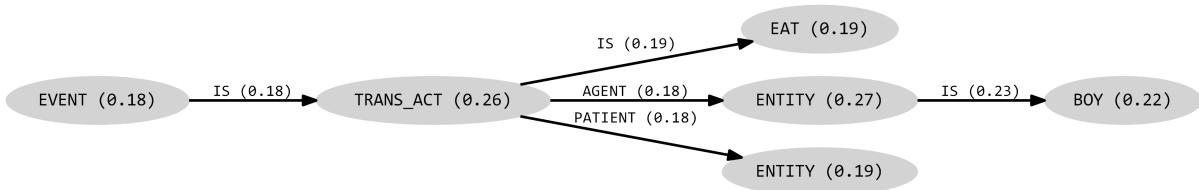


Figure E.3:  $t=101$ . Semantic WM. SemRep generated on the basis of the cooperating construction instances shown in fig. E.2: BOY\_193, SVO\_194, & EAT\_1025. BOY is assigned the role of AGENT of the EAT transitive action. It only predicts that the PATIENT of the action should be an ENTITY. (Compare to the case where World Knowledge is available to refine the prediction).

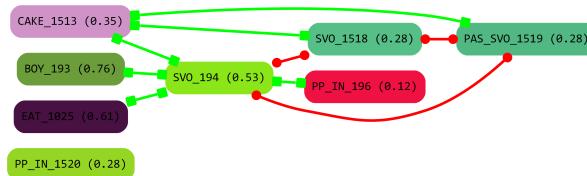


Figure E.4:  $t=121$ . Grammatical WM. Input received “boy eat cake in ...”. Lexical construction instance CAKE\_1513 is invoked and is incorporated in the cooperative assemblage previously used to generate a form-meaning mapping. Upon receiving “in” the PP\_IN\_196 construction instance also becomes part of the assemblage: this prediction triggered upon receiving the first input finally finds confirmation. New SVO, PAS\_SVO and PP\_IN instances are invoked that generate a new set of competitions. Those simply set up low probability predictions that, upon receiving the “cake” input, a new utterance had started (the speaker could have stopped mid-sentence and restarted a new one). Since this won’t be the case, those instances will simply decay in activation until they are pruned out.



Figure E.5:  $t=140$ . Semantic WM. SemRep updated on the basis of the cooperating construction instances shown in fig. E.4. The patient is now specified to be a CAKE, a 'IN' location ENTITY is stipulated but not yet specified.

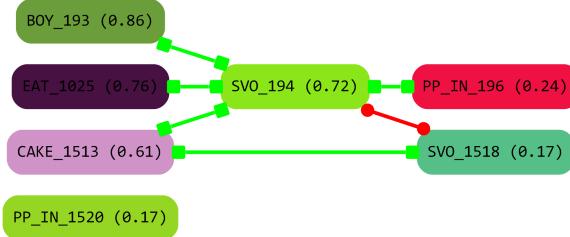


Figure E.6:  $t=171$ . Grammatical WM. No new input since the state shown in fig. E.4. Here the PAS\_SVO instance that had set up a new prediction upon receiving "cake" has already been pruned out.

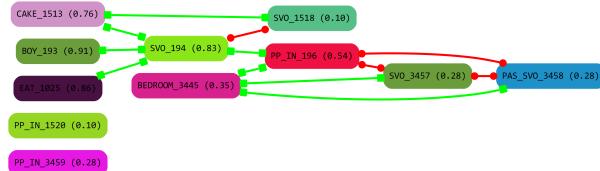


Figure E.7:  $t=221$ . Grammatical WM. Input received "boy eat cake in **bedroom**". Lexical construction instance BEDROOM\_3445 is invoked and is incorporated in the cooperative assemblage previously used to generate a form-meaning mapping. Here again, argument structure predictions are set up, in case a new utterance has been started.

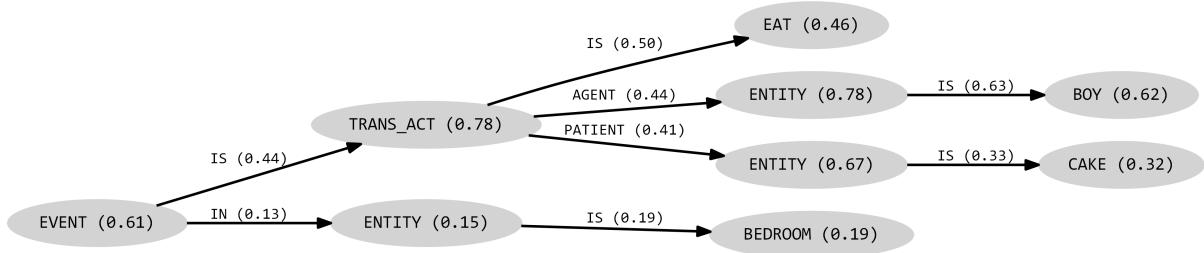


Figure E.8:  $t=221$ . Semantic WM. SemRep updated on the basis of the cooperating construction instances shown in fig. E.7. The location ENTITY positioned in the ‘IN’ role is now specified to be a BEDROOM. The SemRep provides the correct semantic representation (consisting mainly of thematic role assignments) for the input utterance “(the) boy eat(s) (the) cake in (the) bedroom”.

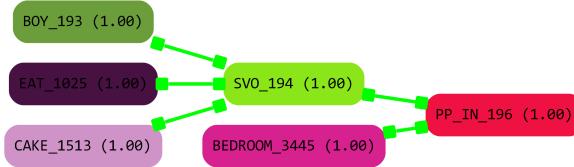


Figure E.9:  $t=891$ . Grammatical WM. Final winning construction instance assemblage.

### E.1.2 Simulation Summaries

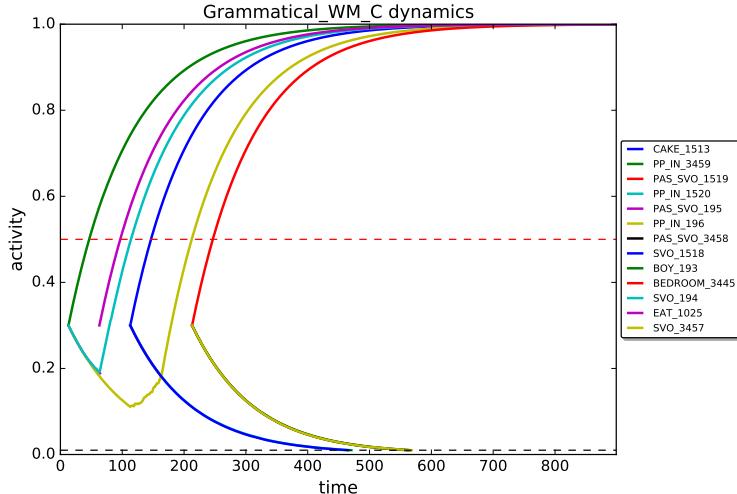


Figure E.10: Activation levels of the construction instances active in Grammatical WM as a function of time. Top dashed line corresponds to the confidence threshold. Bottom dashed line marks to the pruning threshold. See above for representations of the Grammatical WM states over time.

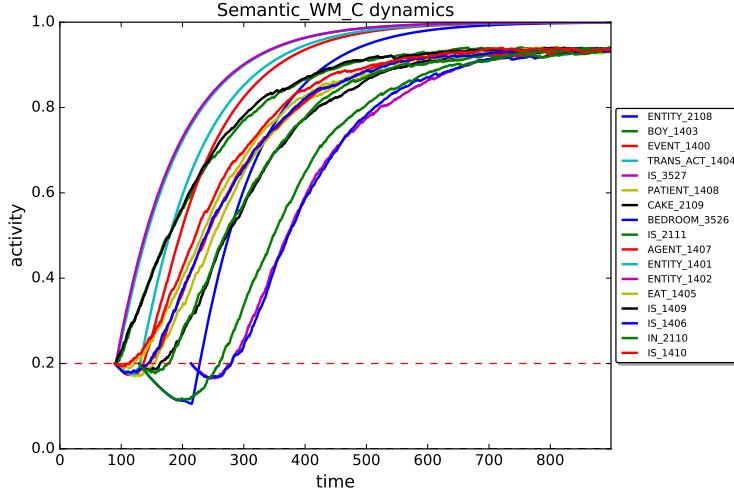


Figure E.11: Activation levels of the concept instances active in Semantic WM as a function of time. Dashed line marks to the pruning threshold. See above for representations of the Semantic WM states over time.

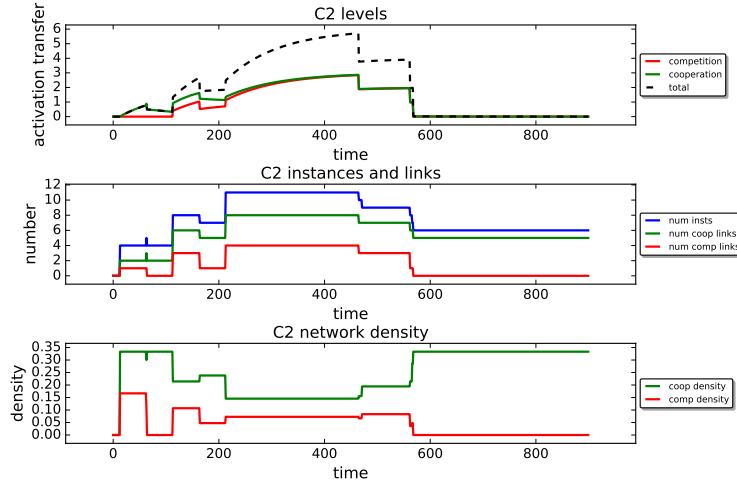


Figure E.12: Higher level analyses of the C2 dynamics taking place in the Grammatical WM through the simulation. Top: C2 levels shows the sum of the activation signals that passes through the C2 network as a function of time (total, dashed line), and also broken down between activation passed through cooperation and competition links (green and red respectively). Middle: Number of instances and links active at each time. Bottom: C2 network density as a function of time (separated into density of cooperation network and competition network).

## E.2 Simulation 1: Grammatical Route Only

**Input:** “the boy is eat -ed by the cake” (counterfactual input)

### E.2.1 WM states

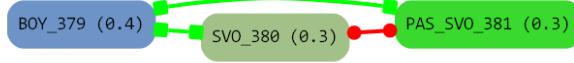


Figure E.13:  $t=21$ . Grammatical WM. Input received: “boy...”. The lexical construction instance triggered by BOY\_379 has been invoked. it enters in cooperation (green links) with argument structure constructions instances PAS\_SVO\_381 and SVO\_380 that predict a passive and active voice transitive action respectively. The two are in competition (red links). (The predictions are limited due to the limited size of the grammar used in the simulations.)

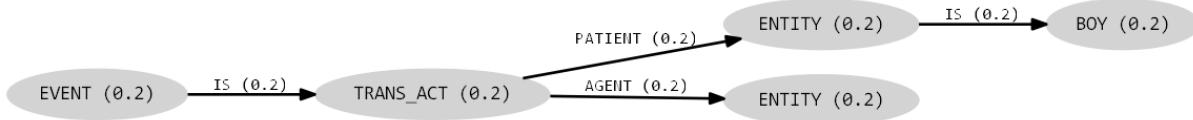


Figure E.14:  $t=112$ . Semantic WM. Input received: “boy **is** ...”. SemRep generated on the basis of the cooperating construction instances shown in fig. E.13: BOY\_379 & PAS\_SVO\_381. This assemblage is chosen since PAS\_SVO\_381 is now winning over SVO\_380 since “is” supports the passive voice hypothesis for a transitive action and not the active voice hypothesis (predicate construction here is not considered for simplicity). BOY is assigned the role of PATIENT of a still to be specified transitive action. It only predicts that the AGENT of the action should be an ENTITY. (Compare to the case where World Knowledge is available to refine the prediction).

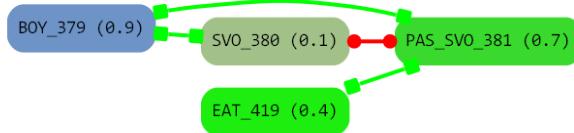


Figure E.15:  $t=221$ . Grammatical WM. Input received “boy is eat -ed by ...”. EAT\_419 lexical construction instance is now invoked in Grammatical WM. The PAS\_SVO\_381 passive voice argument structure construction instance is winning the competition since it has gathered much more support than the active voice construction (this is reflected in their respective activation levels).

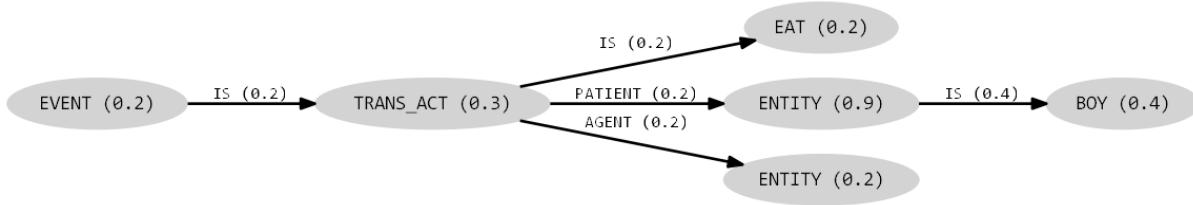


Figure E.16:  $t=312$ . Semantic WM. SemRep updated on the basis of the cooperating construction instances shown in fig. E.15: BOY\_380, PAS\_SVO\_381, & EAT\_419. Compared to the state shown in fig. E.14, the event is now specified to involve an EAT action.

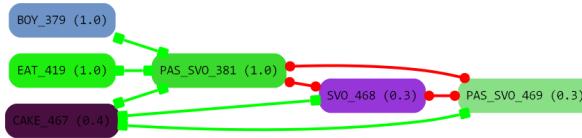


Figure E.17:  $t=521$ . Grammatical WM. Input received "boy is eat -ed by cake". Lexical construction instance CAKE\_494 is invoked and is incorporated in the cooperative assemblage previously used to generate a form-meaning mapping. New SVO and PAS\_SVO instances are invoked that generate a new set of competitions. Those simply set up low probability predictions that, upon receiving the "cake" input, a new utterance had started (the speaker could have stopped mid-sentence and restarted a new one). Since this won't be the case, those instances will simply decay in activation until they are pruned out.

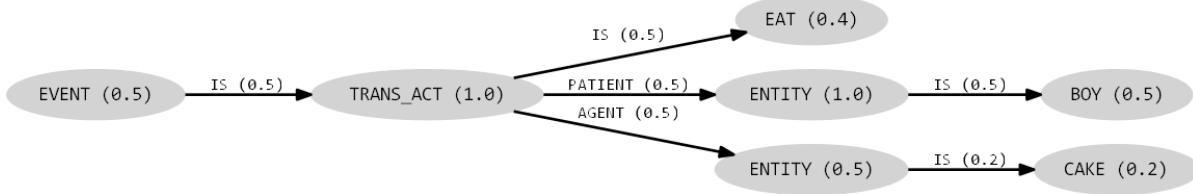


Figure E.18:  $t=612$ . Semantic WM. SemRep updated on the basis of the cooperating construction instances shown in fig. E.17. The AGENT ENTITY is now specified to be a CAKE. The SemRep provides the correct semantic representation (consisting mainly of thematic role assignments) for the counterfactual input utterance "(the) boy is eat -ed by (the) cake".

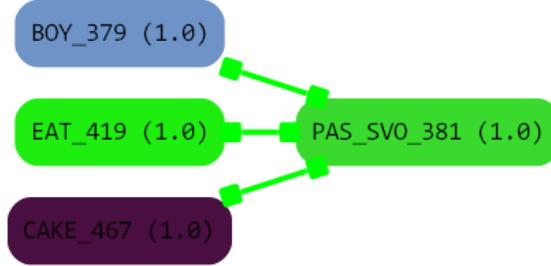


Figure E.19: t=991

### E.2.2 Simulation Summaries

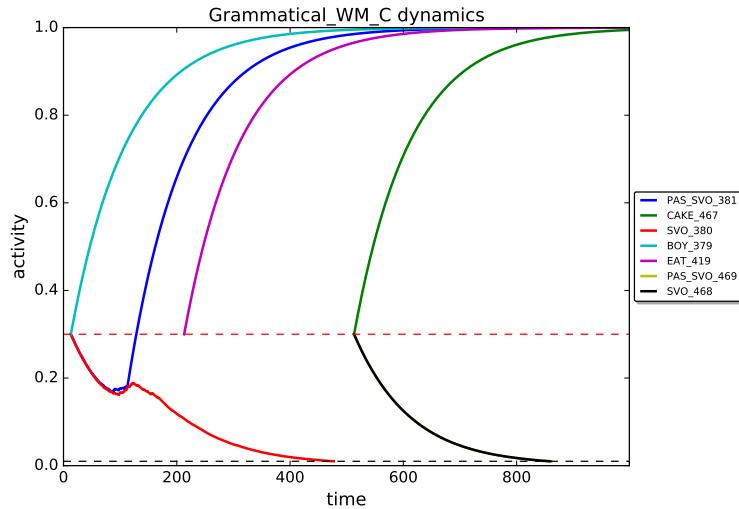


Figure E.20: Activation levels of the construction instances active in Grammatical WM as a function of time. Top dashed line corresponds to the confidence threshold. Bottom dashed line marks to the pruning threshold. See above for representations of the Grammatical WM states over time.

## E.3 Simulation 2: World Knowledge Route Only (+ Lexical Constructions)

**Input:** “the boy is eat -ed by the cake” (counterfactual input)

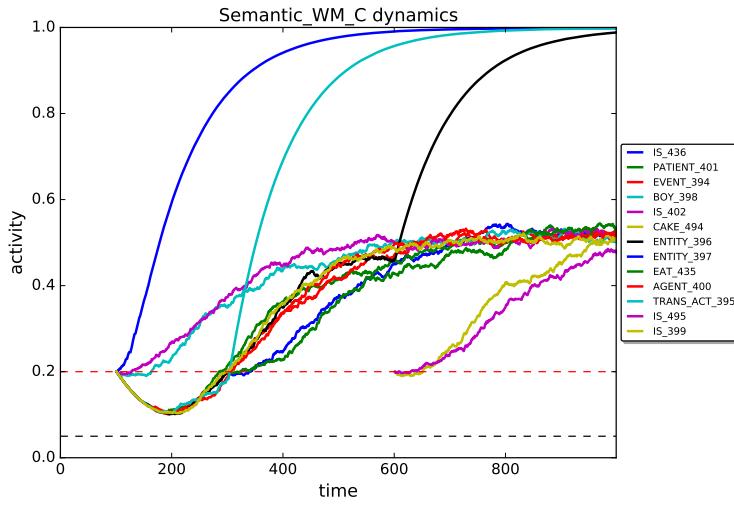


Figure E.21: Activation levels of the concept instances active in Semantic WM as a function of time. Dashed line marks to the pruning threshold. See above for representations of the Semantic WM states over time.

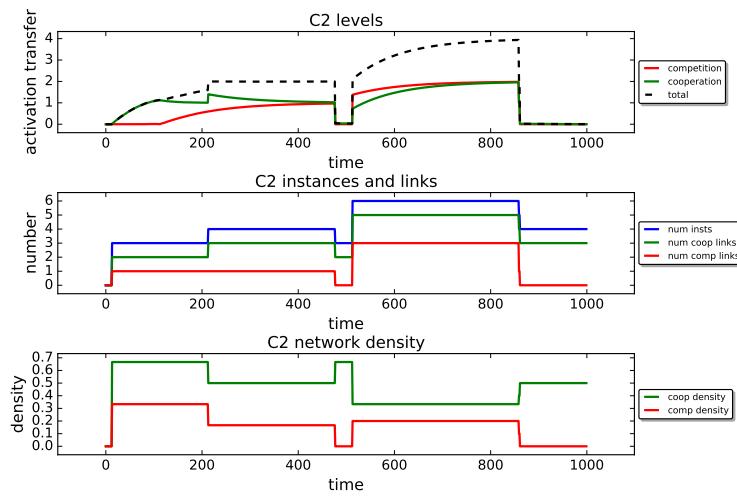


Figure E.22: Higher level analyses of the C2 dynamics taking place in the Grammatical WM through the simulation. Top: C2 levels shows the sum of the activation signals that passes through the C2 network as a function of time (total, dashed line), and also broken down between activation passed through cooperation and competition links (green and red respectively). Middle: Number of instances and links active at each time. Bottom: C2 network density as a function of time (separated into density of cooperation network and competition network).

### E.3.1 WM states

BOY\_381 (0.4)

Figure E.23: t=21. Grammatical WM. Input received: “boy...”. Only the lexical construction instance triggered by “boy”, BOY\_381, has been invoked. Here, no argument structure constructions are invoked to start building a (predictive) grammatical form-meaning mapping cooperative structure.

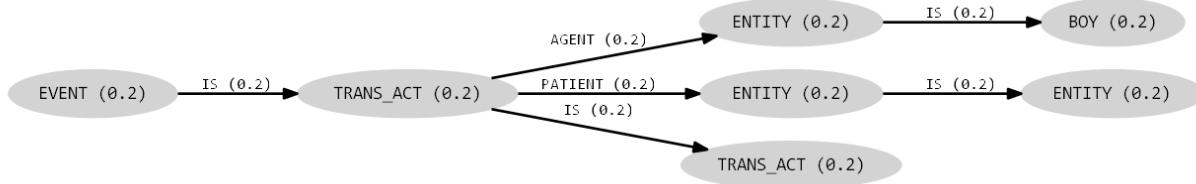


Figure E.24: t=79. Semantic WM. SemRep generated both by the lexical construction instance BOY\_381, only instance active in Grammatical WM and by the state of the World Knowledge WM. “boy” input as triggered the invocation of the “animate agent” frame schema instance in World Knowledge WM. This instance predicts the event frame seen in the SemRep here. It predicts that BOY is likely to be the AGENT of a transitive action event.

BOY\_381 (0.8)

EAT\_407 (0.4)

Figure E.25: t=221. Grammatical WM. Input received: “boy **is eat -ed by...**”. Only “eat” triggers the invocation of a new lexical construction instance EAT\_407. Here again no argument structure constructions are invoked to build a (predictive) grammatical form-meaning mapping cooperative structure. Grammatical cues in the linguistic input are ignored.

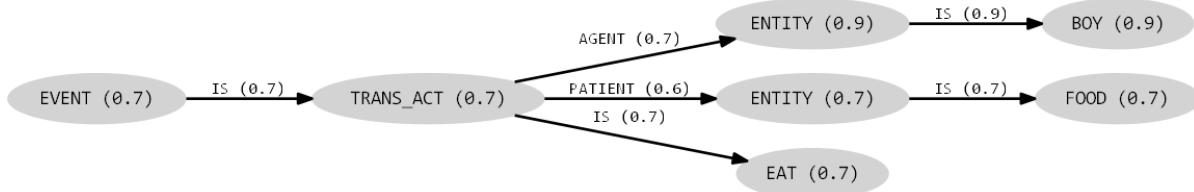


Figure E.26: t=277. Semantic WM. The SemRep is updated by the EAT\_407 construction instance that specifies the action. But much more importantly here, “eat” triggers the invocation of the EAT frame schema instance in World Knowledge WM that stipulates that an “eating event” usually involves an animate agent and a FOOD patient. Therefore here, the patient is already predicted to be a FOOD type. Boy is still placed in the agent role, a hypothesis now comforted both by the “animate agent” and the “eat” frame schema instances, but that reflect the fact that grammatical cues (supporting the passive voice, and hence placing BOY as patient) have not been incorporated into the form-meaning mappings generated.

BOY\_381 (0.9)

EAT\_407 (0.9)

CAKE\_428 (0.3)

Figure E.27: t=521. Grammatical WM. Input received: “boy is eat -ed by **cake**...”. Only “cake” triggers the invocation of a new lexical construction instance CAKE\_428. Here again no argument structure constructions are invoked to build a (predictive) grammatical form-meaning mapping cooperative structure. Grammatical cues in the linguistic input are ignored.

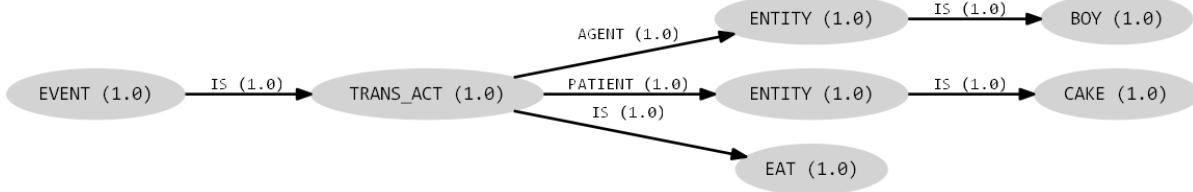


Figure E.28: t=912. Semantic WM. The SemRep is updated by the CAKE\_428 construction instance that specifies the patient since it matches the FOOD patient expectation! The grammatical cues of the passive voice placing CAKE in the agent role have not been used to build the form-meaning mapping. Only world knowledge expectations have guided the thematic role assignment process. The SemRep provides an incorrect semantic representation (reversed thematic role assignment) for the counterfactual input utterance “(the) boy is eat -ed by (the) cake”.

### E.3.2 Simulation Summaries

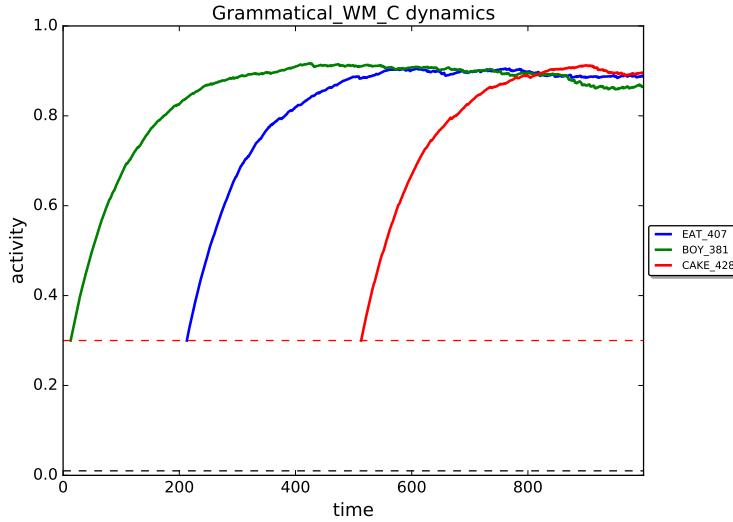


Figure E.29: Activation levels of the construction instances active in Grammatical WM as a function of time. Top dashed line corresponds to the confidence threshold. Bottom dashed line marks to the pruning threshold. See above for representations of the Grammatical WM states over time.

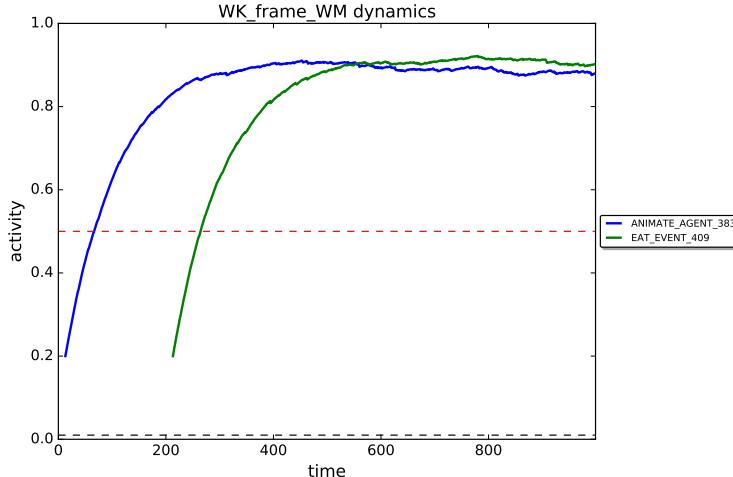


Figure E.30: Activation levels of the world knowledge (event) frame instances active in World Knowledge WM as a function of time. Top dashed line corresponds to the confidence threshold. Bottom dashed line marks to the pruning threshold. See above for a discussion of the World Knowledge WM states over time.

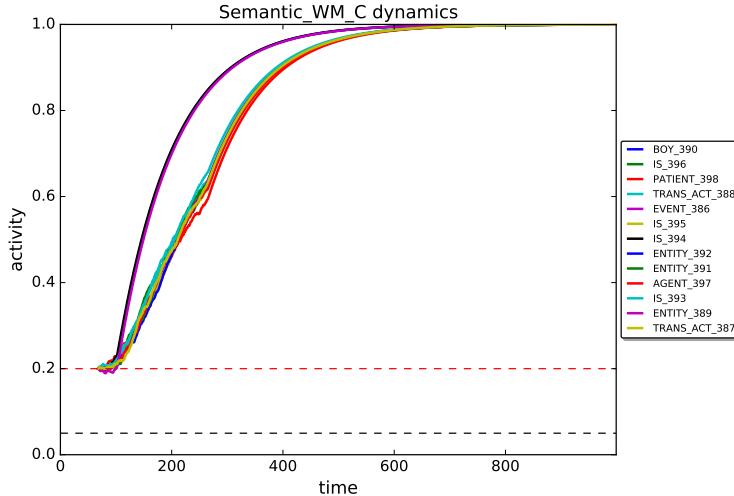


Figure E.31: Activation levels of the concept instances active in Semantic WM as a function of time. Dashed line marks to the pruning threshold. See above for representations of the Semantic WM states over time.

## E.4 Simulation 3: Cooperation Between Routes

**Input:** “the boy eats the cake”

### E.4.1 WM states

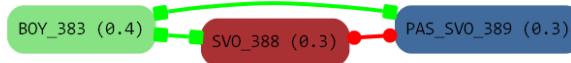


Figure E.32: t=21. Grammatical WM. Input received: “boy...”. The lexical construction instance triggered by BOY\_383 has been invoked. It enters in cooperation (green links) with argument structure constructions instances PAS\_SVO\_389 and SVO\_388 that predict a passive and active voice transitive action respectively. The two are in competition (red links). (The predictions are limited due to the limited size of the grammar used in the simulations.)

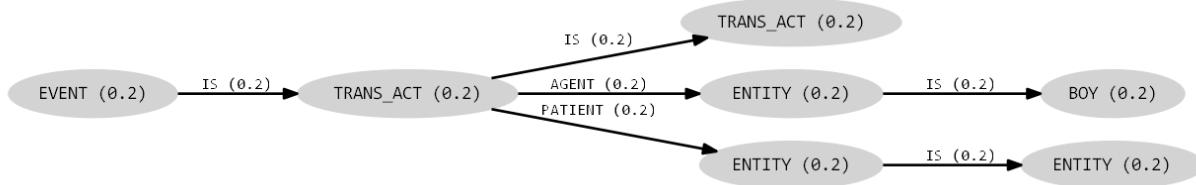


Figure E.33: t=77. Semantic WM. SemRep generated both by the Grammatical Route through the construction instance assemblage comprised of the BOY\_383 and SVO\_388, and by the World Knowledge route through the “animate agent” frame schema instance that has been invoked in World Knowledge WM as a result of the “boy” input. Both routes predict that BOY is likely to be the AGENT of a transitive action event. Therefore the SemRep concept instance receive activation from both the Grammatical and World Knowledge WM, that therefore cooperate and reinforce each other.

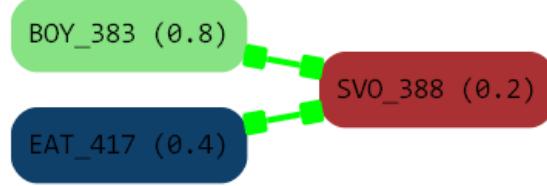


Figure E.34:  $t=121$ . Grammatical WM. Input received “boy eat ...”. EAT\_417 lexical construction instance is now invoked in Grammatical WM. The SVO\_388 active voice argument structure construction instance has won the competition since the passive voice prediction has been disproved by the linguistic input.

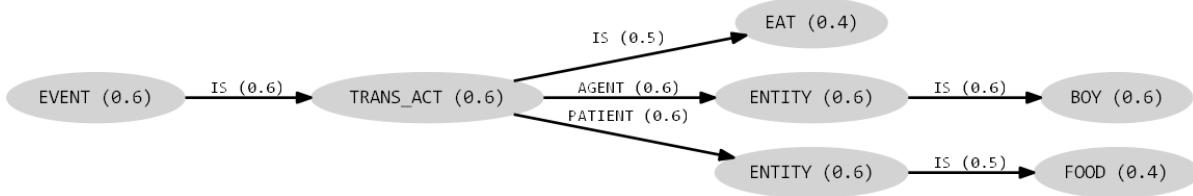


Figure E.35:  $t=177$ . Semantic WM. SemRep updated by both routes. “eat” triggered the invocation of the EAT frame schema instance in World Knowledge WM that stipulates that an “eating event” usually involves an animate agent and a FOOD patient. Therefore here, the patient is already predicted to be a FOOD type. The Grammatical WM has also already processed the “eat” input updated the construction instance assemblage’s activation, specifying the nature of the action as EAT. Here again, both routes support the concept schema instances forming the SemRep.

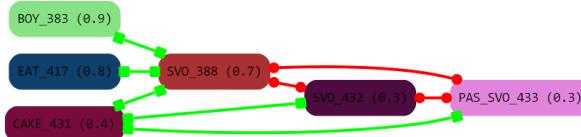


Figure E.36:  $t=221$ . Grammatical WM. Input received “boy eat cake”. Lexical construction instance CAKE\_431 is invoked and is incorporated in the cooperative assemblage previously used to generate a form-meaning mapping. New SVO and PAS\_SVO instances are invoked that generate a new set of competitions. Those simply set up low probability predictions that, upon receiving the “cake” input, a new utterance had started (the speaker could have stopped mid-sentence and restarted a new one). Since this won’t be the case, those instances will simply decay in activation until they are pruned out.

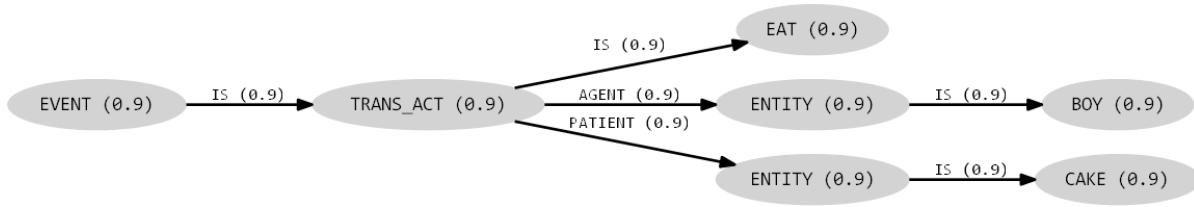


Figure E.37:  $t=312$ . Semantic WM. SemRep updated by the Grammatical WM. The CAKE\_431 construction instance specifies the FOOD patient that had been predicted by the World Knowledge route. Both grammatical and world knowledge have cooperated in guiding the thematic role assignment process. The SemRep provides the correct semantic representation for the input utterance “(the) boy eat(s) (the) cake”.

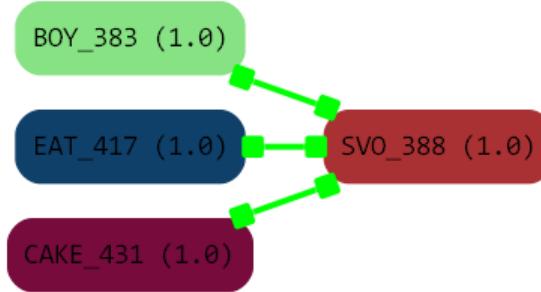


Figure E.38:  $t=991$ . Grammatical WM. Final winning construction instance assemblage.

#### E.4.2 Simulation Summaries

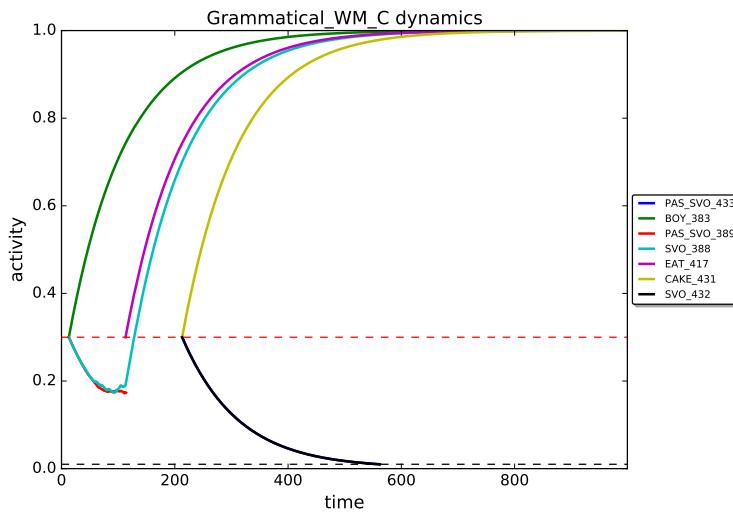


Figure E.39: Activation levels of the construction instances active in Grammatical WM as a function of time. Top dashed line corresponds to the confidence threshold. Bottom dashed line marks to the pruning threshold. See above for representations of the Grammatical WM states over time.

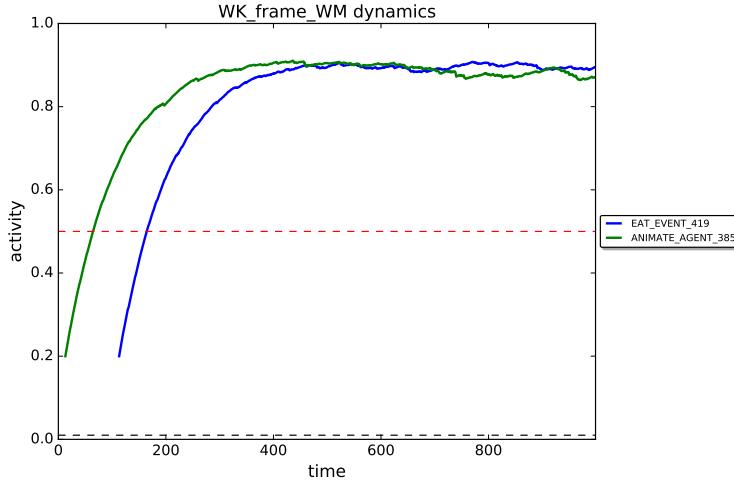


Figure E.40: Activation levels of the world knowledge (event) frame instances active in World Knowledge WM as a function of time. Top dashed line corresponds to the confidence threshold. Bottom dashed line marks to the pruning threshold. See above for a discussion of the World Knowledge WM states over time.

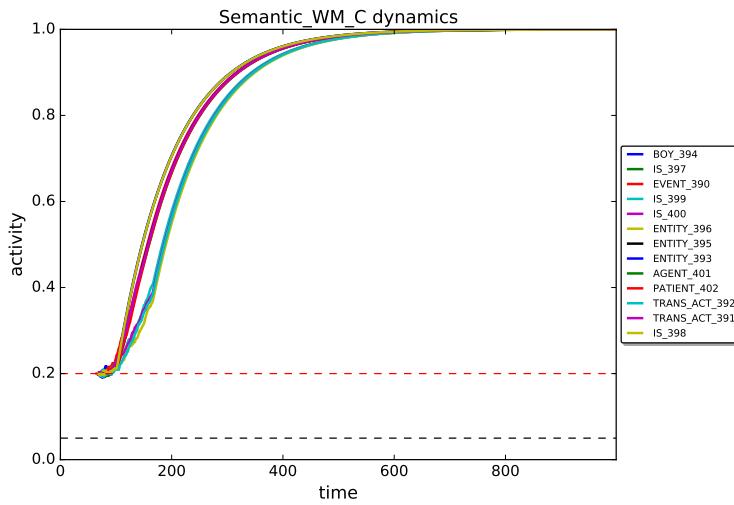


Figure E.41: Activation levels of the concept instances active in Semantic WM as a function of time. Dashed line marks to the pruning threshold. See above for representations of the Semantic WM states over time.

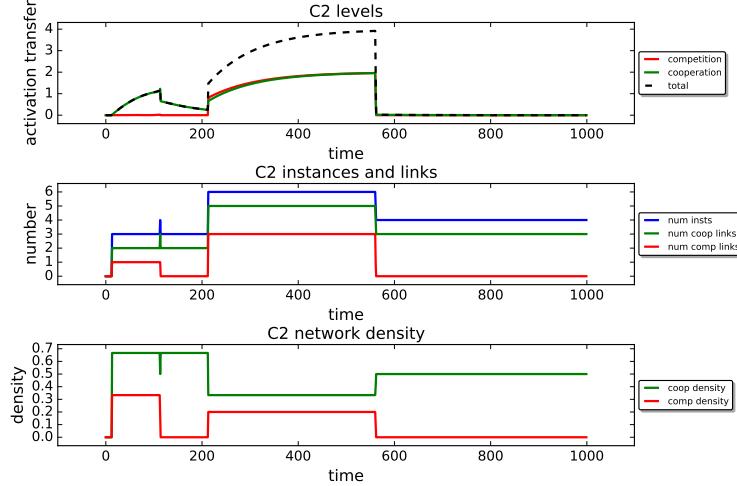


Figure E.42: Higher level analyses of the C2 dynamics taking place in the Grammatical WM through the simulation. Top: C2 levels shows the sum of the activation signals that passes through the C2 network as a function of time (total, dashed line), and also broken down between activation passed through cooperation and competition links (green and red respectively). Middle: Number of instances and links active at each time. Bottom: C2 network density as a function of time (separated into density of cooperation network and competition network).

## E.5 Simulation 4: Competition Between Routes

### E.5.1 WM states

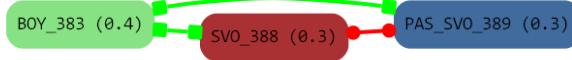


Figure E.43:  $t=21$ . Grammatical WM. Input received: “boy...”. The lexical construction instance triggered by BOY\_383 has been invoked. it enters in cooperation (green links) with argument structure constructions instances PAS\_SVO\_389 and SVO\_388 that predict a passive and active voice transitive action respectively. The two are in competition (red links). (The predictions are limited due to the limited size of the grammar used in the simulations.) The situation is exactly similar to that of fig. E.32

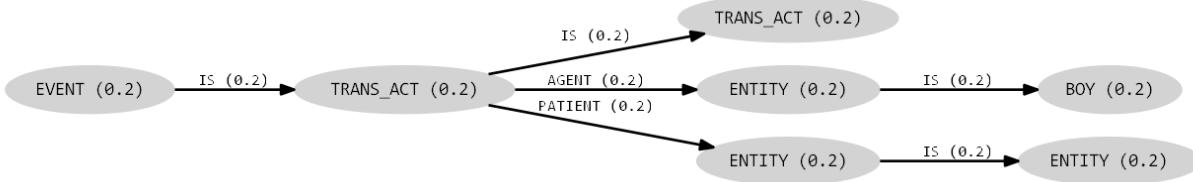


Figure E.44:  $t=77$ . Semantic WM. SemRep generated by the World Knowledge Route only. The Grammatical route has stated to process the input but has not yet reached its confidence level required for it to update the state of the Semantic WM. The “animate agent” frame schema instance has been invoked in World Knowledge WM as a result of the “boy” input. It predicts that BOY is likely to be the AGENT of a transitive action event.

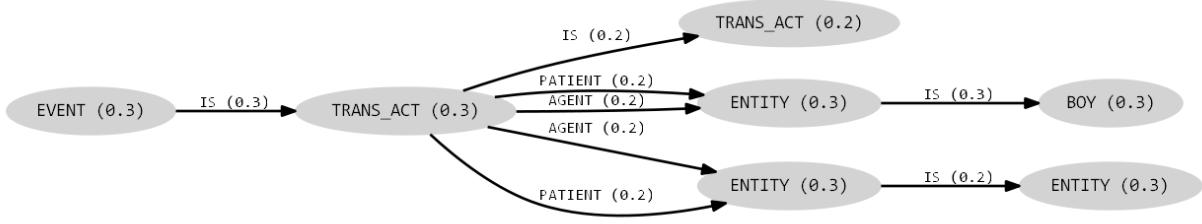


Figure E.45:  $t=112$ . Semantic WM. The SemRep is now updated by the Grammatical route through the construction instance assemblage comprised of the BOY\_383 and PAS.SVO\_389. This assemblage map the form onto a meaning in which BOY is set to be the PATIENT of a transitive action. The two route are therefore in **competition** since they each propose opposite thematic role assignment! This is reflected in the SemRep that is now a multigraph. Where the World Knowledge route had generated a AGENT conceptual relation (edge) for the BOY entity, the Grammatical route adds a parallel PATIENT edge (and conversely the Grammatical route adds an AGENT relation where the World Knowledge had placed a PATIENT one). Those multi-edges in the SemRep reflects **competing concept relation schemas** (competition links are not shown). it is worth noting that there is no guarantee that if the AGENT wins one of the competition, then the PATIENT necessarily needs to win the other one. This reflects the fact that the interpretation need not be coherent (“good enough comprehension” paradigm.)

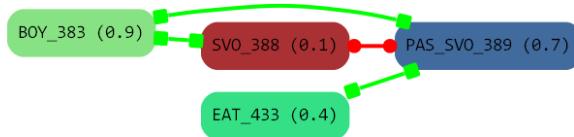


Figure E.46:  $t=221$ . Grammatical WM. Input received “boy is eat -ed by ...”. EAT\_433 lexical construction instance is now invoked in Grammatical WM. The PAS\_SVO\_389 passive voice argument structure construction instance is winning the competition since it has gathered much more support than the active voice construction (this is reflected in their respective activation levels).

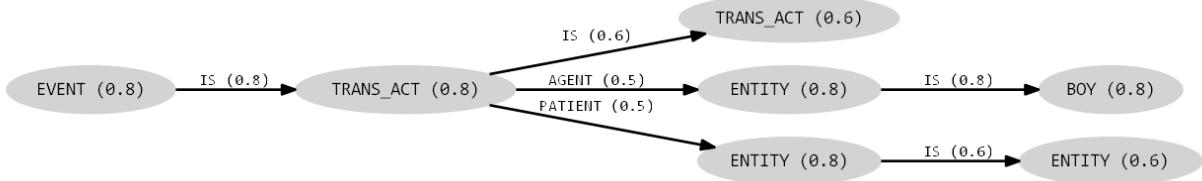


Figure E.47:  $t=253$ . Semantic WM. Compared with the situation in fig. E.45, the competitions have both yielded a winner that in this case reflect the thematic role assignment supported by the world knowledge route. This is due to the fact that in this case the weight of the world knowledge route has been set to a value bigger than that of the grammatical route.

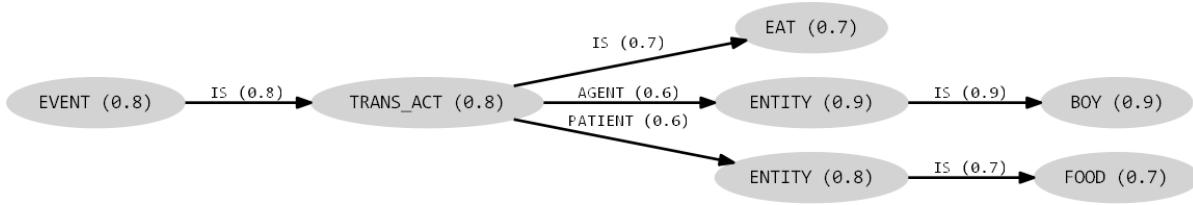


Figure E.48:  $t=278$ . Semantic WM. SemRep updated by both routes. “eat” triggered the invocation of the EAT frame schema instance in World Knowledge WM that stipulates that an “eating event” usually involves an animate agent and a FOOD patient. Therefore here, the patient is predicted to be a FOOD type. The Grammatical WM has also already processed the “eat” input updated the construction instance assemblage’s activation, specifying the nature of the action as EAT. The interpretation of the SemRep is supported by both routes everywhere except for the AGENT and PATIENT concept relation instances that are only supported by the world knowledge route.

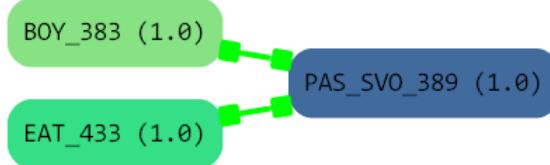


Figure E.49:  $t=512$ . Grammatical WM. Compared with fig. E.46, PAS\_SVO\_389 won.

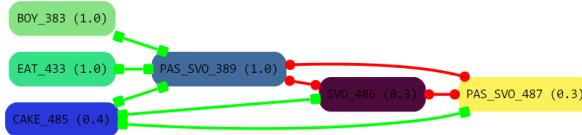


Figure E.50:  $t=521$ . Grammatical WM. Input received “boy is eat -ed by cake”. Lexical construction instance CAKE\_485 is invoked and is incorporated in the cooperative assemblage previously used to generate a form-meaning mapping. New SVO and PAS\_SVO instances are invoked that generate a new set of competitions. Those simply set up low probability predictions that, upon receiving the “cake” input, a new utterance had started (the speaker could have stopped mid-sentence and restarted a new one). Since this won’t be the case, those instances will simply decay in activation until they are pruned out.

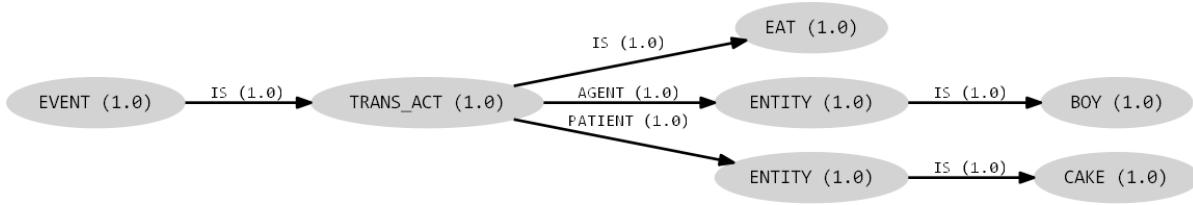


Figure E.51:  $t=649$ . Semantic WM. SemRep updated by the Grammatical WM. The CAKE\_485 construction instance specifies the FOOD patient that had been predicted by the World Knowledge route. Both grammatical and world knowledge have competed, each proposing an opposite thematic role assignment. The World Knowledge route ended up dominating the cooperative computation process and the final SemRep provides an incorrect semantic representation for the counterfactual input utterance “(the) boy is eat-ed by (the) cake”.

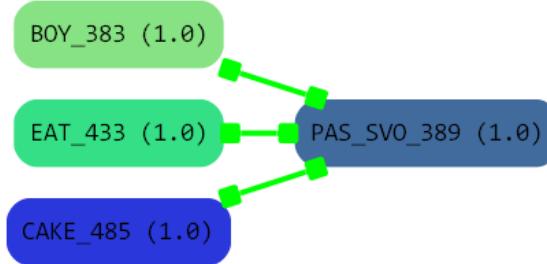


Figure E.52:  $t=912$ . Grammatical WM. Final winning construction instance assemblage.

### E.5.2 Simulation Summaries

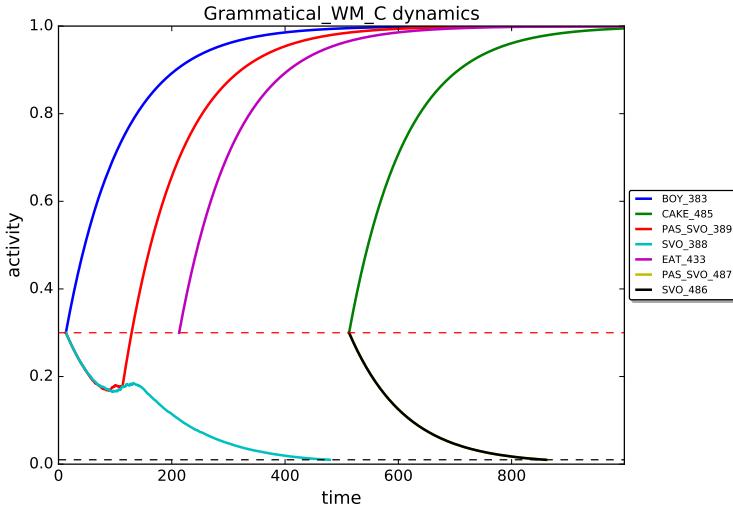


Figure E.53: Activation levels of the construction instances active in Grammatical WM as a function of time. Top dashed line corresponds to the confidence threshold. Bottom dashed line marks to the pruning threshold. See above for representations of the Grammatical WM states over time.

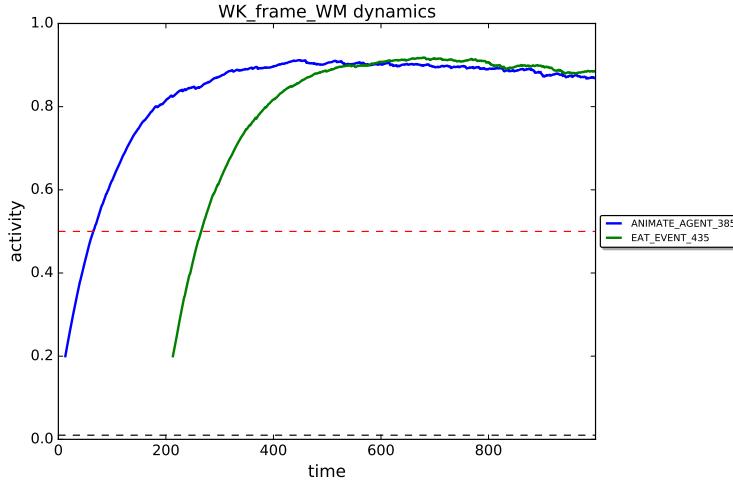


Figure E.54: Activation levels of the world knowledge (event) frame instances active in World Knowledge WM as a function of time. Top dashed line corresponds to the confidence threshold. Bottom dashed line marks to the pruning threshold. See above for a discussion of the World Knowledge WM states over time.

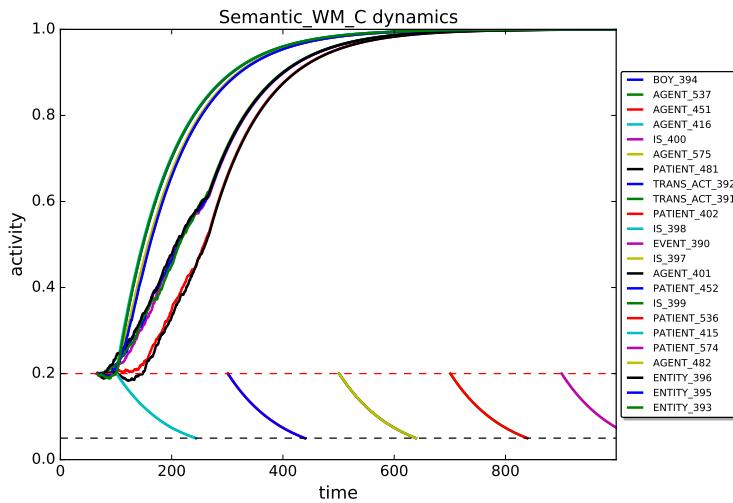


Figure E.55: Activation levels of the concept instances active in Semantic WM as a function of time. Dashed line marks to the pruning threshold. See above for representations of the Semantic WM states over time.

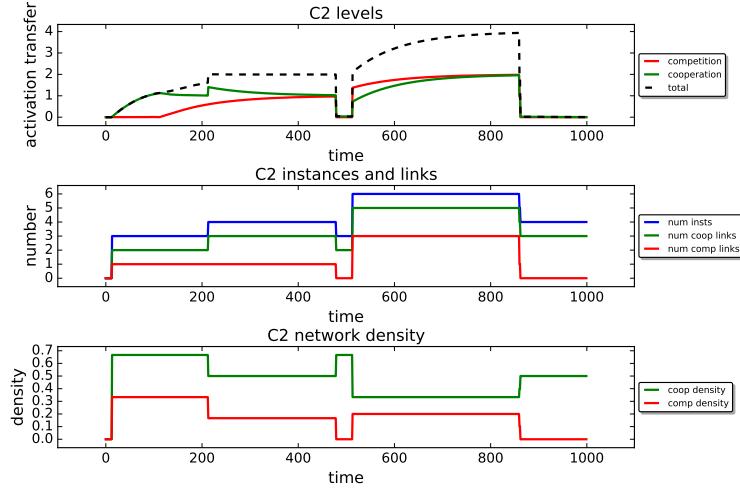


Figure E.56: Higher level analyses of the C2 dynamics taking place in the Grammatical WM through the simulation. Top: C2 levels shows the sum of the activation signals that passes through the C2 network as a function of time (total, dashed line), and also broken down between activation passed through cooperation and competition links (green and red respectively). Middle: Number of instances and links active at each time. Bottom: C2 network density as a function of time (separated into density of cooperation network and competition network).

## Appendix F

# Synthetic ERP: Mathematical, Physical and Computational Foundations

### F.1 General physical Formulation of the Forward Problem

Solving the forward model for EEG recording consists in determining the electrical field generated by neural current sources and measured by electrodes on the scalp. The 4 Maxwell equations give the most general expression of the problem. Given the relatively slow changes of the electric and magnetic fields generated by neural sources (below 100Hz), we can use the quasi-static approximation of Maxwell's equations:

$$(A.1) \left\{ \begin{array}{l} \nabla \cdot \mathbf{E} = \frac{\rho}{\epsilon \epsilon_0} \\ \nabla \cdot \mathbf{B} = 0 \\ \nabla \times \mathbf{E} = \mathbf{0} \\ \nabla \times \mathbf{B} = \mu_0 \mathbf{J} \end{array} \right.$$

as well as

$$(A.2) \nabla \cdot \mathbf{J} = 0 \text{ (Conservation of charge)}$$

Where  $\mathbf{E}$  is the electric field,  $\mathbf{B}$  the magnetic field,  $\mathbf{J}$  the current density,  $\epsilon$  the permittivity of the milieu,  $\rho$  the charge density, and  $\mu_0, \epsilon_0$  are respectively the permeability and permittivity of free space.

It is convenient to separate the primary currents  $\mathbf{J}^P$ , which correspond to the currents generated directly by the neural sources and creating the electric field, from the secondary or return currents  $\mathbf{J}^S$  which are caused by the existence of the field. Following Ohms law we write  $\mathbf{J}^S = \sigma \mathbf{E}$ , where  $\sigma$  is the conductivity (which need not be isotropic or homogeneous in the general case). It follows that:

$$(A.3) \mathbf{J} = \mathbf{J}^P + \sigma \mathbf{E}$$

Finally, since  $\mathbf{E}$  is irrotational, we can define a scalar potential  $V$  such that:

$$(A.4) \mathbf{E} = -\nabla V$$

From there the *Poisson equation for the electric potential*:

$$(A.5) \nabla \cdot (\sigma \nabla V) = \nabla \cdot \mathbf{J}^P$$

We are interested in the potential only at the positions on the scalp where electrodes are located. Solving the forward problem for EEG consists in solving the Poisson equation for primary currents generated by the brain's activity. To do so, one needs to stipulate: (a) what the primary current sources are: this is the role of the *current dipole model* (see Dipole Modeling of Current Sources); (b) what the conductivities are: this is the role of the *head model* (see Conductor Modeling: the Head Model); a method (analytic or numerical) to solve the Poisson equation: we review here the *Boundary Element Method* (BEM) (see Numerical Method: Boundary Element Method (BEM)).

## F.2 Dipole Modeling of Current Sources

Although the activity of every neuron generates currents, both in the dendrites and in the axons, the activity of individual neurons is not sufficient to yield a detectable electric field at the surface of the scalp. Only the local summation of the currents generated by many neurons can result in measurable changes in the electroencephalogram. For currents to be additive, neurons need to be synchronously active and have a spatial arrangement that enables the amplification of the field. Such a configuration is found in the **pyramidal neurons** of the cortex whose dendritic currents are thought to be the main sources of EEG signals. Neighboring pyramidal neurons have apical dendrites arranged parallel to one another and perpendicular to the cortical surface in a palisade-like configuration (cf. Figure 14, left). Dendrites and not axons can display significant synchronous activity since the duration of post synaptic potentials is much longer than that of action potentials, favoring overlapping activation between cells. Finally the excitatory synapses tend to be localized at the apex of the dendritic tree of pyramidal neurons, while the inhibitory synapses tend to cluster around the soma.

Focusing on excitatory synapses, excitatory post-synaptic potentials (EPSPs) result in the pyramidal neurons in currents flowing both in the dendrites and in the extra-cellular environment that can be modeled as in Figure 13 (a). In turn, such currents can be modeled as a current sink at the apex and a current source at the soma, equal in strength, separated by a distance  $L$ . Figure 13 (b) represents such a model with  $\mathbf{r}_a$  the position of the apical sink,  $\mathbf{r}_s$  the position of the source at the soma,  $\mathbf{J}$  the current density, and  $\mathbf{e}_d$  a unit vector pointing from the sink to the source. This configuration can further be modeled as current dipole  $\{\mathbf{d}, \mathbf{r}_d\}$  as presented in Figure 13 (c). The dipole is assigned to the position  $\mathbf{r}_d = \frac{\mathbf{r}_s + \mathbf{r}_a}{2}$ . Its moment is defined as  $\|\mathbf{d}\| = d = L \cdot I$ . The dipole is oriented along  $\mathbf{e}_d$  from the sink to the source.

Modeling of a current sink and source equal in strength by a dipole can be derived from the multiple expansion of the formulation of the field. In this formulation, the first order dipole approximate of the source gives a good account of the field when the distance at which the field is measured is large compared to  $L$ . A current dipole can therefore be used at a microscopic level to model the electric activity of a single pyramidal cell.

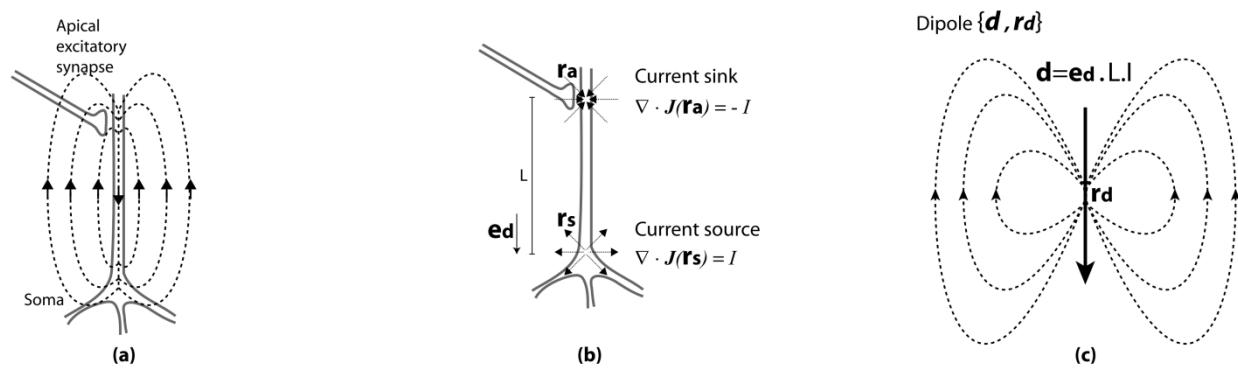


Figure 1. Modeling of a pyramidal neuron as a current dipole. (a) Currents generated by an apical excitatory synapse in the pyramidal neuron. Excitatory synapses tend to be localized in the apical region of the dendritic tree of pyramidal neurons, while inhibitory synapses tend to be localized near the soma. Currents are generated both in the dendrite and in the extracellular environment. (b) Model of the same neuron as a current sink and a current source located at the apex and at the soma of the neuron

respectively. The source and sink are modeled as punctual, separated by a distance  $L$ , and “pumping” an equivalent amount of current  $I$ . (c) Such a configuration can be modeled as a current dipole oriented from the sink to the source, located between the two sources, and whose moment is equal to the product of the current “pumped” by the distance between sources. Such a dipole approximation is valid when the distance at which the field is measured is large compared to  $L$ .

On the left of Figure 14, is schematized a layer of pyramidal neurons, each one associated with a current dipole. The specific spatial configuration result in all the neuron-level dipoles to be oriented perpendicular to the cortical surface. To the extent that the curvature of the cortex can be neglected, the dipoles can be approximated as locally collinear.

As shown in Figure 14 (right), a patch of cortex can therefore be represented by a mesoscopic-level dipole summarizing the activity of the pyramidal neurons. However, such an extra step in modeling the current sources is valid only inasmuch as: (1) the deviation from collinearity in the dipoles due to cortical curvature is negligible, (2) the dipoles are neuron-level dipoles are close to each other relative to the distance of measurement of the field, and (3) the activity of the neurons is hypothesized to be synchronous. Given these hypotheses, a current dipole can be used to model the dendritic currents generated by a population of pyramidal cells in a small patch of cortex. The 6 parameters that define the dipole can in theory be specified: it is located at the center of the patch, oriented perpendicular to the cortical surface pointing inward, its amplitude can be derived from the synaptic activity of the given neuronal population.

Primary current sources can from here be modeled either as *focal single dipoles* (in the case of a few very local activations, a common hypothesis for modeling epileptic seizures), or as *dipoles distributions* (in which case large patches of cortex are modeled as a distribution of dipoles anatomically constrained by the cortical topology).

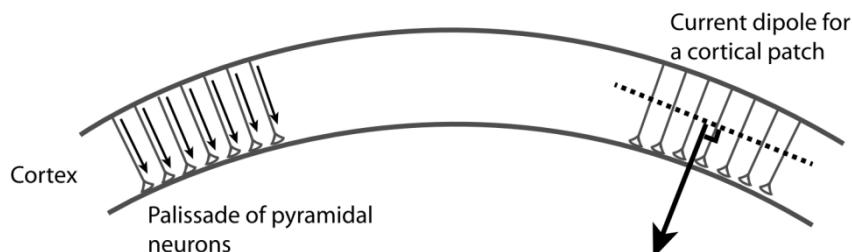


Figure 2. Mesoscopic current dipole model of a patch of cortex. On the right side is depicted the unique organization of pyramidal cells in a palisade-like layer in which the apical dendrites are all oriented perpendicular to the cortical surface. Each neuron is here modeled as a current dipole (cf. Figure 13). On the left side, the whole patch of cortex is modeled as a unique dipole representing the summed contributions of each individual neuron. This summation is made possible by the configuration of apical dendrites. The resulting dipole is perpendicular to the cortical surface and located at the center of the cortical patch. Such a mesoscopic dipole model rests on the hypothesis that any deviation from collinearity of neuron-level dipoles is small (ie curvature of cortex is negligible), the distance between neuron-level dipoles is small compared to the distance between the cortical patch and the locus of measurement of the field, and finally that the neurons’ activities coincide at least partially in time. This possibility to

sum up individual neuron current contributions is a necessary condition for a current source to be strong enough to yield an electric field that can be measured using EEG.

### F.3 Conductor Modeling: the Head Model

Solving Eq. (A.5) requires not only the definition of the primary sources (see Dipole Modeling of Current Sources) but also the definition of the conductivity tensor  $\sigma(\vec{r})$  of the media in which the electric field propagates. To do so we make the following hypotheses: (1) the head as a conductor can be modeled as a series of embedded conduction volumes, typically brain, skull, and scalp (but more complex models can be considered); (2) the boundary between these volumes can be modeled by realistic meshes extracted from anatomical MRI (spherical head models can also be used that allow analytical solutions); (3) the conductivity of each conduction volume can be considered homogeneous and isotropic.

The last point allows the use of the Boundary Element Method to find numerical solutions to Eq. (A.5) (see Numerical Method: Boundary Element Method (BEM)). However, the isotropy hypothesis has been challenged in the case of the skull and white matter. In addition, the values of the conductivities are still debated and tend to depend on the method used to measure them (for a review see (Hallez et al 2007)).

Figure 15 describes the head model we used in our work. The absence of a model of subcortical elements is a limitation of most head models. In addition, the use of a head model in the computation of EEG signals begs the question of “what head” should be used. We here use a realistic head model derived from high-resolution anatomical MRI data acquired for a single subject. But no standard has so far been developed for the head model.

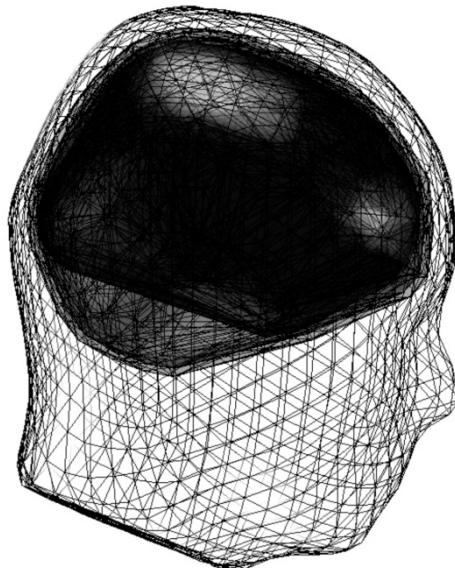


Figure 1. The head model. The head model defines the media in which the electric field propagates. We use a realistic head model composed of 4 conduction volumes. From the innermost volume to the outermost: brain, cerebrospinal fluid, skull, and scalp. These volumes are separated by boundaries defined as triangular meshes: grey matter surface, inner skull, outer skull, and scalp. The meshes are computed from 27 high-resolution anatomical MRI images of a single individual (MNI meshes provided by BrainStorm and extracted using BrainVisa software). Except for the scalp mesh, we used the convex hull associated with the MNI 27 meshes to alleviate computation. In our work, we make the hypothesis that each volume as a homogenous and isotropic conductivity.

The implementation of Phase 2 of Synthetic ERP reported here uses a 4-compartment head model based on the MNI Collins MRI scans (Evans et al 1993, Mazziotta et al 1995) which provides meshes respectively representing the surfaces defined by the grey matter, the inner skull, the outer skull, and the outer surface of the scalp. The ventricles are ignored in the present model. The meshes are triangular meshes provided by the Brainstorm default anatomy (Tadel et al 2011b) and extracted from the MNI Collins MRI scans using BrainVisa software. If the exact geometry of the brain surface is required to compute the orientation of the dipoles, the computation of the electromagnetic field can be greatly alleviated with a minimal impact on the solution by simplifying the geometry of the conduction volumes. For this reason, the head meshes which define the volume boundaries of the head model for the computation of the forward solution are the convex hulls associated with the MNI Collins brain, inner skull, and outer skull but the scalp mesh was kept in its Brainstorm version for display purposes. We used the conductivity measurements provided by (Oostendorp et al 2000) (Brain: 0.22, CSF: 1.79, Skull: 0.015, Scalp: 0.22 S/m).

Few ERP experiments report the stereotactic positions of the electrodes used during the EEG recording. Synthetic ERP therefore uses default 65 electrodes 10/10 standard electrode systems and electrodes positions provided by Brainstorm (for a review of the existing electrode systems see (Jurcak et al 2007)).

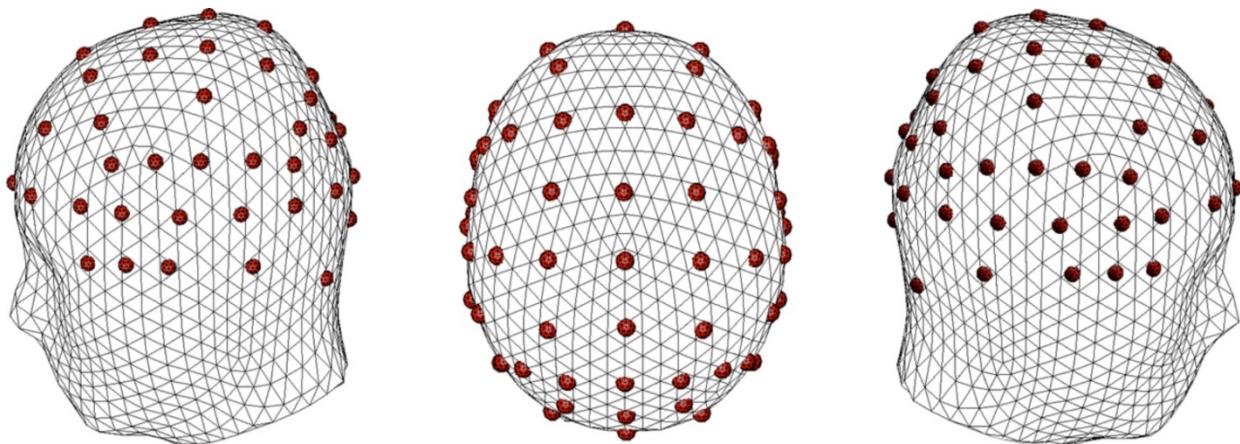


Figure 2. Sensors positions. Sensor positions are defined as the default MNI coordinates of the 10/10 65 channels EEG electrode cap as defined by Brainstorm and associated with the MNI Collins head meshes. The electric field, output of the forward model, is computed at the sensors' positions.

## F.4 Algebraic Formulation

For multiple points of measurements of the electric potential (electrodes) and multiple sources (dipoles), it is useful to give the forward problem a compact algebraic formulation. Given a dipole with position  $\mathbf{r}_{dip}$  and moment  $\mathbf{d}$ , the electric potential measured at an electrode with position  $\mathbf{r}$  solution of the Poisson equation (6) can be written:

$$(D.1) V(\mathbf{r}) = g(\mathbf{r}, \mathbf{r}_{dip}, \mathbf{d})$$

The Poisson equation is linear. If we note  $\mathbf{d} = d \cdot \mathbf{e}_{dip}$ , where  $\|\mathbf{e}_{dip}\| = 1$ , we can write:

$$(D.2) V(\mathbf{r}) = g(\mathbf{r}, \mathbf{r}_{dip}, \mathbf{e}_{dip}) \cdot d$$

The electric potential generated by a dipole at the location of an electrode depends linearly on the amplitude of the dipole, i.e. on the neural activity. However, it does not depend linearly on the position and orientation of the dipole (defined by the geometry of the cortex).

For **N dipoles**  $\{\mathbf{r}_{dip}^j, \mathbf{e}_{dip}^j, d^j\}, j \in \llbracket 1, N \rrbracket$ , and **K electrodes** with positions  $\mathbf{r}^k, k \in \llbracket 1, K \rrbracket$ , given the linearity of the Poisson equation, we can write:

$$(D.3) \forall k \in \llbracket 1, K \rrbracket, V(\mathbf{r}^k) = \sum_{j=1}^N g(\mathbf{r}^k, \mathbf{r}_{dip}^j, \mathbf{e}_{dip}^j) \cdot d^j$$

Therefore:

$$(D.4) \mathbf{V} = \mathbf{G}\mathbf{D}, \text{ with:}$$

$$\mathbf{V} = \begin{bmatrix} V(\mathbf{r}^1) \\ \dots \\ V(\mathbf{r}^K) \end{bmatrix} \in \mathbb{R}^K, \text{ electrode measurements vector, function of the electrodes positions.}$$

$$\mathbf{G} = \begin{bmatrix} g(\mathbf{r}^1, \mathbf{r}_{dip}^1, \mathbf{e}_{dip}^1) & \dots & g(\mathbf{r}^1, \mathbf{r}_{dip}^N, \mathbf{e}_{dip}^N) \\ \dots & \ddots & \dots \\ g(\mathbf{r}^K, \mathbf{r}_{dip}^1, \mathbf{e}_{dip}^1) & \dots & g(\mathbf{r}^K, \mathbf{r}_{dip}^N, \mathbf{e}_{dip}^N) \end{bmatrix} \in \mathbb{R}^{K \times N}, \text{ gain matrix function of the sources positions,}$$

orientations, and of the electrodes positions.

$$\mathbf{D} = \begin{bmatrix} d^1 \\ \dots \\ d^N \end{bmatrix} \in \mathbb{R}^N, \text{ dipole amplitudes vector.}$$

If the positions of electrodes are fixed as well as the position and orientation of the sources with only the amplitudes of the sources varying in time, equation (D.4) can be rewritten

$$(D.4') \mathbf{V}(t) = \mathbf{G}\mathbf{D}(t)$$

Since the gain matrix  $\mathbf{G}$  does not depend on time in these conditions, it can be computed only once for a set of sources. The EEG recordings  $\mathbf{V}(t)$  is simply a linear combination of the **source waveform functions**  $d^j(t), j \in \llbracket 1, N \rrbracket$ .

## F.5 Numerical Method: Boundary Element Method (BEM)

Although analytical solutions can be found for the Poisson equation (see General Physical Formulation of the Forward Problem, Eq. A.5) in the case of spherical head models, the use of realistic head models requires the use of numerical methods. In the case of homogeneous and isotropic conductivities, which remains a common assumption for the forward problem, Eq. (A.5) can be numerically solved using the Boundary Element Method (BEM). In more complex cases (conductivities non-isotropic and/or non-homogeneous), other methods are available such as Finite Element Method (FEM). The strength of the BEM rests in the fact that the potential at any point on a boundary can be determined by its values on the other boundaries. Hence only these boundary values need to be numerically computed which reduces the size of the problem. Since for EEG we are measuring the electric potential on the scalp, we are only interested in the values on this boundary. FEM approaches on the other hand require computing numerical values for the electric potential at every point in space. In a nutshell, the BEM approach consists in computing the values of the electric potential on the conductors' volumes boundaries by approximating the continuous boundaries as a tessellation of small boundary elements.

The BEM rests on the integral formulation of the Poisson equation in the case of  $N_V$  conductor volumes of conductivities  $\sigma_k$  separated by  $N_S$  surface boundaries  $S_i$ . Due to the linearity of the problem (see preceding section), we will here, without loss of generality, consider the case of a single source dipole  $\{\mathbf{r}_{dip}, \mathbf{d}\}$ . It has been shown elsewhere (Sarvas 1987) that the integral formulation of Eq. (A.5) can be written:

$$(E.1) V(\mathbf{r}) = \frac{2\sigma_0}{\sigma_k^- + \sigma_k^+} V_0(\mathbf{r}) + \frac{1}{2\pi} \sum_{j=1}^{N_S} \frac{\sigma_j^- - \sigma_j^+}{\sigma_k^- + \sigma_k^+} \left( \oint_{\mathbf{r}' \in S_j} \frac{V(\mathbf{r}')}{\|\mathbf{r} - \mathbf{r}'\|^2} \frac{\mathbf{r} - \mathbf{r}'}{\|\mathbf{r} - \mathbf{r}'\|} dS_j \right), \text{ for } \mathbf{r} \in S_k.$$

Where  $V_0$  is the potential that would be generated by the dipole source in an infinite medium whose conductivity is that of the medium in which the source is located (the brain in our case).  $\sigma_0$  is the conductivity of this medium (here the conductivity of the brain). For a point  $\mathbf{r} \in S_n$ ,  $\sigma_n^+$  and  $\sigma_n^-$  refer respectively to the conductivity outside and inside the boundary defined by  $S_n$ .

In the case of a current density distribution modeled as a single dipole  $\{\mathbf{r}_{dip}, \mathbf{d}\}$ ,  $V_0(\mathbf{r})$  has a straightforward analytical value:

$$(E.2) V_0(\mathbf{r}) = \frac{1}{4\pi\sigma_0} \frac{\mathbf{d} \cdot (\mathbf{r} - \mathbf{r}_{dip})}{\|\mathbf{r} - \mathbf{r}_{dip}\|^3}$$

The boundary element method consists in first approximating the surface integrals in Eq. (E.1) by **tessellating each surface boundary  $S_k$**  into  $n_{S_k}$  triangles surface elements  $\Delta_i^{S_k}$ ,  $i \in [\![1, n_{S_k}]\!]$ . Eq. (E.1) can then be approximated by:

$$(E.3) V(\mathbf{r}) = \frac{2\sigma_0}{\sigma_k^- + \sigma_k^+} V_0(\mathbf{r}) + \frac{1}{2\pi} \sum_{j=1}^{N_S} \frac{\sigma_j^- - \sigma_j^+}{\sigma_k^- + \sigma_k^+} \sum_{i=1}^{n_{S_j}} \left( \int_{\mathbf{r}' \in \Delta_i^{S_j}} \frac{V(\mathbf{r}')}{\|\mathbf{r} - \mathbf{r}'\|^2} \frac{\mathbf{r} - \mathbf{r}'}{\|\mathbf{r} - \mathbf{r}'\|} dS_j \right)$$

The second step consists in approximating  $V^j(\mathbf{r})$  on each boundary  $S_j$  by  $\tilde{V}^j(\mathbf{r})$  defined as the linear combination of simple basis functions  $h_p(\mathbf{r})$ ,  $p \in [\![1, n_{S_j}]\!]$ .

$$(E.4) \tilde{V}^j(\mathbf{r}) = \sum_{p=1}^{n_{S_j}} V_{j,p} h_p(\mathbf{r})$$

We can then rewrite equation (E.3):

$$(E.5) \quad V(\mathbf{r}) = \frac{2\sigma_0}{\sigma_k^- + \sigma_k^+} V_0(\mathbf{r}) + \frac{1}{2\pi} \sum_{j=1}^{N_S} \frac{\sigma_j^- - \sigma_j^+}{\sigma_k^- + \sigma_k^+} \sum_{i=1}^{n_{S_j}} \sum_{p=1}^{n_{S_j}} V_{j,p} \left( \int_{\mathbf{r}' \in \Delta_i^{S_j}} \frac{h_p(\mathbf{r}')}{\|\mathbf{r} - \mathbf{r}'\|^2} \frac{\mathbf{r} - \mathbf{r}'}{\|\mathbf{r} - \mathbf{r}'\|} dS_j \right)$$

A common choice for the basis function is to consider the set of basis functions such that on surface  $S_j$

$$h_p(\mathbf{r}) = \begin{cases} 1 & \text{if } \mathbf{r} \in \Delta_p^{S_j} \\ 0 & \text{otherwise} \end{cases}$$

which corresponds to an approximation that the electric potential is constant on each triangular boundary element.

The final approximation consists in defining so called *collocation points* at which the basis functions will be numerically estimated. The collocation points are typically the centroids of the triangular boundary elements.

Writing equation (E.5) for each one of these collocation points result in a set of equations that can be solved to determine the coefficient  $V_{j,p}$  for each triangular element  $p$  on each surface  $S_j$ .

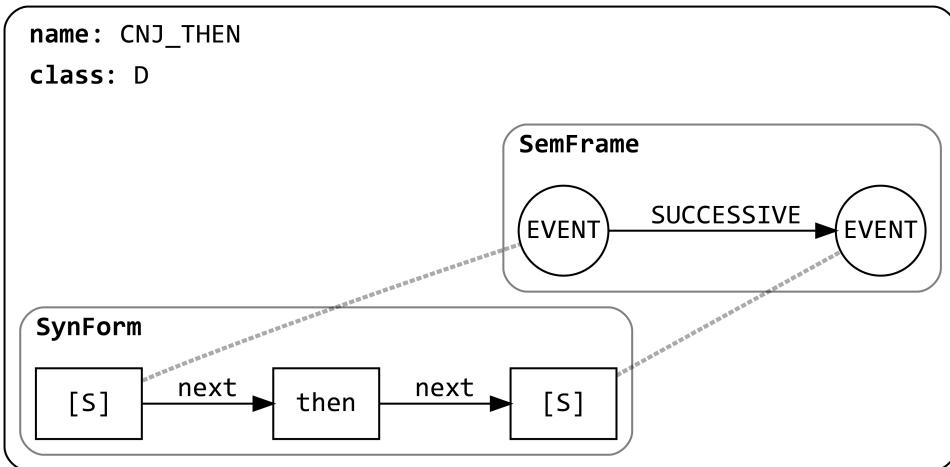
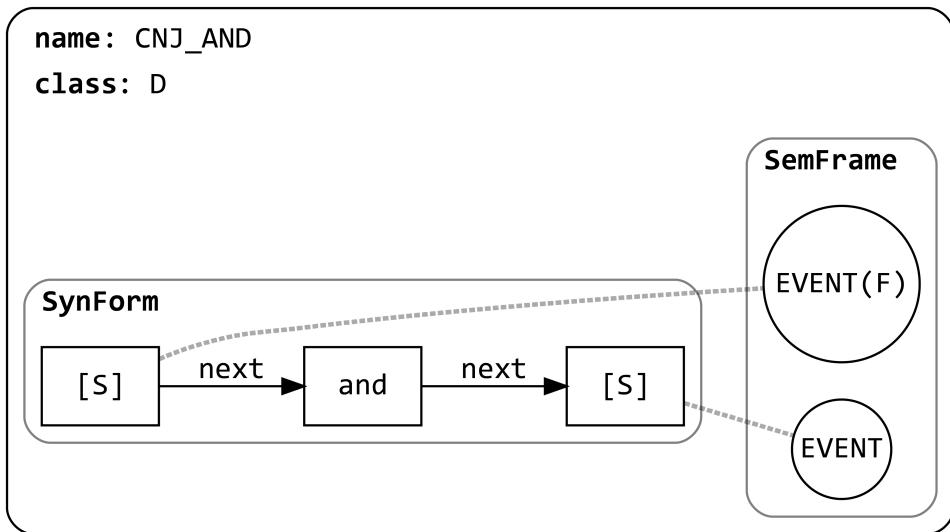
The forward problem corresponds in this case in solving a system of linear equations, which requires the inversion of a dense matrix with a size of the order of number of boundary elements defined. The size of the matrix is relatively small since we do not need to compute parameters for each point in the volume and depends on the coarseness of the tessellations used.

Going back to Eq. (D.4') (see Algebraic Formulation), the matrix inversion need only to be done once for a given location and orientation of the dipole sources (and location of the sensors) to define the gain matrix.

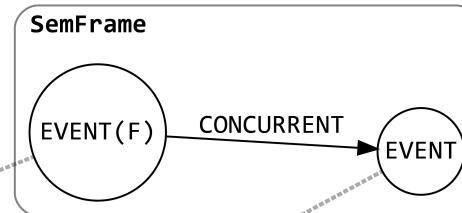
In this paper, the forward model is computed using the FieldTrip (Oostenveld et al 2011) MatLab implementation of OpenMEG version of BEM (Gramfort et al 2011).

## Appendix G

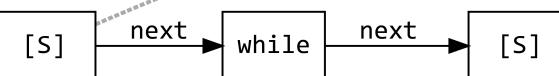
# Abstract Constructions



**name:** CNJ\_WHILE  
**class:** D

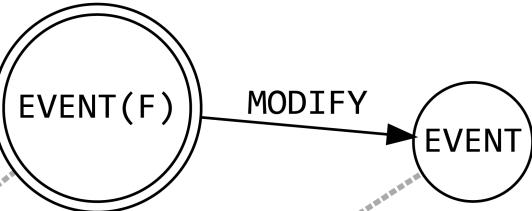


**SynForm**



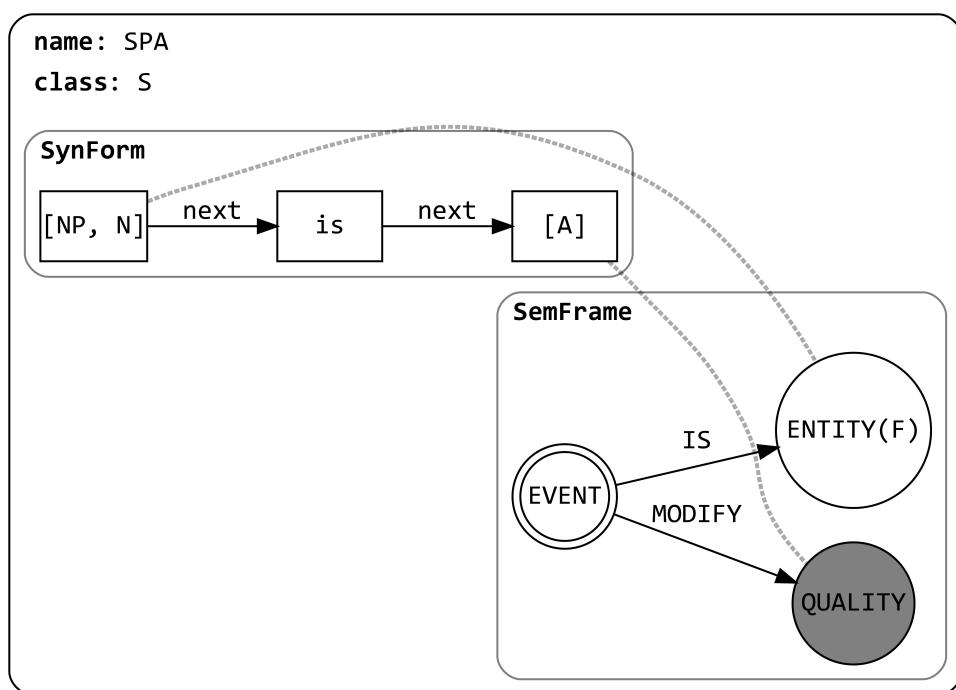
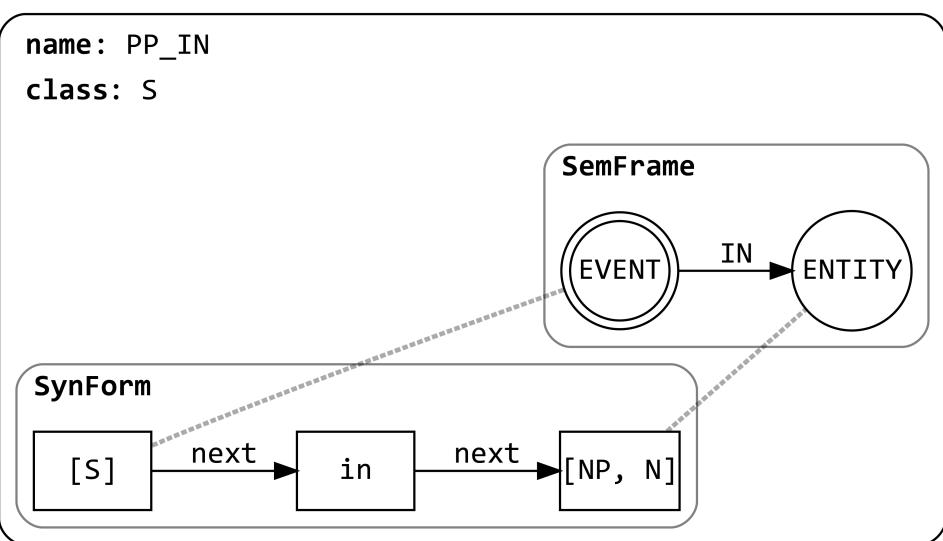
**name:** S\_PP  
**class:** S

**SemFrame**

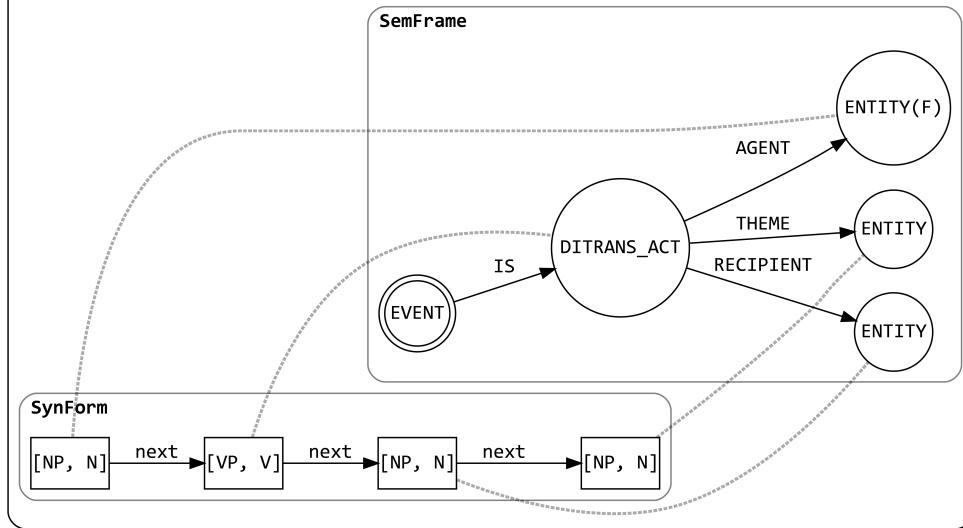


**SynForm**

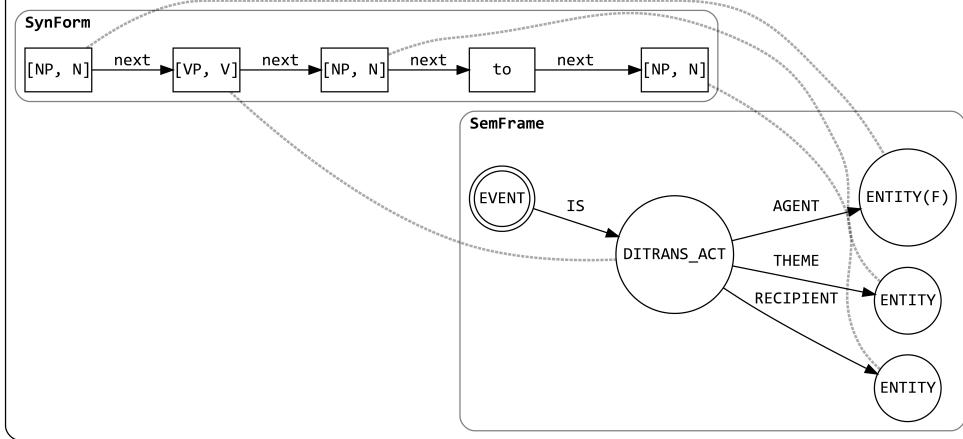




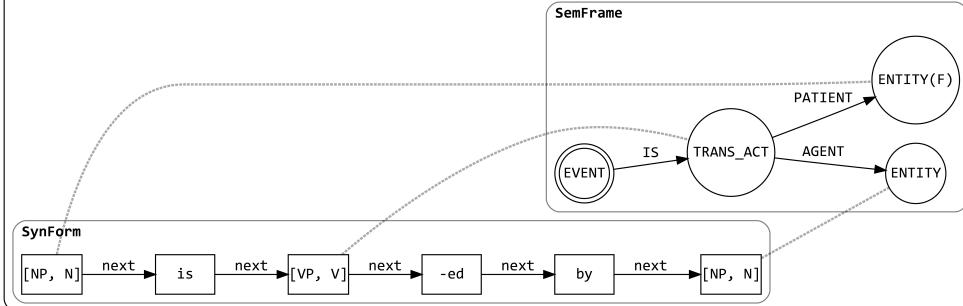
**name:** DOUBLE\_OBJ  
**class:** S



**name:** OBLIQUE\_DATIVE  
**class:** S



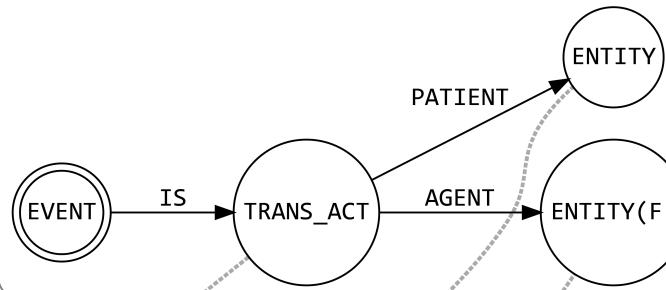
**name:** PAS\_SVO  
**class:** S



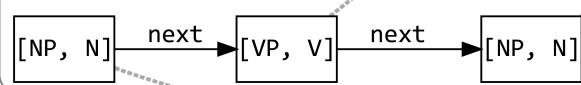
**name:** SVO

**class:** S

**SemFrame**



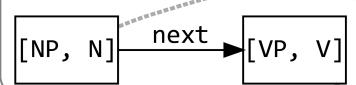
**SynForm**



**name:** SV

**class:** S

**SynForm**



**SemFrame**

