

MACQUARIE UNIVERSITY, SYDNEY, AUSTRALIA

Centre for Language Technology

Department of Computing

Division of Information and Communication Sciences

and

UNIVERSITÉ DE PROVENCE, FRANCE

U.F.R. M.I.M.

ÉCOLE DOCTORALE DE MATHÉMATIQUES ET INFORMATIQUE E.D. 184

Laboratoire Parole et Langage

# Modelling Syntactic Gradience with Loose Constraint-based Parsing



## Modélisation de la gradience syntaxique par analyse relâchée à base de contraintes

Jean-Philippe PROST

Submitted in Partial Fulfilment of Joint Institutional Requirements for the  
Double-badged Degree of

DOCTOR OF PHILOSOPHY

and

DOCTEUR DE L'UNIVERSITÉ DE PROVENCE

*Spécialité : Informatique*

October 2008



UNIVERSITÉ DE PROVENCE, FRANCE  
U.F.R. M.I.M.  
ÉCOLE DOCTORALE DE MATHÉMATIQUES ET INFORMATIQUE E.D. 184  
Laboratoire Parole et Langage

et

MACQUARIE UNIVERSITY, SYDNEY, AUSTRALIA  
Centre for Language Technology  
Department of Computing  
Division of Information and Communication Sciences

## THÈSE en COTUTELLE

présentée pour obtenir le grade de

DOCTEUR DE L'UNIVERSITÉ DE PROVENCE

*Spécialité : Informatique*

et

DOCTOR OF PHILOSOPHY

# Modélisation de la gradience syntaxique par analyse relâchée à base de contraintes



## Modelling Syntactic Gradience with Loose Constraint-based Parsing

Jean-Philippe PROST

soutenue publiquement le 10 décembre 2008

### JURY

|                          |                               |                          |
|--------------------------|-------------------------------|--------------------------|
| Pr. Alexis NASR          | Université de la Méditerranée | <i>Président du Jury</i> |
| Pr. Denys DUCHIER        | Université d'Orléans          | <i>Rapporteur</i>        |
| Dr. Gerald PENN          | University of Toronto         | <i>Rapporteur</i>        |
| Dr. Eric de la CLERGERIE | INRIA                         | <i>Examineur</i>         |
| Dr. Philippe BLACHE      | Université de Provence        | <i>Co-directeur</i>      |
| Dr. Diego MOLLÁ ALIOD    | Macquarie University          | <i>Co-directeur</i>      |
| Dr. Mark DRAS            | Macquarie University          | <i>Directeur adjoint</i> |



# Abstract

The grammaticality of a sentence has conventionally been treated in a binary way: either a sentence is grammatical or not. A growing body of work, however, focuses on studying intermediate levels of acceptability, sometimes referred to as *gradience*. To date, the bulk of this work has concerned itself with the exploration of human assessments of syntactic gradience. This dissertation explores the possibility to build a robust computational model that accords with these human judgements.

We suggest that the concepts of *Intersective Gradience* and *Subsective Gradience* introduced by Aarts for modelling graded judgements be extended to cover deviant language. Under such a new model, the problem then raised by gradience is to classify an utterance as a member of a specific category according to its syntactic characteristics. More specifically, we extend Intersective Gradience (IG) so that it is concerned with choosing the most suitable syntactic structure for an utterance among a set of candidates, while Subsective Gradience (SG) is extended to be concerned with calculating to what extent the chosen syntactic structure is typical from the category at stake. IG is addressed in relying on a criterion of optimality, while SG is addressed in rating an utterance according to its grammatical acceptability. As for the required syntactic characteristics, which serve as features for classifying an utterance, our investigation of different frameworks for representing the syntax of natural language shows that they can easily be represented in Model-Theoretic Syntax; we choose to use Property Grammars (PG), which offers to model the *characterisation* of an utterance. We present here a fully automated solution for modelling syntactic gradience, which characterises any well formed or ill formed input sentence, generates an optimal parse for it, then rates the utterance according to its grammatical acceptability.

Through the development of such a new model of gradience, the main contribution of this work is three-fold.

First, we specify a model-theoretic logical framework for PG, which bridges the gap observed in the existing formalisation regarding the constraint satisfaction and constraint relaxation mechanisms, and how they relate to the projection of a category

during the parsing process. This new framework introduces the notion of *loose satisfaction*, along with a formulation in first-order logic, which enables reasoning about the characterisation of an utterance.

Second, we present our implementation of Loose Satisfaction Chart Parsing (LSCP), a dynamic programming approach based on the above mechanisms, which is proven to always find the full parse of optimal merit. Although it shows a high theoretical worst time complexity, it performs sufficiently well with the help of heuristics to let us experiment with our model of gradience.

And third, after postulating that human acceptability judgements can be predicted by factors derivable from LSCP, we present a numeric model for rating an utterance according to its syntactic gradience. We measure a good correlation with grammatical acceptability by human judgements. Moreover, the model turns out to outperform an existing one discussed in the literature, which was experimented with parses generated manually.

**Keywords** Gradience, acceptability, grammaticality, optimality, configuration, Model-Theoretic Syntax, Property Grammars, characterisation, constraint-based chart parsing, robustness, loose constraint satisfaction.

## Résumé

La grammaticalité d’une phrase est habituellement conçue comme une notion binaire : une phrase est soit grammaticale, soit agrammaticale. Cependant, bon nombre de travaux se penchent de plus en plus sur l’étude de degrés d’acceptabilité intermédiaires, auxquels le terme de *gradience* fait parfois référence. À ce jour, la majorité de ces travaux s’est concentrée sur l’étude de l’évaluation humaine de la gradience syntaxique. Cette étude explore la possibilité de construire un modèle robuste qui s’accorde avec ces jugements humains.

Nous suggérons d’élargir au langage mal formé les concepts de *Gradience Intersective* et de *Gradience Subsective*, proposés par Aarts pour la modélisation de jugements graduels. Selon ce nouveau modèle, le problème que soulève la gradience concerne la classification d’un énoncé dans une catégorie particulière, selon des critères basés sur les caractéristiques syntaxiques de l’énoncé. Nous nous attachons à étendre la notion de Gradience Intersective (GI) afin qu’elle concerne le choix de la meilleure solution parmi un ensemble de candidats, et celle de Gradience Subsective (GS) pour qu’elle concerne le calcul du degré de typicité de cette structure au sein de sa catégorie. La GI est alors modélisée à l’aide d’un critère d’optimalité, tandis que la GS est modélisée par le calcul d’un degré d’acceptabilité grammaticale. Quant aux caractéristiques syntaxiques requises pour permettre de classer un énoncé, notre étude de différents cadres de représentation pour la syntaxe du langage naturel montre qu’elles peuvent aisément être représentées dans un cadre de syntaxe modèle-théorique (*Model-Theoretic Syntax*). Nous optons pour l’utilisation des Grammaires de Propriétés (GP), qui offrent, précisément, la possibilité de modéliser la *caractérisation* d’un énoncé. Nous présentons ici une solution entièrement automatisée pour la modélisation de la gradience syntaxique, qui procède de la caractérisation d’une phrase bien ou mal formée, de la génération d’un arbre syntaxique optimal, et du calcul d’un degré d’acceptabilité grammaticale pour l’énoncé.

À travers le développement de ce nouveau modèle, la contribution de ce travail comporte trois volets.

Premièrement, nous spécifions un système logique pour les GP qui permet la révision de sa formalisation sous l’angle de la théorie des modèles. Il s’attache notamment à formaliser les mécanismes de satisfaction et de relâche de contraintes mis en œuvre dans les GP, ainsi que la façon dont ils permettent la projection d’une catégorie lors du processus d’analyse. Ce nouveau système introduit la notion de *satisfaction relâchée*, et une formulation en logique du premier ordre permettant de raisonner au sujet d’un énoncé.

Deuxièmement, nous présentons notre implantation du processus d’*analyse syntaxique relâchée à base de contraintes* (*Loose Satisfaction Chart Parsing*, ou LSCP), dont nous prouvons qu’elle génère toujours une analyse syntaxique complète et optimale. Cette approche est basée sur une technique de programmation dynamique (*dynamic programming*), ainsi que sur les mécanismes décrits ci-dessus. Bien que d’une complexité élevée, cette solution algorithmique présente des performances suffisantes pour nous permettre d’expérimenter notre modèle de gradience.

Et troisièmement, après avoir postulé que la prédiction de jugements humains d’acceptabilité peut se baser sur des facteurs dérivés de la LSCP, nous présentons un modèle numérique pour l’estimation du degré d’acceptabilité grammaticale d’un énoncé. Nous mesurons une bonne corrélation de ces scores avec des jugements humains d’acceptabilité grammaticale. Qui plus est, notre modèle s’avère obtenir de meilleures performances que celles obtenues par un modèle préexistant que nous utilisons comme référence, et qui, quant à lui, a été expérimenté à l’aide d’analyses syntaxiques générées manuellement.

**Mots-clés** Gradience, acceptabilité, grammaticalité, optimalité, configuration, syntaxe modèle-théorique (*Model-Theoretic Syntax*), Grammaires de Propriétés, analyse syntaxique tabulaire par contraintes, robustesse, satisfaction relâchée de contraintes.



# Contents

|  |          |
|--|----------|
| Abstract                                 | v        |
| Résumé                                   | vii      |
| Table of Contents                        | ix       |
| List of Tables                           | xv       |
| List of Figures                          | xvii     |
| List of Algorithms                       | xix      |
| Preface                                  | xxi      |
| Acknowledgements                         | xxiii    |
| Remerciements                            | xxv      |
| <b>1 Introduction</b>                    | <b>1</b> |
| <b>2 Background</b>                      | <b>5</b> |
| 2.1 Introduction . . . . .               | 5        |
| 2.2 Epistemology of Gradience . . . . .  | 6        |
| 2.2.1 Subsective Gradience . . . . .     | 8        |
| 2.2.2 Intersective Gradience . . . . .   | 9        |
| 2.2.3 Constructional Gradience . . . . . | 9        |

|          |   |           |
|----------|---|-----------|
| 2.2.4    | Markedness . . . . .  | 12        |
| 2.3      | Gradience and Frameworks for KR . . . . .                           | 13        |
| 2.3.1    | Generative-Enumerative Syntax <i>vs.</i> Model-Theoretic Syntax . . | 13        |
| 2.3.2    | Optimality Theory . . . . .   | 21        |
| 2.3.3    | Construction Grammar . . . . .                                      | 26        |
| 2.3.4    | Preliminary Conclusions on Knowledge Representation . . . . .       | 28        |
| 2.4      | Models of Syntactic Gradience . . . . .                             | 30        |
| 2.4.1    | Aarts' Model . . . . .  | 30        |
| 2.4.2    | Linear Optimality Theory (LOT) (Keller) . . . . .                   | 33        |
| 2.5      | Implementing Syntactic Gradience . . . . .                          | 38        |
| 2.5.1    | Constraint Dependency Grammar (CDG) (Maruyama) . . . . .            | 38        |
| 2.5.2    | Weighted CDG (Schröder et al.) . . . . .                            | 41        |
| 2.5.3    | Property Grammars (Blache) . . . . .                                | 46        |
| 2.5.4    | SeedParser (VanRullen) . . . . .                                    | 48        |
| 2.5.5    | Dahl and Blache . . . . .   | 52        |
| 2.5.6    | Morawietz and Blache . . . . .                                      | 54        |
| 2.5.7    | Configuration Task . . . . .  | 54        |
| 2.5.8    | eXtended Dependency Grammar (XDG) (Duchier) . . . . .               | 57        |
| 2.5.9    | Estratat and Henocque . . . . .                                     | 59        |
| 2.6      | Conclusion . . . . .  | 60        |
| 2.6.1    | Summary . . . . .   | 60        |
| 2.6.2    | Pending Questions . . . . .   | 63        |
| <b>3</b> | <b>A Model-theoretic Framework for PG</b>                           | <b>67</b> |
| 3.1      | Introduction . . . . .  | 67        |
| 3.2      | A Logical System for Property Grammars . . . . .                    | 70        |
| 3.2.1    | Syntax . . . . .  | 70        |
| 3.2.2    | Semantics . . . . .   | 71        |
| 3.2.3    | Formulation of the PG Constraint Types . . . . .                    | 74        |
| 3.2.4    | Grammar . . . . .   | 81        |
| 3.2.5    | Satisfaction . . . . .  | 85        |

|          |   |            |
|----------|---|------------|
| 3.3      | Properties of $\Xi$ and Discussion . . . . .              | 92         |
| 3.3.1    | A Discriminant for Multiple Loose Models . . . . .        | 92         |
| 3.3.2    | Constituent Structure . . . . .                           | 93         |
| 3.3.3    | Grammaticality . . . . .                                  | 94         |
| 3.3.4    | Monotonicity . . . . .                                    | 95         |
| 3.3.5    | Some Related Works . . . . .                              | 98         |
| 3.4      | Conclusion . . . . .                                      | 101        |
| <b>4</b> | <b>Loose Constraint-based Parsing</b>                     | <b>103</b> |
| 4.1      | Introduction . . . . .                                    | 103        |
| 4.2      | Presenting And Representing Syntactic Knowledge . . . . . | 108        |
| 4.3      | Problem Specification . . . . .                           | 109        |
| 4.3.1    | Problem Statement . . . . .                               | 109        |
| 4.3.2    | Outcome . . . . .   | 109        |
| 4.4      | Algorithmic Solution . . . . .                            | 111        |
| 4.4.1    | Correctness . . . . .                                     | 112        |
| 4.4.2    | Sketch of the Process . . . . .                           | 112        |
| 4.4.3    | Algorithm . . . . .                                       | 114        |
| 4.4.4    | Merit Function . . . . .                                  | 126        |
| 4.4.5    | Consulting The Grammar . . . . .                          | 132        |
| 4.4.6    | Algorithm Walkthrough . . . . .                           | 133        |
| 4.5      | Optimality . . . . .                                      | 141        |
| 4.6      | Complexity . . . . .                                      | 142        |
| 4.7      | The Corpus of Acceptability Judgements . . . . .          | 145        |
| 4.8      | Heuristic . . . . .                                       | 147        |
| 4.8.1    | Fixing a Satisfaction Threshold . . . . .                 | 147        |
| 4.9      | Evaluation . . . . .                                      | 151        |
| 4.9.1    | What and How to Evaluate? . . . . .                       | 152        |
| 4.9.2    | Evaluation #1: EASY . . . . .                             | 153        |
| 4.9.3    | Evaluation #2: quasi-expressions . . . . .                | 160        |
| 4.9.4    | Elements of Accuracy . . . . .                            | 163        |

|          |  |            |
|----------|--|------------|
| 4.9.5    | Conclusion . . . . .                                   | 170        |
| 4.10     | Conclusion . . . . .                                   | 171        |
| <b>5</b> | <b>A Computational Model For Gradience</b>             | <b>173</b> |
| 5.1      | Introduction . . . . .                                 | 173        |
| 5.2      | Reference Corpus . . . . .                             | 175        |
| 5.3      | Modelling Syntactic Gradience . . . . .                | 176        |
| 5.3.1    | Merit <i>vs.</i> Rate . . . . .                        | 179        |
| 5.3.2    | Postulates . . . . .                                   | 181        |
| 5.4      | Rating Models . . . . .                                | 188        |
| 5.4.1    | Scoring Terms . . . . .                                | 188        |
| 5.4.2    | Combining Terms into Scoring Functions . . . . .       | 190        |
| 5.4.3    | Rating Functions . . . . .                             | 192        |
| 5.5      | Empirical Investigation . . . . .                      | 194        |
| 5.5.1    | Model Calibration . . . . .                            | 196        |
| 5.5.2    | Data Sample from Blache <i>et al.</i> (2006) . . . . . | 197        |
| 5.5.3    | Top Scores . . . . .                                   | 198        |
| 5.5.4    | Interpretation . . . . .                               | 203        |
| 5.6      | Conclusion . . . . .                                   | 213        |
| 5.6.1    | Further Work . . . . .                                 | 215        |
| <b>6</b> | <b>Conclusion</b>                                      | <b>217</b> |
| 6.1      | Summary . . . . .                                      | 218        |
| 6.2      | Further Work . . . . .                                 | 223        |
| 6.2.1    | On Scaling Up . . . . .                                | 223        |
| 6.2.2    | On Modelling Gradience . . . . .                       | 225        |
| 6.2.3    | Generalisation and Optimisation of LSCP . . . . .      | 226        |
|          | <b>Bibliography</b>                                    | <b>229</b> |
|          | <b>Index</b>   | <b>241</b> |
| <b>A</b> | <b>PG Construction Grammar</b>                         | <b>251</b> |

|                                     |            |
|-------------------------------------|------------|
| <b>B PG EASY Grammar</b>            | <b>263</b> |
| <b>C Instructions to Annotators</b> | <b>281</b> |



# List of Tables

|      |  |     |
|------|--|-----|
| 3.1  | Legend and graphic conventions used in this dissertation . . . . .               | 83  |
| 3.2  | $\Gamma$ : An Example PG Grammar for French (1) . . . . .                        | 84  |
| 3.3  | $\Gamma$ : An Example PG Grammar for French (2) . . . . .                        | 85  |
| 4.1  | Chart of unlabelled sub-structures . . . . .                                     | 118 |
| 4.2  | Parsing <i>Chloé aime le chocolat</i> (Chloe likes [the] chocolate): chart . . . | 135 |
| 4.3  | Parsing <i>*Chloé aime chocolat le</i> (Chloe likes chocolate [the]): chart . .  | 137 |
| 4.4  | Error patterns . . . . .   | 146 |
| 4.5  | Values for the Threshold by Error Pattern . . . . .                              | 151 |
| 4.6  | Values for the Construction-specific Threshold . . . . .                         | 151 |
| 4.7  | Evaluation of <i>Numbat</i> (EASY) . . . . .                                     | 155 |
| 4.8  | Evaluation of the shallow parser <i>ShP</i> (EASY) . . . . .                     | 156 |
| 4.9  | Evaluation of the stochastic parser <i>StP</i> (EASY) . . . . .                  | 157 |
| 4.10 | Cross bracket measures for <i>Numbat</i> by constituent type (EASY) . . . .      | 158 |
| 4.11 | Precision/Recall (Evaluation #2) . . . . .                                       | 162 |
| 5.1  | Error patterns . . . . .   | 177 |
| 5.2  | Human judgements of acceptability and Cohesion, per error pattern . .            | 181 |
| 5.3  | Calibration of adjustments and weights for all three models . . . . .            | 196 |
| 5.4  | Correlations on small data sample . . . . .                                      | 198 |
| 5.5  | Human judgements of acceptability: reference scores . . . . .                    | 201 |
| 5.6  | Gradient of $g$ -scores . . . . .  | 202 |
| 5.7  | Gradient of $\gamma$ -scores . . . . .   | 202 |

|     |   |     |
|-----|---|-----|
| 5.8 | Gradient of $\gamma'$ -scores . . . . . | 203 |
|-----|---|-----|



# List of Figures

|      |  |     |
|------|--|-----|
| 3.1  | Loosely consistent constituent structure for an ill-formed sentence . . .                      | 92  |
| 4.1  | Example of lookup table for the grammar (sample) . . . . .                                     | 133 |
| 4.2  | <i>Numbat</i> 's output for case of missing Adjective . . . . .                                | 164 |
| 4.3  | <i>Numbat</i> 's output for case of false substantive adjective . . . . .                      | 166 |
| 4.4  | <i>Numbat</i> 's output for case of false substantive adjective . . . . .                      | 166 |
| 4.5  | <i>Numbat</i> 's output for case of redundant past participle . . . . .                        | 167 |
| 4.6  | <i>Numbat</i> 's output for case of missing Verb . . . . .                                     | 169 |
| 4.7  | <i>Numbat</i> 's output for case of missing Verb . . . . .                                     | 170 |
| 5.1  | Cohesion <i>vs.</i> Acceptability: model fit over the full corpus . . . . .                    | 182 |
| 5.2  | Propagation Postulate . . . . .  | 187 |
| 5.3  | Grammaticality (Index) <i>vs.</i> Acceptability: model fit over small data<br>sample . . . . . | 199 |
| 5.4  | Grammaticality (Index) <i>vs.</i> Acceptability: model fit over the full corpus                | 199 |
| 5.5  | Coherence <i>vs.</i> Acceptability: model fit over the full corpus . . . . .                   | 200 |
| 5.6  | Taxed Coherence <i>vs.</i> Acceptability: model fit over the full corpus . . .                 | 200 |
| 5.7  | Partial parses automatically generated for Type 5.3 (VP-violation) . . .                       | 204 |
| 5.8  | Partial parses automatically generated for Type 5.3 (VP-violation) . . .                       | 205 |
| 5.9  | Partial parses automatically generated for Type 5.3 (VP-violation) . . .                       | 206 |
| 5.10 | Automatically generated parse for Type 2.3 (NP-violation) . . . . .                            | 207 |
| 5.11 | Automatically generated parse for Type 2.3 (NP-violation) . . . . .                            | 207 |
| 5.12 | Substantive adjectives: constituent structures . . . . .                                       | 208 |

|      |   |     |
|------|---|-----|
| 5.13 | Automatically generated parse for Type 2.4 (NP-violation) . . . . . | 210 |
| 5.14 | Automatically generated parse for Type 4.4 (PP-violation) . . . . . | 211 |
| 5.15 | Automatically generated alternative parses . . . . .                | 212 |
| 5.16 | Automatically generated parse for Type 4.4 (PP-violation) . . . . . | 212 |

# List of Algorithms

|   |  |     |
|---|--|-----|
| 1 | Probabilistic CKY . . . . .                | 115 |
| 2 | Loose Satisfaction Chart Parsing . . . . . | 116 |
| 3 | Set Partitioning . . . . .                 | 120 |
| 4 | Characterisation Function . . . . .        | 123 |



# Preface

The research presented in this thesis is the original work of the author except where otherwise indicated. Some parts of the thesis include revised versions of published papers. This work has not been submitted for a higher degree to any other University or Institution than Macquarie University (Australia) and Université de Provence (France).

Jean-Philippe Prost



## Acknowledgements

Foremost, I wish to thank my supervisors Philippe Blache from Université de Provence, and Diego Mollá Aliod and Mark Dras from Macquarie University for their trust, their encouragement, their constant support, their valuable feedback, and their unnumerable advice.

I would also like to thank the members of the examination committee for agreeing to review my work and for their insightful comments: Denys Duchier, Gerald Penn, and Eric Villemonte de la Clergerie.

This work was supported by an international Macquarie University Research Scholarship (iMURS), and by various other travel and visiting grants from Macquarie University, Université de Provence, and Ambassade de France en Australie (*French Embassy in Australia*).

I was lucky enough, during the entire course of this project, to call *home* two different countries and two different institutions, which gave me the great opportunity to meet a very large number of people from different backgrounds. They all had a considerable influence on this work, in one way or another.

From the Australian Connection, special thanks must go to Robert Dale for his always amazing wisdom. I would also like to thank people from Macquarie University and from the Centre for Language Technology: Agnieszka Baginska, Sarah Bedford, Steve Cassidy, Christophe Doche, Frederic Ehrler, Dominique Estival, Maree Graham, Mark Lauer, Joanne Pizzato, Luiz Pizzato, Vanessa Long, Cécile Paris, Brett Powley, Matt Roberts, Rolf Schwitter, Tony Sloane, Kate Stefanov, Dom Verity, and Menno Van Zaanen. I am particularly indebted to my fellow Markists, who made our weekly Reading Group meetings a necessary niche of sanity, humour and wit during The Big Journey: Elena Akhmatova, Mary Gardiner, Ben Hutchinson, Andrew Lampert, Pawel Mazur, Jette Viethen, Stephen Wan, Simon Zwarts, and obviously Mark Dras.

From the French Connection, I am also especially grateful to all the anonymous annotators involved in parts of this work for their time and expertise, as well as to a number of people from *Laboratoire Parole et Langage* for their help and useful feedback: Emmanuel Bellengier, Céline De Looze, Françoise Douay, Gaëlle Ferré, Marie-Laure Guénot, Barbara Hemforth, Cécile Petitjean, Cristel Portes, Tristan VanRullen, and Stéphane Rauzy.

This work also owes a debt to all the volunteers who kindly and courageously agreed to proofread earlier versions of this dissertation: Gail Sinclair, Matt Roberts, Ben Hutchinson, David Wren, Jette Viethen, Jason Malae, Sarah Bedford, Marc Tilbrook and Barbara Mifsud, as well as to various other people for fruitful discussions and comments, and for reviewing parts of this work: Henning Christiansen, Veronica Dahl, Ted Gibson, and Eric Würbel.

A very special mention goes to mum, encouraging and supportive as ever, who certainly did not expect having me back home with my mood and obstination after so long! I certainly do not forget either Anne-Lise, Chloé, Jean-Marc, and Véronique, and I want to warmly thank them here for their unconditional and incredible understanding and support.



## Remerciements

Avant tout, je souhaite remercier mes directeurs de thèse, Philippe Blache de l'Université de Provence, ainsi que Diego Mollá Aliod et Mark Dras de Macquarie University, pour leur confiance, leurs encouragements, leur soutien permanent, et pour la pertinence de leurs innombrables commentaires et remarques.

Je souhaite également remercier les membres du jury pour avoir accepté de corriger mon travail, et pour leurs commentaires experts: Denys Duchier, Gerald Penn et Eric Villemonte de la Clergerie.

Ces travaux ont été supportés financièrement par une allocation internationale de recherche de Macquarie University (iMURS), ainsi que par diverses autres subventions pour déplacements et visites de la part de Macquarie, de l'Université de Provence, et de l'Ambassade de France en Australie.

J'ai eu le privilège, tout au long de ce projet, de me sentir chez moi dans deux pays différents, et deux institutions différentes, ce qui m'a donné l'occasion de faire un très grand nombre de rencontres, aussi variées qu'enrichissantes. Toutes ont eu, d'une façon ou d'une autre, une influence considérable sur ce travail.

Du côté australien, une mention spéciale à l'adresse de Robert Dale pour sa surprenante érudition. Je souhaite également remercier les gens de Macquarie University et du *Centre for Language Technology*, tout particulièrement : Agnieszka Baginska, Sarah Bedford, Steve Cassidy, Christophe Doche, Frederic Ehrler, Dominique Estival, Maree Graham, Mark Lauer, Joanne Pizzato, Luiz Pizzato, Vanessa Long, Cécile Paris, Brett Powley, Matt Roberts, Rolf Schwitter, Tony Sloane, Kate Stefanov, Dom Verity, et Menno Van Zaanen. J'ai une dette toute particulière envers tous les Markistes, qui ont fait de nos réunions hebdomadaires une source nécessaire de revitalisation morale, d'humour et d'esprit tout au long de la Grande Aventure : Elena Akhmatova, Mary Gardiner, Ben Hutchinson, Andrew Lampert, Pawel Mazur, Jette Viethen, Stephen Wan, Simon Zwarts, et bien évidemment Mark Dras.

Du côté français, je suis tout particulièrement reconnaissant envers tous les annotateurs anonymes qui ont participé à cette étude. Leur temps et leur expertise m'ont été précieux. Je remercie également les membres du Laboratoire Parole et Langage pour leur aide et leurs commentaires forts utiles : Emmanuel Bellengier, Céline De Looze, Françoise Douay, Gaëlle Ferré, Marie-Laure Guénot, Barbara Hemforth, Cécile Petitjean, Cristel Portes, Tristan VanRullen, et Stéphane Rauzy.

Ce travail a également une dette non-négligeable envers tous les volontaires qui ont si gentiment et si courageusement accepté de relire différentes versions préliminaires de cette dissertation : Gail Sinclair, Matt Roberts, Ben Hutchinson, David Wren, Jette Viethen, Jason Malae, Sarah Bedford, Marc Tilbrook and Barbara Mifsud. Je remercie également Henning Christiansen, Veronica Dahl, Ted Gibson, et Eric Würbel

pour diverses discussions et commentaires très utiles, et pour avoir accepté de se pencher sur certaines parties de ce travail.

Enfin, une mention toute spéciale à ma mère dont les encouragements et le soutien sont restés sans faille, et qui ne s'attendait certainement pas à devoir à nouveau héberger mes humeurs et mon obstination après si longtemps ! Je n'oublie, bien sûr, pas non plus Anne-Lise, Chloé, Jean-Marc, et Véronique, que je veux remercier ici pour leur compréhension, et leur incroyable et inconditionnel soutien.

*Uno lengo es un clapas ; es uno antico fundamento ounte chasque passant a tra sa pèço d'or o d'argènt o de couire ; es un mounumen inmènse ounte chasco famiho a carreja sa pèiro, ounte chasco ciéuta a basti soun pieloun, ounte uno raço entiero a travaia de cors e d'amo pendènt de cènt e de milo an.*

*Uno lengo, en un mot, es la revelacioun de la vido vidanto, la manifestacioun de la pensado umano, l'estrumen subre-sant di civilisacioun e lou testamen parlant di soucieta morto o vivo.*

(Frederi Mistrau, 1877. *La lengo dóu Miejour*. Discours à l'Assemblado de Santo Estello d'Avignoun, 21 de Mai 1877)

*Une langue est un bloc : c'est un antique fondement où chaque passant a jeté sa pièce d'or, d'argent ou de cuivre : c'est un monument immense où chaque cité a bâti son pilier, où une race entière a travaillé de corps et d'âme pendant des centaines et des milliers d'années.*

*Une langue, en un mot, est la révélation de toute une vie, la manifestation de la pensée humaine, l'instrument sacro-saint des civilisations et le testament parlant des sociétés mortes ou vivantes.*

(Translated from Prouvencau by the *Lexilogos* web site,  
[http://www.lexilogos.com/provencal\\_mistral\\_discours.htm](http://www.lexilogos.com/provencal_mistral_discours.htm),  
as of 22 January 2008)

*A language is a block: it is an antic fundament where every passer-by laid their gold, silver or copper coin: it is an immense monument where every city built its pillar, where an entire race worked from their body and soul during hundreds and thousands of years.*

*A language, in a word, is the revelation of a lifetime, the demonstration of human thought, the sacro-sanct instrument of civilisations and the talking legacy from living or dead societies.*

(Personal translation)

