



## TIPA. Travaux interdisciplinaires sur la parole et le langage

27 | 2008

Apprentissage et acquisition d'une langue seconde

---

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

Thèse en cotutelle de doctorat en Informatique, dirigée par Philippe Blache et Diego Molla Aliod (Macquarie University, Sydney, Australie), soutenue le 10 décembre 2008

*Modelling Syntactic Gradience with Loose Constraint-based Parsing*

Jean-Philippe Prost

---



#### Édition électronique

URL : <https://journals.openedition.org/tipa/423>

DOI : [10.4000/tipa.423](https://doi.org/10.4000/tipa.423)

ISSN : 2264-7082

#### Éditeur

Laboratoire Parole et Langage

#### Édition imprimée

Date de publication : 1 juin 2008

ISSN : 1621-0360

#### Référence électronique

Jean-Philippe Prost, « Modélisation de la gradience syntaxique par analyse relâchée à base de contraintes », *TIPA. Travaux interdisciplinaires sur la parole et le langage* [En ligne], 27 | 2008, mis en ligne le 04 décembre 2012, consulté le 08 septembre 2021. URL : <http://journals.openedition.org/tipa/423> ; DOI : <https://doi.org/10.4000/tipa.423>

---

Ce document a été généré automatiquement le 8 septembre 2021.



La revue *TIPA. Travaux interdisciplinaires sur la parole et le langage* est mise à disposition selon les termes de la licence Creative Commons Attribution - Pas d'Utilisation Commerciale - Pas de Modification 4.0 International.

---

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

Thèse en cotutelle de doctorat en Informatique, dirigée par Philippe Blache et Diego Molla Aliod (Macquarie University, Sydney, Australie), soutenue le 10 décembre 2008

*Modelling Syntactic Gradience with Loose Constraint-based Parsing*

**Jean-Philippe Prost**

---

- 1 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.
- 2 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 fondés 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.

- 3 À travers le développement de ce nouveau modèle, la contribution de ce travail comporte trois volets.
- 4 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 sur la formalisation des 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.
- 5 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 s'appuie 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.
- 6 Et troisièmement, après avoir postulé que la prédiction de jugements humains d'acceptabilité peut se fonder 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.

\*\*\*

- 7 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.
- 8 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.

- 9 Through the development of such a new model of gradience, the main contribution of this work is three-fold.
- 10 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.
- 11 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.
- 12 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.

---

## INDEX

**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

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