# Acceptability Prediction by Means of Grammaticality Quantification

Philippe Blache, Barbara Hemforth, Stéphane Rauzy

## HAL Id: hal-00135422
## https://hal.archives-ouvertes.fr/hal-00135422

Submitted on 7 Mar 2007

# Acceptability Prediction by Means of Grammaticality Quantification

**Philippe Blache, Barbara Hemforth & Stéphane Rauzy**
Laboratoire Parole & Langage
CNRS - Université de Provence
29 Avenue Robert Schuman
13621 Aix-en-Provence, France
`{blache,hemforth,rauzy}@lpl.univ-aix.fr`

## Abstract

We propose in this paper a method for quantifying sentence grammaticality. The approach based on *Property Grammars*, a constraint-based syntactic formalism, makes it possible to evaluate a grammaticality index for any kind of sentence, including ill-formed ones. We compare on a sample of sentences the grammaticality indices obtained from PG formalism and the acceptability judgements measured by means of a psycholinguistic analysis. The results show that the derived grammaticality index is a fairly good tracer of acceptability scores.

## 1   Introduction

Syntactic formalisms make it possible to describe precisely the question of grammaticality. When a syntactic structure can be associated to a sentence, according to a given grammar, we can decide whether or not the sentence is grammatical. In this conception, a language (be it natural or not) is *produced* (or generated) by a grammar by means of a specific mechanism, for example *derivation*. However, when no structure can be built, nothing can be said about the input to be parsed except, eventually, the origin of the failure. This is a problem when dealing with non canonical inputs such as spoken language, e-mails, non-native speaker productions, etc. From this perspective, we need robust approaches that are at the same time capable of describing precisely the form of the input, the source of the problem and to continue the parse. Such capabilities render it possible to arrive at a precise evaluation of the grammaticality of the input. In other words, instead of deciding on the grammaticality of the input, we can give an *indication* of its grammaticality, quantified on the basis of the description of the properties of the input.

This paper addresses the problem of ranking the grammaticality of different sentences. This question is of central importance for the understanding of language processing, both from an automatic and from a cognitive perspective. As for NLP, ranking grammaticality makes it possible to control dynamically the parsing process (in choosing the most adequate structures) or to find the best structure among a set of solutions (in case of non-deterministic approaches). Likewise the description of cognitive processes involved in language processing by human has to explain how things work when faced with unexpected or non canonical material. In this case too, we have to explain why some productions are more acceptable and easier to process than others.

The question of ranking grammaticality has been addressed from time to time in linguistics, without being a central concern. Chomsky, for example, mentioned this problem quite regularly (see for example (Chomsky75)). However he rephrases it in terms of "degrees of 'belongingness' to the language", a somewhat fuzzy notion both formally and linguistically. More recently, several approaches have been proposed illustrating the interest of describing these mechanisms in terms of constraint violations. The idea consists in associating weights to syntactic constraints and to evaluate, either during or after the parse, the weight of violated constraints. This approach is at the basis of *Linear Optimality Theory* (see (Keller00), and (Sorace05) for a more general perspective) in which grammaticality is judged on the basis of the total weights of violated constraints. It is then possible to rank different candidate struc-

tures. A similar idea is proposed in the framework of *Constraint Dependency Grammar* (see (Menzel98), (Schröder02)). In this case too, acceptability is function of the violated constraints weights.

However, constraint violation cannot in itself constitute a measure of grammaticality without taking into account other parameters as well. The type and the number of constraints that are satisfied are of central importance in acceptability judgment: a construction violating 1 constraint and satisfying 15 of them is more acceptable than one violating the same constraint but satisfying only 5 others. In the same way, other informations such as the position of the violation in the structure (whether it occurs in a deeply embedded constituent or higher one in the structure) plays an important role as well.

In this paper, we propose an approach overcoming such limitations. It takes advantage of a fully constraint-based syntactic formalism (called *Property Grammars*, cf. (Blache05b)) that offers the possibility of calculating a grammaticality index, taking into account automatically derived parameters as well as empirically determined weights. This index is evaluated automatically and we present a psycholinguistic study showing how the parser predictions converge with acceptability judgments.

## 2 Constraint-based parsing

Constraints are generally used in linguistics as a control process, verifying that a syntactic structure (e.g. a tree) verifies some well-formedness conditions. They can however play a more general role, making it possible to express syntactic information without using other mechanism (such as a generation function). *Property Grammars* (noted hereafter *PG*) are such a fully constraint-based formalism. In this approach, constraints stipulate different kinds of relation between categories such as linear precedence, imperative co-occurrence, dependency, repetition, etc. Each of these syntactic relations corresponds to a type of constraint (also called property):

- *Linear precedence*: $Det \prec N$ (a determiner precedes the noun)

- *Dependency*: $AP \rightsquigarrow N$ (an adjectival phrase depends on the noun)

- *Requirement*: $V[inf] \Rightarrow to$ (an infinitive comes with *to*)

- *Exclusion*: $seems \nLeftrightarrow ThatClause[subj]$ (the verb *seems* cannot have *That* clause subjects)

- *Uniqueness* : $Uniq_{NP}\{Det\}$ (the determiner is unique in a *NP*)

- *Obligation* : $Oblig_{NP}\{N, Pro\}$ (a pronoun or a noun is mandatory in a *NP*)

- *Constituency* : $Const_{NP}\{Det, AP, N, Pro\}$ (set of possible constituents of *NP*)

In PG, each category of the grammar is described with a set of properties. A grammar is then made of a set of properties. Parsing an input consists in verifying for each category of description the set of corresponding properties in the grammar. More precisely, the idea consists in verifying, for each subset of constituents, the properties for which they are relevant (i.e. the constraints that can be evaluated). Some of these properties are satisfied, some others possibly violated. The result of a parse, for a given category, is the set of its relevant properties together with their evaluation. This result is called *characterization* and is formed by the subset of the satisfied properties, noted $P^+$, and the set of the violated ones, noted $P^-$.

For example, the characterizations associated to the *NPs* "*the book*" and "*book the*" are respectively of the form:
$P^+=\{Det \prec N;\ Det \rightsquigarrow N;\ N \nLeftrightarrow Pro;\ Uniq(Det), Oblig(N), etc.\}$, $P^-=\emptyset$
$P^+=\{Det \rightsquigarrow N;\ N \nLeftrightarrow Pro;\ Uniq(Det), Oblig(N), etc.\}$, $P^-=\{Det \prec N\}$

This approach allows to characterize any kind of syntactic object. In PG, following the proposal made in *Construction Grammar* (see (Fillmore98), (Kay99)), all such objects are called constructions. They correspond to a phrase (NP, PP, etc.) as well as a syntactic turn (cleft, wh-questions, etc.). All these objects are described by means of a set of properties (see (Blache05b)).

In terms of parsing, the mechanism consists in exhibiting the potential constituents of a given construction. This stage corresponds, in constraint solving techniques, to the search of an assignment satisfying the constraint system. The particularity in PG comes from constraint relaxation. Here, the goal is not to find the assignment satisfying the constraint system, but the best assignment (i.e. the one satisfying as much as possible the system). In this way, the PG approach permits to deal with more or less grammatical sentences. Provided that

some control mechanisms are added to the process, PG parsing can be robust and efficient (see (Blache06)) and parse different material, including spoken language corpora.

Using a constraint-based approach such as the one proposed here offers several advantages. First, constraint relaxation techniques make it possible to process any kind of input. When parsing non canonical sentences, the system identifies precisely, for each constituent, the satisfied constraints as well as those which are violated. It furnishes the possibility of parsing any kind of input, which is a pre-requisite for identifying a graded scale of grammaticality. The second important interest of constraints lies in the fact that syntactic information is represented in a non-holistic manner or, in other words, in a decentralized way. This characteristic allows to evaluate precisely the syntactic description associated with the input. As shown above, such a description is made of sets of satisfied and violated constraints. The idea is to take advantage of such a representation for proposing a quantitative evaluation of these descriptions, elaborated from different indicators such as the number of satisfied or violated constraints or the number of evaluated constraints.

The hypothesis, in the perspective of a gradience account, is to exhibit a relation between a quantitative evaluation and the level of grammaticality: the higher the evaluation value, the more grammatical the construction. The value is then an indication of the quality of the input, according to a given grammar. In the next section we propose a method for computing this value.

## 3 Characterization evaluation

The first idea that comes to mind when trying to quantify the quality of a characterization is to calculate the ratio of satisfied properties with respect to the total set of evaluated properties. This information is computed as follows:

Let $C$ a construction defined in the grammar by means of a set of properties $S_C$, let $A_C$ an assignment for the construction $C$,

- $P^+$ = set of satisfied properties for $A_C$

- $P^-$ = set of violated properties for $A_C$

- $N^+$ : number of satisfied properties $N^+ = card(P^+)$

- $N^-$ : number of violated properties $N^- = card(P^-)$

- *Satisfaction ratio* ($SR$): the number of satisfied properties divided by the number of evaluated properties $SR = \frac{N^+}{E}$

The SR value varies between 0 and 1, the two extreme values indicating that no properties are satisfied (*SR=0*) or none of them are violated (*SR=1*). However, SR only relies on the evaluated properties. It is also necessary to indicate whether a characterization uses a small or a large subpart of the properties describing the construction in the grammar. For example, the *VP* in our grammar is described by means of 25 constraints whereas the *PP* only uses 7 of them. Let's imagine the case where 7 constraints can be evaluated for both constructions, with an equal *SR*. However, the two constructions do not have the same quality: one relies on the evaluation of all the possible constraints (in the *PP*) whereas the other only uses a few of them (in the *VP*). The following formula takes these differences into account :

- $E$ : number of relevant (i.e. evaluated) properties $E = N^+ + N^-$

- $T$= number of properties specifying construction $C = card(SC)$

- Completeness coefficient ($CC$) : the number of evaluated properties divided by the number of properties describing the construction in the grammar $CC = \frac{E}{T}$

These purely quantitative aspects have to be contrasted according to the constraint types. Intuitively, some constraints, for a given construction, play a more important role than some others. For example, linear precedence in languages with poor morphology such as English or French may have a greater importance than obligation (i.e. the necessity of realizing the head). To its turn, obligation may be more important than uniqueness (i.e. impossible repetition). In this case, violating a property would have different consequences according to its relative importance. The following examples illustrate this aspect:

(1)   a. *The the man who spoke with me is my brother.*
      b. *The who spoke with me man is my brother.*

In (1a), the determiner is repeated, violating a uniqueness constraint of the first *NP*, whereas (1c) violates a linearity constraint of the same *NP*.

Clearly, (1a) seems to be more grammatical than (1b) whereas in both cases, only one constraint is violated. This contrast has to be taken into account in the evaluation. Before detailing this aspect, it is important to note that this intuition does not mean that constraints have to be organized into a ranking scheme, as with the *Optimality Theory* (see (Prince93)). The parsing mechanism remains the same with or without this information and the hierarchization only plays the role of a process control.

Identifying a relative importance of the types of constraints comes to associate them with a weight. Note that at this stage, we assign weights to constraint types, not directly to the constraints, differently from other approaches (cf. (Menzel98), (Foth05)). The experiment described in the next section will show that this weighting level seems to be efficient enough. However, in case of necessity, it remains possible to weight directly some constraints into a given construction, overriding thus the default weight assigned to the constraint types.

The notations presented hereafter are used to describe constraint weighting. Remind that $P^+$ and $P^-$ indicate the set of satisfied and violated properties of a given construction.

- $p_i^+$ : property belonging to $P^+$

- $p_i^-$ : property belonging to $P^-$

- $w(p)$ : weight of the property of type $p$

- $W^+$ : sum of the satisfied properties weights

$$W^+ = \sum_{i=1}^{N^+} w(p_i^+)$$

- $W^-$ : sum of the violated properties weights

$$W^- = \sum_{i=1}^{N^-} w(p_i^-)$$

One indication of the relative importance of the constraints involved in the characterization of a construction is given by the following formula:

- *QI*: the quality index of a construction

$$QI = \frac{W^+ - W^-}{W^+ + W^-}$$

The $QI$ index varies then between -1 and 1. A negative value indicates that the set of violated constraints has a greater importance than the set of satisfied one. This does not mean that more constraints are violated than satisfied, but indicates the importance of the violated ones.

We now have three different indicators that can be used in the evaluation of the characterization: the satisfaction ratio (noted $SR$) indicating the ratio of satisfied constraints, the completeness coefficient (noted $CC$) specifying the ratio of evaluated constraints, and the quality index (noted $QI$) associated to the quality of the characterization according to the respective degree of importance of evaluated constraints. These three indices are used to form a global precision index (noted $PI$). These three indicators do not have the same impact in the evaluation of the characterization, they are then balanced with coefficients in the normalized formula:

- $PI = \frac{(k \times QI) + (l \times SR) + (m \times CC)}{3}$

As such, $PI$ constitutes an evaluation of the characterization for a given construction. However, it is necessary to take into account the "quality" of the constituents of the construction as well. A construction can satisfy all the constraints describing it, but can be made of embedded constituents more or less well formed. The overall indication of the quality of a construction has then to integrate in its evaluation the quality of each of its constituents. This evaluation depends finally on the presence or not of embedded constructions. In the case of a construction made of lexical constituents, no embedded construction is present and the final evaluation is the precision index *PI* as described above. We will call hereafter the evaluation of the quality of the construction the "*grammaticality index*" (noted $GI$). It is calculated as follows:

- Let $d$ the number of embedded constructions

- If $d = 0$ then $GI = PI$, else

$$GI = PI \times \frac{\sum_{i=1}^{d} GI(C_i)}{d}$$

In this formula, we note $GI(C_i)$ the grammaticality index of the construction $C_i$. The general formula for a construction $C$ is then a function of its precision index and of the sum of the grammaticality indices of its embedded constituents. This

formula implements the propagation of the quality of each constituent. This means that the grammaticality index of a construction can be lowered when its constituents violate some properties. Reciprocally, this also means that violating a property at an embedded level can be partially compensated at the upper levels (provided they have a good grammaticality index).

## 4 Grammaticality index from PG

We describe in the remainder of the paper predictions of the model as well as the results of a psycholinguistic evaluation of these predictions. The idea is to evaluate for a given set of sentences on the one hand the grammaticality index (done automatically), on the basis of a PG grammar, and on the other hand the acceptability judgment given by a set of subjects. This experiment has been done for French, a presentation of the data and the experiment itself will be given in the next section. We present in this section the evaluation of grammaticality index.

Before describing the calculation of the different indicators, we have to specify the constraints weights and the balancing coefficients used in PI. These values are language-dependent, they are chosen intuitively and partly based on earlier analysis, this choice being evaluated by the experiment as described in the next section. In the remainder, the following values are used:

| Constraint type | Weight |
|---|---|
| *Exclusion, Uniqueness, Requirement* | 2 |
| *Obligation* | 3 |
| *Linearity, Constituency* | 5 |

Concerning the balancing coefficients, we give a greater importance to the quality index (coefficient $k=2$), which seems to have important consequences on the acceptability, as shown in the previous section. The two other coefficients are significatively less important, the satisfaction ratio being at the middle position (coefficient $l=1$) and the completeness at the lowest (coefficient $m=0,5$).

Let's start with a first example, illustrating the process in the case of a sentence satisfying all constraints.

(2) Marie a emprunté un très long chemin pour le retour.
*Mary took a very long way for the return.*

The first NP contains one lexical constituent, Mary. Three constraints, among the 14 describing the *NP*, are evaluated and all satisfied: *Oblig(N)*, stipulating that the head is realized, *Const(N)*, in-

dicating the category *N* as a possible constituent, and *Excl(N, Pro)*, verifying that *N* is not realized together with a pronoun. The following values come from this characterization:

| N+ | N- | E | T | W+ | W- | QI | SR | CC | PI | GI |
|---|---|---|---|---|---|---|---|---|---|---|
| 3 | 0 | 3 | 14 | 10 | 0 | 1 | 1 | 0.21 | 1.04 | 1.04 |

We can see that, according to the fact that all evaluated constraints are satisfied, *QI* and *SR* equal 1. However, the fact that only 3 constraints among 14 are evaluated lowers down the grammatical index. This last value, insofar as no constituents are embedded, is the same as *PI*.

These results can be compared with another constituent of the same sentence, the *VP*. This construction also only contains satisfied properties. Its characterization is the following : *Char(VP)=Const(Aux, V, NP, PP) ; Oblig(V) ; Uniq(V) ; Uniq(NP) ; Uniq(PP) ; Aux⇒V[part] ; V≺NP ; Aux≺V ; V≺PP*. On top of this set of evaluated constraints (9 among the possible 25), the VP includes two embedded constructions : a *PP* and a *NP*. A grammaticality index has been calculated for each of them: *GI(PP) = 1.24 GI(NP)=1.23*. The following table indicates the different values involved in the calculation of the *GI*.

| N+ | N- | E | T | W+ | W- | QI | SR | CC | PI |
|---|---|---|---|---|---|---|---|---|---|
| 9 | 0 | 9 | 25 | 31 | 0 | 1 | 1 | 0.36 | 1.06 |

| GI_Emb_Const | GI |
|---|---|
| 1.23 | 1.31 |

The final *GI* of the *VP* reaches a high value. It benefits on the one hand from its own quality (indicated by *PI*) and on another hand from that of its embedded constituents. In the end, the final *GI* obtained at the sentence level is function of its own *PI* (very good) and the *NP* and *VP GI*s, as shown in the table:

| N+ | N- | E | T | W+ | W- | QI | SR | CC | PI |
|---|---|---|---|---|---|---|---|---|---|
| 5 | 0 | 5 | 9 | 17 | 0 | 1 | 1 | 0.56 | 1.09 |

| GI_Emb_Const | GI |
|---|---|
| 1.17 | 1.28 |

Let's compare now these evaluations with those obtained for sentences with violated constraints, as in the following examples:

(3) a. Marie a emprunté très long chemin un pour le retour.
*Mary took very long way a for the return.*

b. Marie a emprunté un très chemin pour le retour.
*Mary took a very way for the return.*

In (2a), 2 linear constraints are violated: a determiner follows a noun and an *AP* in "*très long chemin un*". Here are the figures calculated for this *NP*:

| N+ | N- | E | T | W+ | W- | QI | SR | CC | PI | GI |
|---|---|---|---|---|---|---|---|---|---|---|
| 8 | 2 | 10 | 14 | 23 | 10 | 0.39 | 0.80 | 0.71 | 0.65 | 0.71 |

The *QI* indicator is very low, the violated constraints being of heavy weight. The grammaticality index is a little bit higher because a lot of constraints are also satisfied. The *NP GI* is then propagated to its dominating construction, the *VP*. This phrase is well formed and also contains a well-formed construction (*PP*) as sister of the *NP*. Note that in the following table summarizing the *VP* indicators, the *GI* product of the embedded constituents is higher than the *GI* of the *NP*. This is due to the well-formed *PP* constituent. In the end, the *GI* index of the *VP* is better than that of the ill-formed *NP*:

| N+ | N- | E | T | W+ | W- | QI | SR | CC | PI |
|----|----|----|----|----|----|----|----|----|----|
| 9 | 0 | 9 | 25 | 31 | 0 | 1 | 1 | 0.36 | 1.06 |

| GI_Emb_Const | GI |
|----|----|
| 0.97 | 1.03 |

For the same reasons, the higher level construction *S* also compensates the bad score of the *NP*. However, in the end, the final *GI* of the sentence is much lower than that of the corresponding well-formed sentence (see above).

| N+ | N- | E | T | W+ | W- | QI | SR | CC | PI |
|----|----|----|----|----|----|----|----|----|----|
| 5 | 0 | 5 | 9 | 17 | 0 | 1 | 1 | 0.56 | 1.09 |

| GI_Emb_Const | GI |
|----|----|
| 1.03 | 1.13 |

The different figures of the sentence (2b) show that the violation of a unique constraint (in this case the *Oblig(Adj)* indicating the absence of the head in the *AP*) can lead to a global lower *GI* than the violation of two heavy constraints as for (2a). In this case, this is due to the fact that the AP only contains one constituent (a modifier) that does not suffice to compensate the violated constraint. The following table indicates the indices of the different phrases. Note that in this table, each phrase is a constituent of the following (i.e. *AP* belongs to *NP* itself belonging to *VP*, and so on).

| | N+ | N- | E | T | W+ | W- | QI | SR | CC | PI |
|----|----|----|----|----|----|----|----|----|----|----|
| AP | 2 | 1 | 3 | 7 | 7 | 3 | 0.40 | 0.67 | 0.43 | 0.56 |
| NP | 10 | 0 | 10 | 14 | 33 | 0 | 1 | 1 | 0.71 | 1.12 |
| VP | 9 | 0 | 9 | 25 | 31 | 0 | 1 | 1 | 0.36 | 1.06 |
| S | 5 | 0 | 5 | 9 | 17 | 0 | 1 | 1 | 0.56 | 1.09 |

| | GI_Emb_Const | GI |
|----|----|----|
| AP | 1 | 0.56 |
| NP | 0.56 | 0.63 |
| VP | 0.93 | 0.99 |
| S | 1.01 | 1.11 |

## 5  Judging acceptability of violations

We ran a questionnaire study presenting participants with 60 experimental sentences like (11) to (55) below. 44 native speakers of French completed the questionnaire giving acceptability judgements following the Magnitude Estimation technique. 20 counterbalanced forms of the questionnaire were constructed. Three of the 60 experimental sentences appeared in each version in each form of the questionnaire, and across the 20 forms, each experimental sentence appeared once in each condition. Each sentence was followed by a question concerning its acceptability. These 60 sentences were combined with 36 sentences of various forms varying in complexity (simple main clauses, simple embeddings and doubly nested embeddings) and plausibility (from fully plausible to fairly implausible according to the intuitions of the experimenters). One randomization was made of each form.

*Procedure*: The rating technique used was magnitude estimation (ME, see (Bard96)). Participants were instructed to provide a numeric score that indicates how much better (or worse) the current sentence was compared to a given reference sentence (Example: If the reference sentence was given the reference score of 100, judging a target sentence five times better would result in 500, judging it five times worse in 20). Judging the acceptability ratio of a sentence in this way results in a scale which is open-ended on both sides. It has been demonstrated that ME is therefore more sensitive than fixed rating-scales, especially for scores that would approach the ends of such rating scales (cf. (Bard96)). Each questionnaire began with a written instruction where the subject was made familiar with the task based on two examples. After that subjects were presented with a reference sentence for which they had to provide a reference score. All following sentences had to be judged in relation to the reference sentence. Individual judgements were logarithmized (to arrive at a linear scale) and normed (z-standardized) before statistical analyses.

Global mean scores are presented figure 1. We tested the reliability of results for different randomly chosen subsets of the materials. Constructions for which the judgements remain highly stable across subsets of sentences are marked by an asterisk (*rs > 0.90; p < 0.001*). The mean reliability across subsets is *rs > 0.65 (p < 0.001)*.

What we can see in these data is that in particular violations within prepositional phrases are not judged in a very stable way. The way they are judged appears to be highly dependent on the preposition used and the syntactic/semantic context. This is actually a very plausible result, given that heads of prepositional phrases are closed class items that are much more predictable in many syntactic and semantic environments than heads of

noun phrases and verb phrases. We will therefore base our further analyses mainly on violations within noun phrases, verb phrases, and adjectival phrases. Results including prepositional phrases will be given in parentheses. Since the constraints described above do not make any predictions for semantic violations, we excluded examples 25, 34, 45, and 55 from further analyses.

# 6 Acceptability versus grammaticality index

We compare in this section the results coming from the acceptability measurements described in section 5 and the values of grammaticality indices obtained as proposed section 4.

From the sample of 20 sentences presented in figure 1, we have discarded 4 sentences, namely sentence 25, 34, 45 and 55, for which the property violation is of semantic order (see above). We are left with 16 sentences, the reference sentence satisfying all the constraints and 15 sentences violating one of the syntactic constraints. The results are presented figure 2. Acceptability judgment (ordinate) versus grammaticality index (abscissa) is plotted for each sentence. We observe a high coefficient of correlation ($\rho = 0.76$) between the two distributions, indicating that the grammaticality index derived from PG is a fairly good tracer of the observed acceptability measurements.

The main contribution to the grammaticality index comes from the quality index $QI$ ($\rho = 0.69$) while the satisfaction ratio $SR$ and the complete-

| No violations | |
|---|---|
| 11. Marie *a emprunté un très long chemin pour le retour* | 0.465 |

| NP-violations | |
|---|---|
| 21. *Marie a emprunté très long chemin un pour le retour* | -0.643 * |
| 22. *Marie a emprunté un très long chemin chemin pour le retour* | -0.161 * |
| 23. *Marie a emprunté un très long pour le retour* | -0.871 * |
| 24. *Marie a emprunté très long chemin pour le retour* | -0.028 * |
| 25. *Marie a emprunté un très heureux chemin pour le retour* | -0.196 * |

| AP-violations | |
|---|---|
| 31. *Marie a emprunté un long très chemin pour le retour* | -0.41 * |
| 32. *Marie a emprunté un très long long chemin pour le retour* | -0.216 - |
| 33. *Marie a emprunté un très chemin pour le retour* | -0.619 - |
| 34. *Marie a emprunté un grossièrement long chemin pour le retour* | -0.058 * |

| PP-violations | |
|---|---|
| 41. Marie *a emprunté un très long chemin le retour pour* | -0.581 - |
| 42. *Marie a emprunté un très long chemin pour pour le retour* | -0.078 - |
| 43. *Marie a emprunté un très long chemin in le retour* | -0.213 - |
| 44. Marie *a emprunté un très long chemin pour* | -0.385 - |
| 45. *Marie a emprunté un très long chemin dans le retour* | -0.415 - |

| VP-violations | |
|---|---|
| 51. *Marie un très long chemin a emprunté pour le retour* | -0.56 * |
| 52.*Marie a emprunté emprunté un très long chemin pour le retour* | -0.194 * |
| 53.Marie *un très long chemin pour le retour* | -0.905 * |
| 54. *Marie emprunté un très long chemin pour le retour* | -0.322 * |
| 55. *Marie a persuadé un très long chemin pour le retour* | -0.394 * |

Figure 1: Acceptability results

ness coefficient CC contributions, although significant, are more modest ($\rho = 0.18$ and $\rho = 0.17$ respectively).

We present in figure 3 the correlation between acceptability judgements and grammaticality indices after the removal of the 4 sentences presenting *PP* violations. The analysis of the experiment described in section 5 shows indeed that acceptability measurements of the PP-violation sentences is less reliable than for others phrases. We thus expect that removing these data from the sample will strengthen the correlation between the two distributions. The coefficient of correlation of the 12 remaining data jumps to $\rho = 0.87$, as expected.
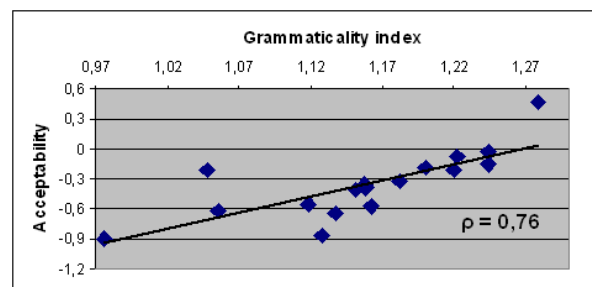


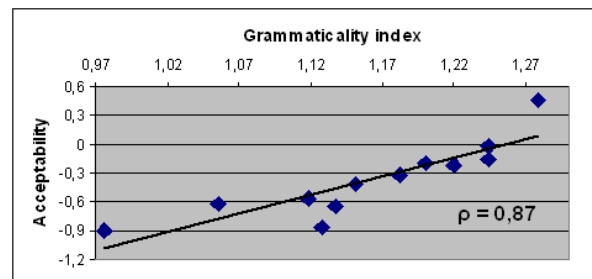Figure 2: Correlation between acceptability judgement and grammaticality index



Figure 3: Correlation between acceptability judgement and grammaticality index removing PP violations

Finally, the adequacy of the PG grammaticality indices to the measurements was investigated by means of resultant analysis. We adapted the parameters of the model in order to arrive at a good fit based on half of the sentences materials (randomly chosen from the full set), with a correlation of $\rho = 0.85$ ($\rho = 0.76$ including *PP*s) between the grammaticality index and acceptability judgements. Surprisingly, we arrived at the best fit with only two different weights: A weight of 2 for *Exclusion*, *Uniqueness*, and *Requirement*, and a weight of 5 for *Obligation*, *Linearity*, and *Constituency*. This result converges with the hard

and soft constraint repartition idea as proposed by (Keller00).

The fact that the grammaticality index is based on these properties as well as on the number of constraints to be evaluated, the number of constraints to the satisfied, and the goodness of embedded constituents apparently results in a fined grained and highly adequate prediction even with this very basic distinction of constraints.

Fixing these parameters, we validated the predictions of the model for the remaining half of the materials. Here we arrived at a highly reliable correlation of $\rho = 0.86$ ($\rho = 0.67$ including *PP*s) between PG grammaticality indices and acceptability judgements.

## 7 Conclusion

The method described in this paper makes it possible to give a quantified indication of sentence grammaticality. This approach is direct and takes advantage of a constraint-based representation of syntactic information, making it possible to represent precisely the syntactic characteristics of an input in terms of satisfied and (if any) violated constraints. The notion of *grammaticality index* we have proposed here integrates different kind of information: the quality of the description (in terms of well-formedness degree), the density of information (the quantity of constraints describing an element) as well as the structure itself. These three parameters are the basic indicators of the grammaticality index.

The relevance of this method has been experimentally shown, and the results described in this paper illustrate the correlation existing between the prediction (automatically calculated) expressed in terms of GI and the acceptability judgment given by subjects.

This approach also presents a practical interest: it can be directly implemented into a parser. The next step of our work will be its validation on large corpora. Our parser will associate a grammatical index to each sentence. This information will be validated by means of acceptability judgments acquired on the basis of a sparse sampling strategy.

## References

Bard E., D. Robertson & A. Sorace (1996) "Magnitude Estimation of Linguistic Acceptability", Language 72:1.

Blache P. & J.-P. Prost (2005) "Gradience, Constructions and Constraint Systems", in H. Christiansen & al. (eds), *Constraint Solving and NLP*, Lecture Notes in Computer Science, Springer.

Blache P. (2005) "Property Grammars: A Fully Constraint-Based Theory", in H. Christiansen & al. (eds), *Constraint Solving and NLP*, Lecture Notes in Computer Science, Springer.

Blache P. (2006) "A Robust and Efficient Parser for Non-Canonical Inputs", in proceedings of *Robust Methods in Analysis of Natural Language Data*, EACL workshop.

Chomsky N.. (1975) *The Logical Structure of Linguistic Theory*, Plenum Press

Croft W. & D. Cruse (2003) *Cognitive Linguistics*, Cambridge University Press.

Foth K., M. Daum & W. Menzel (2005) "Parsing Unrestricted German Text with Defeasible Constraints", in H. Christiansen & al. (eds), *Constraint Solving and NLP*, Lecture Notes in Computer Science, Springer.

Fillmore C. (1998) "Inversion and Contructional Inheritance", in *Lexical and Constructional Aspects of Linguistic Explanation*, Stanford University.

Kay P. & C. Fillmore (1999) "Grammatical Constructions and Linguistic Generalizations: the *what's x doing y* construction", Language.

Keller F. (2000) *Gradience in Grammar. Experimental and Computational Aspects of Degrees of Grammaticality*, Phd Thesis, University of Edinburgh.

Keller F. (2003) "A probabilistic Parser as a Model of Global Processing Difficulty", in proceedings of *ACCSS-03*

Menzel W. & I. Schroder (1998) "Decision procedures for dependency parsing using graded constraints", in S. Kahane & A. Polguère (eds), Proc. Coling-ACL Workshop on Processing of Dependency-based Grammars.

Prince A. & Smolensky P. (1993) *Optimality Theory: Constraint Interaction in Generative Grammars*, Technical Report RUCCS TR-2, Rutgers Center for Cognitive Science.

Sag I., T. Wasow & E. Bender (2003) *Syntactic Theory. A Formal Introduction*, CSLI.

Schröder I. (2002) *Natural Language Parsing with Graded Constraints*. PhD Thesis, University of Hamburg.

Sorace A. & F. Keller (2005) "Gradience in Linguistic Data", in *Lingua*, 115.