Gradience, Constructions and Constraint Systems

Philippe Blache, Jean-Philippe Prost

HAL Id: hal-00134201

https://hal.science/hal-00134201

Submitted on 1 Mar 2007

Gradience, Constructions and Constraint Systems

Philippe Blache[1] and Jean-Philippe Prost[2, 1]

[1] LPL-CNRS
Université de Provence
29 Avenue Robert Schuman
13621 Aix-en-Provence, France
pb@lpl.univ-aix.fr
[2] Centre for Language Technology
Macquarie University
Sydney NSW 2109, Australia
jpprost@ics.mq.edu.au

Abstract. This paper is concerned with the question of quantifying gradient degrees of acceptability by introducing the notion of *Density* in the context of constructional constraint language processing. We first present here our framework for language processing, where all linguistic knowledge is represented by means of constraints. The grammar itself is a constraint system. A constraint is a relation among categories, which encodes a linguistic property. But in contrast to more traditional constraint-based approaches, a constraint can hold and be assessed independently from the structure. In this context, we then introduce the notion of density, based on proportions of satisfied and violated linguistic properties. Our intuition is that density can be used as a means to measure fuzzy notions such as syntactic complexity or as a criterion to identify gradient levels of acceptability. We present and discuss early experimental results concerning density.

1 Introduction

One important question to be addressed by modern linguistics concerns the variability of linguistic constructions. Some are very frequent, some are quite rare. Some are easy to explain, some others are hard. And finally, some are fully grammatical whereas some others are less grammatical. These different aspects have been addressed, usually separately, from a psycholinguistic perspective. Some elements of explanation are given in terms of sentence complexity and gradience. Complexity has to do with the fact that some utterances can be interpreted more easily than others (see for example [1] Gibson, 2000). In the literature, *Gradience*[3] usually refers to categorial vagueness and reflects the fact that gradient phenomena are observed 'at all levels of linguistic analysis, from phonology, morphology and lexis on the one hand, to syntax and semantics on the other' ([3] Aarts, 2004) (see also in particular [4] Bresnan, 2003, [5] Keller, 2000, and [6] Sorace, 2004). Bas Aarts in [3] (2004) is even more specific and distinguishes between *Subjective Gradience* for 'degree of resemblance to a categorial prototype' and *Intersective Gradience* for 'degree of inter-categorial resemblance'.

---

[3] The term seems to appear in the english linguistic literature since 1961 with Bolinger in his book *Generality, Gradience, and the All–or–None* ([2] Bolinger, 1961).

Several linguistic theories address explicitly such questions, in particular from the perspective of dealing with ill-formed inputs. Some elements of answer can be found for example in the *Optimality Theory* (see [7] Prince, 1993), in the *Model-Theoretic Syntax* approach (see [8] Pullum, 2001) or in *Construction Grammar* (see [9] Fillmore, 1998, [10] Goldberg, 1995). One of the main challenges in these approaches is the characterization of gradience in linguistic data. The basic idea consists in hierarchizing the linguistic information according to some "importance" criterion. However, such importance is difficult to define. In some approaches such as probabilistic grammars, it relies on frequency information (see [11] Keller, 2003): each rule is weighted according to its frequency acquired on treebanks. The syntactic constructions are specified according to the weights of the different rules. In some other approaches, explored in this paper, the idea is to propose some objective information relying on a symbolic representation.

The aim of this paper is to present a work still in progress concerning the notion of density and to put forward some elements of discussion about it. The basic intuition is that we can probably measure how dense is an utterance in terms of linguistic information and then use this value of density for different purposes, such as evaluating the syntactic complexity of an utterance, or maybe also predicting where the information is located within a text. In the context of Information Retrieval or Question/Answering, for example, we can possibly make the assumption that the more dense a part of a text, the more likely the part is to carry information of some relevance. In the present paper we mainly focus on discussing to what extent does the density suit the need for representing and measuring gradience. In this respect we will see how the density might be used to rank utterances according to their degree of acceptability.

In the first and second parts of this paper we present a constructional fully constraint-based approach representing all kinds of information by means of constraints, which makes it possible to quantify information. In a third part we introduce a notion of density and discuss its suitability with regard to quantifying gradience. Then in a fourth part we present an early experiment where we compare different ways of computing density of utterances. We discuss the different methods and give a first intuitive interpretation of different figures for each of the methods. The last two parts are concerned with further work and conclusion.

2 Constructions

Several modern approaches in linguistics argue in favor of a contextual description of linguistic phenomena. This way of representing information is typically proposed in Construction Grammar (see [9] Fillmore, 1998 or [12] Kay, 1999), in Property Grammars (see [13] Blache, 2000), etc. In these approaches, a specific phenomenon is characterized by a convergence of different properties. In [12] Kay, 1999,

> A *construction* (e.g., the Subject Auxilary Inversion construction) is a set of conditions licensing a class of actual constructs[4] of a language (e.g. the class of English inverted clauses, (...)).

---

[4] where the *constructs* are the words, phrases and sentences of the language licensed by the *constructions*.

For example, taking into account the syntactic level alone, a dislocated construction in French is roughly characterized by the realization of an NP, before or after a main clause, together with an anaphoric clitic inside this clause. Other properties can be added to this framework, for example concerning the lexical selection that can be specific to this construction. The important point is that a construction is specified by a given set of properties, some of them being relevant only for this construction. In other words, given the fact that, as proposed in Property Grammars, a property can be conceived as a constraint, a construction is defined by a constraint system. Thus what is important at the construction level is not a property taken separately from the others, but the interaction of all the constraints (see the definition of a construction using Property Grammars in section 3.3). Contextuality is then implemented by means of such an interaction: defining a linguistic form consists in specifying on the one hand a minimum set of properties which characterizes the particular construction and on the other hand some additional properties according to different context for the construction. As, for example, in the description of passive in French, where an accusative pronoun in the VP must agree with the subject—must be reflexive (see (1)[5] and (2) for example). This property is characteristic of the construction; other constraints are also needed in the construction definition in order to fully describe all the possible properties for all the licensed construtcts in different contexts. Consequently, a fully grammatical construct can be licensed by only satisfying a subset of the constraint system associated with a particular construction.

(1) Je me le suis dit (*I myself it aux-1st said*)
(2) *Je te le suis dit (*I you it aux-1st said*)

Such an approach presents several advantages. First, it is possible to express constraints at very different granularity levels for the specification of a given construction. Constraints can hold on lexical categories, on phrasal categories, or even on more complex constructions. Moreover, different kinds of constraints, coming from different linguistic domains such as prosody, semantics, pragmatics, etc. can contribute to the definition of a construction. It is one of the main arguments in favor of construction approaches. This aspect is exemplified in the following utterance (from [14] Mertens, 1993) illustrating a very specific construction in which only very little information is available at the syntactic level and no punctuation at all is available either given that we deal with a spoken utterance. In this case, prosody plays an important role in its interpretation:

(3) lundi lavage mardi repassage mercredi repos
    monday washing tuesday ironing wednesday rest

Finally, a constraint can have in such perspective a relative importance. For some constructions, a constraint can be obligatory whereas the same constraint can be easily relaxed in some other cases.

---

[5] (1) means *I said it to myself.*

3 Constraints

Classically, constraints are used in linguistic theories as a filtering process. This is typically the case with constraint grammars, but also with most recent constraint-based approaches such as HPSG (see [15] Sag, 1999) or Optimality (see [7] Prince, 1993). In HPSG for example, constraints are applied to a structure in order to verify its well-formedness. As a side effect, constraints can also implement feature values instantiation or propagation. The valence principle for example plays exactly this double role: ruling out the structures that don't satisfy the constraint and, in case of unification between a structure and a description in the valence list, instantiating some feature values of the structure. In this case, constraints are seen as structure descriptions; they don't implement information that can possibly be evaluated independently from these structures. This means that structures are first to be built before verifying their values, and syntactic properties are expressed in terms of relations inside such hierarchized constructions.

Constraints are used in a completely different way in Optimality Theory (OT). They also constitute a filtering process, but the constraints belong to a system containing two main pieces of information: the basic information specified by the constraint itself, expressed in universal and imperative terms and a second-level (rather implicit) information expressed by ranking. In such a system, the fact that a constraint is satisfied or not is in the end less important than its position in the ranking. Moreover, all constraints are stipulated taking into account the fact that other opposite constraints also belong to the system. This is a kind of negative way of using constraints that are in fact stipulated so as to be violated.

There is a common basis of these different uses of constraints. In all cases, they need to be interpreted into a system. In other words, they cannot be evaluated in isolation, but in reference to the entire system. This is what Pullum stresses as being a holistic way of seeing syntactic information in which a syntactic property cannot be interpreted in itself. This is a limitation in the perspective of finding syntactic characterizations of unrestricted material: in many cases, especially when parsing spoken languages, syntactic information is sparse. For example in (3), which is transcribed from a corpus of spoken language, the relation among the different elements is difficult to express at the syntactic level.

In such cases, information comes from the interaction of different linguistic domains, in particular morphology and prosody more than other kinds of information. And in such cases, classical approaches fail to build a description. There is also another drawback. In the case of OT for example, ranking makes it possible to order different candidates. Such ranking expresses a level of well-formedness, according to the grammar. However, there is no direct relation between well-formedness and more general notions such as understandability, acceptability, sentence complexity, etc. What is important to explain from a cognitive point of view is what kind of utterances are more easily interpretable and why.

We think that constraints can play an important role in this perspective provided that they are expressed in a non holistic manner. In such a perspective, each constraint must implement a syntactic property and be expressed independently from the others. Obviously, constraints have to interact, but they can always be evaluated. This characteristic is highlighted by Pullum in [8] 2001, as being one interest of a Model-Theoretical ap-

proach in comparison with a deductive one: it is possible to give some information in all cases, whatever the input form.

Since Maruyama who first applied pure constraint satisfaction techniques to natural language processing in [16] 1990, other non-generative approaches of the kind have been presented, such as the work on dependency parsing from Duchier [17], 1999; the Weighted Constraint Dependency Grammar from [18] Heinecke et al., 1998; or Constraint Handling Rule Grammars from [19] Christiansen, 2005. However none of these approaches directly addresses the problem of gradience.

3.1 Property Grammars (PG)

We briefly describe here a framework for such an approach called *Property Grammars* (see [13] Blache, 2000). In this approach, all information is represented by means of constraints (also called properties). These constraints are relations among categories expressed as a set of features. A category, unlike in HPSG, doesn't contain hierarchical information among constituents, but only what is called in construction grammar *intrinsic* information. All constraints are expressed independently from the others and represent a specific kind of syntactic information:

- *linear precedence*, which is an order relation among constituents,
- *subcategorization*, which indicates the co-occurrence relations among categories or sets of categories,
- *exclusion*, which specifies an impossibility of co-occurrence between categories,
- *unicity*, which is the impossibility of the repetition of a category,
- *obligation*, which lists the minimal set of obligatory constituents (usually one single constituent) which is the head,
- *dependency*, which represents the semantic relations among categories, in terms of dependency.

These different kinds of information correspond to different properties, respectively: *Linearity, Requirement, Exclusion, Unicity, Obligation, Dependency*. Such information can always be expressed in terms of relations among categories, as shown in the following examples:

- 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: $most \not\Leftrightarrow Adj[super]$ (*most* can not modify an adjective, which morphology has already the mark of a superlative.)

Here is a more formal representation of such information :

let $\mathcal{K}$ be a set of categories, $\mathcal{A}$ be the ordered set of categories for a given input,
let $pos(\mathcal{C}, \mathcal{A})$ be a function that returns the position of $\mathcal{C}$ in $\mathcal{A}$,
let $card(\mathcal{C}, \mathcal{A})$ be a function that returns the number of elements of type $\mathcal{C}$ in $\mathcal{A}$,
let $\{\mathcal{C}_1, \mathcal{C}_2\} \in \mathcal{K}$,

let $comp(\mathcal{C}_1, \mathcal{C}_2)$ be a function that verifies the semantic compatibility of $\mathcal{C}_1$ and $\mathcal{C}_2$ and that completes the semantic structure of $\mathcal{C}_2$ with that of $\mathcal{C}_1$[6]

- LP: $\mathcal{C}_1 \prec \mathcal{C}_2$ holds in $\mathcal{A}$ iff $pos(\mathcal{C}_1, \mathcal{A}) < pos(\mathcal{C}_2, \mathcal{A})$
- Req: $\mathcal{C}_1 \Rightarrow \mathcal{C}_2$ holds in $\mathcal{A}$ iff $\mathcal{C}_1 \notin \mathcal{A}$ or $\mathcal{C}_2 \in \mathcal{A}$
- Excl: $\mathcal{C}_1 \not\Leftrightarrow \mathcal{C}_2$ holds in $\mathcal{A}$ iff $\{\mathcal{C}_1, \mathcal{C}_2\} \cap \mathcal{A} \neq \{\mathcal{C}_1, \mathcal{C}_2\}$
- Uniq: $Uniq(\mathcal{C}_1)$ holds in $\mathcal{A}$ iff $card(\mathcal{C}_1, \mathcal{A}) \leq 1$
- Oblig: $Oblig(\mathcal{C}_1)$ holds in $\mathcal{A}$ iff $card(\mathcal{C}_1, \mathcal{A}) = 1$
- Dep: $\mathcal{C}_1 \rightsquigarrow \mathcal{C}_2$ holds in $\mathcal{A}$ iff $comp(\mathcal{C}_1, \mathcal{C}_2)$ holds

A grammar, in this perspective, is a set of constraints and nothing else. In particular, there is neither ranking in the OT sense nor need of building any structure before being able to evaluate the properties (as in OT with the Gen function or in HPSG with the need of selecting first a hierarchized structure type). Parsing in property grammars consists in evaluating for a given input the entire set of constraints. The characterization of an input is then formed by the set of satisfied constraints and the set of violated ones. More precisely, the grammar contains for each category a subset of constraints. It is then possible to build a characterisation (the set of evaluated constraints) for each category as well as for the entire input.That is, after checking the input against the grammar for the properties relative to, say the NP, we end up with the set of properties satisfied by the input on one hand, and the set of violated properties on the other hand. These two sets characterise the input; they constitute its characterisation, relative to the NP.

3.2 Property Grammars and the Constraint Solving Problem

Unlike a more standard way of seeing a constraint solving problem, which would consist in finding an assignment that satisfies the system of constraints, we start here with a partial assignment—corresponding to the utterance to be parsed—and we want to complete this assignment so as to satisfy the system of constraints—i.e. the grammar. The assignment is partial in the sense that it does not concern all the variables involved in the constraint system as it would be the case in a traditional constraint solving problem, but it is only interested in the lexical entries from the utterance to be analysed. Each of the words from the utterance corresponds to a variable, which is associated with a domain of values made of all the possible lexical categories for this word. More variables are created by propagation along the process, typically when a construction from a higher level is identified, while other variables may be suppressed by simplification. In terms of output of the solving process we are not only interested in the qualitative aspect of the final assignment but equally in knowing which constraints are satisfied and which ones are not, both from a qualitative and a quantitative point of view. The knowledge of the assignment by itself is not enough. Actually, what we want to know is, given a system of constraints and a first partial assignment—representing the input to be analysed— a description of the final system and the complete assignment used to reach it. In a linguistic perspective it means that we want to know the sets of properties

---

[6] The semantic completion follows some schemas such as *subject, complement* or *modifier*. These schema indicate what part of the semantic structure of the modified category must be completed with that of the dependent.

which are satisfied and violated by the input utterance—i.e. the characterisation—when the input is analysed in such way—i.e. for a given assignment. For instance, for the following utterance in French:

la ferme

one possible initial assignment is [la–*Det*, ferme–*N*], which leads to the characterisation of an NP with no violated property. In the final assignment the pair [la, ferme] will be labelled NP. In this case, the utterance means *the farm*.

Another possible initial assignment is [la–*Pro*, ferme–*V*], which leads to the characterisation of an NP for [la–*Pro*] with no violated property, then a VP for the assignment [la–*NP*, ferme–*V*] with no violated property either. In the final assignment the pair [la,ferme] is labelled VP. In this case the utterance means *shut up*.

Of course, as per usual with constraint-based approaches such an elegant statement of the problem comes with a computational cost. We mostly rely on traditional techniques such as propagation and simplification—and also simpagation, in the case of an implementation with Constraint Handling Rules (see [20] Frühwirth, 1995)—to tackle the question. Although addressing complexity issues related to constraint solving is beyond the scope of this paper, it is worth mentioning that density could probably be used to reduce the search space, in particular when faced with ill-formed input. Indeed in such a case relying on a measure of gradient acceptability and thus on density to rank the possible candidate assignments seems a reasonable heuristic. It leads us to predict that density is likely to play an important practical role when it comes to implement constraint-based parsing.
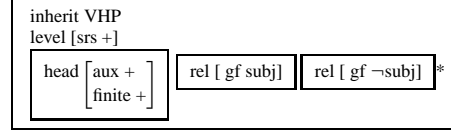
3.3 Constructions and Constraints

A fully constraint-based view such as the one proposed by Property Grammars makes it possible to implement contextual fine-grained information. In the same way as categories are described by means of constraint subsets, other kind of objects such as constructions can also be specified in the same way. The notion of construction is of deep importance, especially in the perspective of finding gradient judgements: the importance of a constraint can vary from one construction to another. In the remainder of this section, we describe how constructions can be represented by means of constraints in the PG framework, taking the example of the Subject-Auxiliary Inversion (SAI) linguistic form in English (see [9] Fillmore, 1998).

In this linguistic form, there are strong linear constraints, especially concerning adverbs as shown in the following examples:

(4) Did you learn your lesson?
(5) Did you really learn your lesson?
(6) Didn't you learn your lesson?
(7) *Did really you learn your lesson?
(8) *Did not you learn your lesson?
(9) *Did you your lesson learn?

This linguistic form is explained by the following SAI construction:

```
┌─────────────────────────────────────────────────────────────┐
│ inherit VHP                                                  │
│ level [srs +]                                                │
│  ┌──────────────┐  ┌──────────────┐ ┌──────────────┐        │
│  │ head ⎡aux +⎤ │  │ rel [ gf subj] │ │ rel [ gf ¬subj] │ * │
│  │      ⎣finite +⎦│  └──────────────┘ └──────────────┘      │
│  └──────────────┘                                            │
└─────────────────────────────────────────────────────────────┘
```

In terms of PG, this linguistic form can be detailed as a set of such constraints:

(C1) $V[aux] \prec NP[subj]$
(C2) $NP[subj] \prec V[\neg fin]$
(C3) $V[aux] \Rightarrow NP[subj]$
(C4) $V[\neg fin] \prec XP[\neg sub]$
(C5) $NP[subj] \prec Adv[neg, \neg contraction]$
(C6) $NP[subj] \prec Adv[\neg neg]$
(C7) $NP \rightsquigarrow V$
(C8) $Adv \rightsquigarrow V$

This subset of constraints {C1,C2,C4,C7,C8} represents the information of the SAI construction as represented in the box above and we can say that these two notations are equivalent. In this example, on top of the general constraints describing the SAI construction it is necessary to specify some further constraints in order to describe the top-level linguistic form we are interested in—see (4) to (9). Adding new information simply consists in adding new constraints to the set describing the basic construction, at the same level. For example, the negative form has to be contracted here. This constraint is imperative in this linguistic form, whereas it can be optional in other cases. This constitutes one important interest in the perspective of associating constraints with an "importance" degree: such degree may vary according to the construction. Using the soft vs. hard dichotomy between constraints whether its violation leads to respectively mild or serious ungrammaticality, as described in [6] Sorace, 2004 (paragraph 2.1), a constraint can be hard in some constructions or soft in some others.

Note that it is important, in order to discriminate the constructions, for a non-empty subset of constraints from the constraint system specifying the form (i.e., the construction) to be unique. Satisfying this subset is a necessary, but not sufficient, condition for identifying the construction.

4   Gradience and Density

A fully constraint-based representation may also be helpful in identifying criteria for sentence complexity as well as for acceptability. The idea is to make use of the information contained in characterisations in terms of satisfied and violated constraints. More precisely, some figures can be extracted from these characterisations illustrating the difference in the realization of a given category. For example, the ratio of satisfied/violated constraints is obviously of main interest.

(10) Quelles histoires Paul a-t-il écrites ?
What stories Paul did he write[fem-plu]? / What stories did Paul write?
(11) Quelles histoires Paul a-t-il écrit ?
What stories Paul did he write[masc-sing]? / What stories did Paul write?
(12) Quelles histoires a-t-il écrites Paul ?
What stories did he write[fem-plu] Paul? / What stories did he write Paul?
(13) Quelles histoires a-t-il Paul écrites ?
What stories did he Paul write[fem-plu]? / What stories did he Paul write?

These examples are given in order of (un)acceptability which corresponds in our hypothesis to a progressively greater number of violated constraints. Constraints are given here without taking into account specificities of the interrogative construction.

(C9)   NP[obj] $\rightsquigarrow$ VP[ppas]
(C10) NP[subj] $\prec$ VP
(C11) NP[subj] $\prec$ VP, V $\not\Leftrightarrow$ NP[subj]

Even without a precise evaluation of the consequence of constraint violations type by type, this first criterion can constitute an objective element of estimation for acceptability: unacceptability increases with the number of constraint violations (This observation corresponds to Keller's property of *Cumulativity*). This indication seems trivial, but directly comes from the possibility of representing separately the different types of syntactic information by means of properties. Such estimation is for example not possible with a phrase-structure representation and even difficult using classical constraint-based approaches such as HPSG.

However, it is necessary to have a finer-grained use of such information. In particular, the number of constraints may vary from one category to another. Some categories, such as Adverbial Phrases, are very static and are described with a limited number of properties. At the other end of the spectrum, the Noun Phrase, which can have many different forms, needs an important number of properties. It is then necessary to distinguish the number of constraint violation in these cases: violating a constraint for an AdvP entails more consequences regarding acceptability than for the NP. Again, this indication is purely quantitative and does not take into account constraint type. It is probably the case that some constraints (for example exclusion) play a more important role with respect to acceptability than dependency for example. However, when taking into consideration interpretability for example, a hard constraint such as unicity with respect to acceptability becomes soft for the interpretation, as shown in the following examples:

(14) Paul reads a book
(15) Paul reads reads a book

The second example is obviously unacceptable but perfectly understandable.
We propose then a first stage in the identification of gradient criterion by means of purely quantitative aspects. This is the role played by the notion of *density*. This information indicates two figures: the number of satisfied properties with respect to the total number of properties that described the object and the same ratio for violated properties.

We note respectively these figures as `dens_sat` and `dens_unsat` with the following definitions:

- `dens_sat` = num. of satisfied properties / total num. of properties
- `dens_unsat` = num. of unsatisfied properties / total num. of properties

Density in itself, unlike the ratio satisfied/violated properties, gives some indication about the quantity of information of a given object. In the case of a high density of satisfied properties, this means that an important number of syntactic characteristics contained in the grammar is realized in the object. In other words, we can say that this object contains, with respect to the grammar, important syntactic information. Reciprocally, a low density of satisfied properties can have different interpretations. In case of a high density of violated constraints, the object is clearly ill-formed and we can suspect a low probability for its acceptability. But it can also be the case that there is a low density for violated constraints. This situation indicated that the object contains little syntactic information. In the example (16), extracted from a corpus analysis, the category, a sentence, is formed with a PP and a VP (the values for densities are reported in table 1).

(16) En renforçant le projet, avançons vers le succès.
in reinforcing the project, let's go toward the success

Cat	dens_sat	dens_unsat	construction
S	0,5	0,125	PP; VP

Table 1. Density values for a PP–VP construction.

Such a construction is not frequent, but some information can be given, according to the grammar. Concerning the violated properties, the non-null density comes from the fact that there is a dependency relation between the VP and an NP subject which is not realized. The level of satisfied properties density comes from the fact that even if the properties involving PP and VP are satisfied, many properties describing S involve an NP. There is then a high number of properties for S that cannot be evaluated, explaining a low `dens_sat`.

These different ratios constitute then a first tool providing some indication about acceptability and interpretability. Acceptability primarily depends on the ratio of satisfied constraints to the number of violated ones. Interpretability can be illustrated by the densities of satisfied and violated constraints. Low densities, as shown above, indicate a low level of syntactic information. More generally, there is a correlation between the quantity of information contained by an object and its interpretability: a high density of satisfied constraints comes with an easier interpretation. In case of low densities, it is necessary to obtain information from other domains such as prosody.

Using these different indications makes it possible to give information about any kind of input, without any restriction to well-formed ones. Moreover, it becomes possible to propose quantitative elements towards gradience in linguistic data concerning

both acceptability and interpretability. Moreover, such elements of information give also some indication about domain interaction. For some utterances, it is necessary to extract information from the different linguistic domains such as morphology, syntax, prosody or pragmatics. In some other cases, the morpho-syntactic level alone contains enough information in order to make an utterance interpretable. In other words, there is a balance among the different domains. Each domain can be characterized with densities such as the one described here for syntax, the balance status being a function of the different densities. A high density of satisfied properties for one domain is an indication of a high level of information. The hypothesis stipulates that in this case, other domains can contain a low level of information without consequence on the interpretability. For example, for some construction, if there is a high density in syntax and semantics, then the density of prosody is not constrained and can take any value. Concretely, this means that intonation is not constrained anymore and can be realized in various ways. On the contrary, when syntactic and semantic densities are not heavy enough, then the prosody density has to be high and the intonation is less variable. This is the case in the example (3) for which prosody plays an important role in the interpretation.

5 Experiment

In the following, we give some indications from different French corpora, calculated from the output of a deterministic property grammar parser (see [21] Blache, 2001, for a description of this parser). One important restriction is that, insofar as the parser used is deterministic, the number of violated constraints has been restricted to a minimal level. In particular, only the linearity, exclusion and unicity constraints have to be satisfied. The density of violated constraints is therefore not relevant for our discussion. We take then only into account the density of satisfied constraints. The aim of this experiment is to extract some figures from different data, for a given grammar and a given parser. It cannot be considered as a parser (or grammar) evaluation.

Three different corpora have been used: the first from the newspaper "Le Monde", with 15,420 words, the two others are transcribed spoken language corpora containing respectively 523 and 1,923 words. These corpora are very small, which is justified by the difficulty in parsing such data. Moreover, they have been filtered: incomplete words for example have been eliminated. However, all repetitions are kept.

The first observation in this experiment is that, even if most of the categories have a null density (for the reasons explained above), there is a huge difference among the densities of satisfied constraints. The table 2 indicates for example some figures concerning the noun phrase in the written text corpus.

In these figures, one can remark that density doesn't grow systematically with grammaticality. For example, the two lowest densities correspond to grammatical constructions (personal pronoun and clitic). This comes from the fact that the noun phrase, which is the most frequent category, has many different constructions and needs a lot of constraints to describe them. In all cases, even when a given construction satisfied all its corresponding constraints, insofar as the total number of constraints for the NP is high, the density is necessarily low. Moreover, the realizations observed here only contain one category. The total number of satisfied properties is then by definition very

Density	Const	Density	Const
0.034483	PPro	0.310345	Det AP PP
0.068966	Clit	0.379310	Det N Rel
0.103448	N	0.413793	Det AP N
0.1724138	ProP AP	0.413793	Det N PP
0.206897	Det AP	0.517241	Det N Rel PP
0.241379	Det PP Rel	0.551724	Det N AP PP
0.275862	Det N	0.655172	Det N PP AP Rel

Table 2. Densities of Noun Phrases

low without having any consequence on the grammaticality (which should be indicated by the ratio satisfied/violated constraints). The same explanation is valid when comparing the realization /Det N/ with /Det N PP/. The first has a lower density whereas one should expect a high one for this basic construction. Frequency information plays then an important role in the use of the density notion. The table 3 indicates the mean density with respect to the frequency of the category in the different corpora. We can

Cat	Frequency	Density
S	0.069917582	0.4733535
AP	0.108379121	0.408556
AdvP	0.048139361	1
NP	0.302047952	0.204571
PP	0.1003996	0.31331
VP	0.218981019	0.341995
Circ	0.064360639	0.718518
Coord	0.071978022	0.4821425
Rel	0.015796703	0.3543475

Table 3. Mean density with respect to frequency

see in this table that the most frequent categories (NP, VP and PP) are also those with the lowest mean density whereas the less frequent ones (Circ, AdvP and Coord) are associated with high densities. The arguments given above concerning the number of constituents, the number of properties and the number of different constructions can be used to explain these differences.

This density parameter has then to be modulated with the frequency of the construction. In all cases, the progression of the density comes with an increasing quantity of information. It is important to notice that the density notion is not directly useful in the identification of sentence complexity. For example, one can consider that a realization of an NP with a relative clause is more complex than a construction /Det N/. However, the first has a higher density than the second, for the reasons explained above. But

from the interpretability point view, these aspects are disconnected. For example, a cleft construction, which is identified as being complex, is easily understandable because of the high number of constraints describing it. The examples in table 4 illustrate some density differences from a construction with few constraints to be satisfied and another containing more information.

	Example	*Density*
(16)	celui rouge *that red*	0,1724138
(17)	Le contenu de la future con- vention qui devrait permettre de régler les problèmes de fond *the content of the future conven-* *tion that may allow to solve the* *fundamental problems*	0,6551724

Table 4. Density when few vs. many constraints

6 Further Work

Intuitively, the notion of density could be refined by weighting the constraints according to their importance. The hard/soft discrimination ([6] Sorace, 2004), for instance, is not accounted for at the moment by the density whereas we have seen previously that the constraints play roles of different importance when it comes to acceptability. Some sort of constraint ranking would also let us model the cumulativity and ganging up effects (i.e. when multiple violations of soft constraints could possibly be more unacceptable than a single violation of a hard constraint) described in [5] Keller, 2000,and [6] Sorace, 2004.
Another object of further investigation concerns the use of weighted densities during the parsing process as an element of disambiguation. Indeed when faced with different possible assignments heuristics could be based on the measure of density for each possibility in order to rank the structures by preference. Subsequently a deterministic approach, of course, could also use this preferred structure to reduce the search space at different stages in the solving process.

7 Conclusion

We introduced in this paper the notion of density of information in the context of a constructional constraint-based framework for language processing. The purpose of density is to quantify the amount of linguistic information available within an utterance. It is meant to be used for various applications. One possible application consists in using density measure during the constraint parsing process in order to reduce the search

space. Another class of applications is related to using density measures as part of Information Retrieval or Question/Answering systems to locate the parts of a text which are the most likely to carry information.

Our intuition is that our approach to modeling language by means of constraints only is a privileged framework for such a measurement. The reason for that is that the underlying representation by constraints lets us manipulate all the linguistic information available for an utterance at the same level regardless of any sort of candidate structure which could be associated to the utterance and also across different domains of language, such as syntax, prosody and pragmatics.

Constraints play the role of a filtering process: linguistic information does not consist in defining all and only the possible constructions, but does consist in indicating for some construction what kind of information can be extracted. This means that in cases where little (or no) information can be extracted from one domain (e.g. from syntax, prosody or pragmatics), we can still rely upon other domains to deal with the utterance.

We believe that within an appropriate theoretical framework for modeling language by means of constraints it is possible for fuzzy notions such as syntactic complexity, or acceptability to be evaluated and quantified by means of a numeric function of the proportion of satisfied and violated linguistic properties. The notion of density by itself still presents weaknesses and the method for its computation should be refined. It would certainly benefit, for instance, from taking into account the freqency of a construction. In spite of these problems, introducing such a notion of density of information is a first attempt to go in that direction.

Acknowledgements

References

1. Gibson, E.: The Dependency Locality Theory: A Distance-Based Theory of Linguistic Complexity. In Marantz, A., Miyashita, Y., ONeil, W., eds.: Image, Language, Brain. Cambridge, Mass., MIT Press (2000) 95–126
2. Bolinger, D.: Generality, Gradience and the All–or–None. The Hague: Mouton (1961)
3. Aarts, B.: Modelling Linguistic Gradience. Studies in Language **28** (2004) 1–49
4. Bresnan, J., Nikitina, T.: On the Gradience of the Dative Alternation. Draft (2003)
5. Keller, F.: Gradience in Grammar - Experimental and Computational Aspects of Degrees of Grammaticality. PhD thesis, University of Edinburgh (2000)
6. Sorace, A., Keller, F.: Gradience in Linguistic Data. Lingua (2004) [in press].
7. Prince, A., Smolensky, P.: Optimality Theory: Constraint Interaction in Generatire Grammar. Technical report, TR-2, Rutgers University Cognitive Science Center, New Brunswick, NJ. (1993)
8. Pullum, G., Scholz, B.: On the Distinction Between Model-Theoretic and Generative-Enumerative Syntactic Frameworks. In de Groote, P., Morrill, G., Rétoré, C., eds.: Logical Aspects of Computational Linguistics: 4th International Conference. Number 2099 in Lecture Notes in Artificial Intelligence, Berlin, Springer Verlag (2001) 17–43

9. Fillmore, C.: Inversion and Constructional Inheritance. In: Lexical and Constructional Aspects of Linguistic Explanation, Stanford University (1998)

10. Goldberg, A.: Constructions: A Construction Grammar Approach to Argument Structure. Chicago University Press (1995)

11. Keller, F.: A Probabilistic Parser as a Model of Global Processing Difficulty. In: Proceedings of ACCSS-03. (2003)

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

13. Blache, P.: Constraints, Linguistic Theories and Natural Language Processing. In Christodoulakis, D., ed.: Natural Language Processing. Volume 1835 of Lecture Notes in Artificial Intelligence. Springer-Verlag (2000)

14. Mertens, P.: Accentuation, intonation et morphosyntaxe. Travaux de Linguistique **26** (1993)

15. Sag, I., Wasow, T.: Syntactic Theory. A Formal Introduction. CSLI (1999)

16. Maruyama, H.: Structural Disambiguation with Constraint Propagation. In: Proceedings 28th Annual Meeting of the ACL, Pittburgh, PA (1990) pp. 31–38

17. Duchier, D.: Axiomatizing Dependency Parsing Using Set Parsing. In: Proceedings 6th Meeting on the Mathematics of Language, Orlando, FL (1999)

18. Heinecke, J., Kunze, J., Menzel, W., Shröder, I.: Eliminative Parsing with Graded Constraints. In: Proceedings 7th International Conferenceon Computational Linguistics, 36th Annual Meeting of the ACL. Volume Coling–ACL '98., Montreal, Canada (1998) pp. 526–530

19. Christiansen, H.: CHR Grammars. Theory and Practice of Logic Programming **Special issue on Constraint Handling Rules** (2005) [to appear].

20. Frühwirth, T.: Constraint Handling Rules. In Podelski, A., ed.: Constraint Programming: Basics and Trends. Springer LNCS 910 (1995)

21. Blache, P.: Les Grammaires de Propriétés : des contraintes pour le traitement automatique des langues naturelles. Hermès Sciences (2001)