Light-and-Deep Parsing: a Cognitive Model of Sentence Processing

Philippe Blache
CNRS & Aix-Marseille Université
Laboratoire Parole et Langage
blache@blri.fr

December 23, 2015

1 Introduction

Understanding how human process language in a very fast and efficient manner remains largely unexplained. The main difficulty is that, in spite of the fact that many disciplines has addressed this question (linguistics, psychology, computer science, and neuroscience), it remains difficult at this stage to describe language as a global system. For example, no linguistic theory entirely answers to the question of explaining how the different sources of linguistic information interact. Most theories, and then most descriptions, only capture partial phenomena, without being capable of providing a general framework bringing together prosody, pragmatics, syntax, semantics, etc. This is the reason why until now, most linguistic theories consider language organization from a modular perspective: linguistic domains are studied and processed separately, their interaction is implemented in a second stage. As a consequence, the lack of a general theory of language, accounting for its different aspects, renders difficult the elaboration of a global processing architecture. This problem has direct consequences in natural language processing: the classical architecture relies on different subtasks: segmenting, labeling, identifying the structures, interpreting, etc. This organization more or less strictly imposes a sequential view of language processing, considering words as being the core of the system. This view does not account for the fact that language is based on complex objects, made of different and heterogenous sources of information, interconnected at different levels, and which interpretation cannot always be done compositionally, each information domain transferring a subset of information to another, but on the contrary taking into account at the same time all sources of information and their interaction.

Cognitive approaches to language processing (hereafter LP) are faced with the same difficulties. Even more crucially than in linguistics, psycholinguistics models mainly rely on a sequential and modular organization. Language is usually considered to be strictly incremental, relying on a word-by-word processing, each word being integrated into a partial syntactic structure, starting from which an interpretation can be calculated. In this organization, the same different steps as in classical NLP architectures is used: segmentation, lexical access, categorization, parsing, interpretation, etc.

This perspective is also adopted in neurolinguistics, trying to identify in a spatial or in a temporal dimension the brain basis of LP. The question there consists in studying what part of the brain are involved in LP and in what manner. What is interesting is that even though the different works focus on only one linguistic dimension (lexicon, lexical semantics, prosody, morphosyntax etc.), they also show that they are strongly dependent from each other.

Bringing closer the different types of knowledge about language processing coming from these disciplines makes it possible to draw a broader and more integrated architecture.

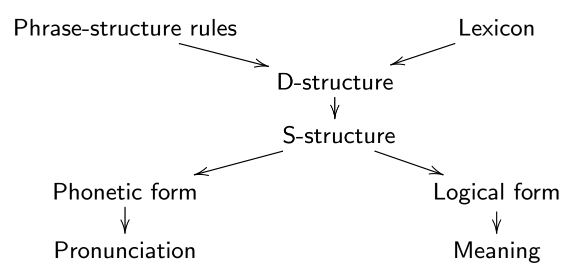
2 An interdisciplinary view of language processing

We propose in this section an overview of language processing through different disciplines (linguistics, psycholinguistics, computational linguistics and neurolinguistics) through several perspectives. We show in particular that classical architectures (modular and serial) are now challenged, in particular when taking into account language as a whole, in its natural environment. The main features of a more flexible approach are presented.

2.1 Classically, LP is modular and incremental

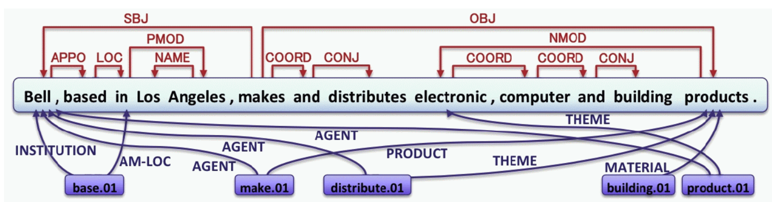
This section present the main features of the classical modular architecture of LP, coming from the generative framework, and that has influenced the other disciplines.

Modularity (a view from linguistics): The classical generative architecture relies on a succession of different modules, each one being specialized for a specific linguistic dimension. The following figure illustrates such organization for "*Government and Binding*", one of the generative theories presented in [Chomsky, 1981]. Starting from the lexicon and the rules (the set of local trees), the process consists in generating an underlying structure, subject to modifications and transformations that lead to another structure, closer to the surface form. From this structure, the phonological and a logical forms are produced, making it possible to access to the meaning:



This organization not only consider the different modules as being separated, but also sequential. Many linguistic theories propose a similar organization in which each domain produces a structure that is taken as the input of another one. One of the reasons comes from the fact that linguistic theories are usually *syntacto-centric*: all domains are considered as adding information to the syntactic structure.

Module interaction (a view from computational linguistics): LP is taken from a specific perspective in NLP because of the constraints of implementation: LP is usually considered as a set of tasks, implementing the different modules in a serial manner. In this architecture, modules are synchronized, the input of one module being the output of the previous one. Up to now, no real answer to the question of the integration of the different sources of linguistic knowledge is given and their interaction is described in terms of specific synchronization rules [Jackendoff, 2007]. More precisely, even though many works have been done concerning the study of the interaction between the domains (e.g. prosody/syntax, syntax/semantics, etc.), the solutions are proposed by giving the priority to one domain, usually syntax. For example, the compositional view of semantics [Werning et al., 2012] is implemented by the construction of a syntactic structure starting from which the interpretation can be calculated [Copestake et al., 2001]. The same kind of approach can be found for the syntax/prosody interface, in which prosodic information is integrated to the syntactic structure [Steedman, 2000]. As presented in the previous section, this is a syntacto-centric organization, that induces an incremental and modular view of LP:. As a consequence, classical NLP architectures are organized around a series of sub-processes: tokenization, tagging, parsing, discourse organization, semantic interpretation. A parser builds complex structure (trees, feature-value structures, etc.) involving information at different levels, as shown in the following example, based on information produced by parsers such as the Stanford [de Marneffe and Manning, 2008], enriched with semantic information:



Other processes can be added to this general schema when studying audio (phonetics, prosody) or multimodality (gestures). Even though the question of parallelization in NLP is regularly addressed [Adriaens and Hahn, 1994, Jindal et al., 2013], the answer is usually given in terms of different parallel processes with meeting points, instead of an integrated view.

Incrementality (a view from computational psycholinguistics): In psycholinguistics, the LP classical architecture relies on the idea that processing is incremental, consisting in integrating each new word into a partial structure under construction [Fodor and Ferreira, 1998, Grodner and Gibson, 2005, Sturt and Lombardo, 2005, Keller, 2010, Altmann and Mirković, 2009, **?**]. In this approach, the basic units are considered to be the words: all information related to the lexical item are accessed when encountering a new word, this information being used in order to integrate this item into a partial syntactic structure (also called the current partial phrase marker, or CPPM). This operation consists in finding a site in the structure where to attach the word. In case of difficulty, the word is integrated where it last severely violates the grammar, following the "*Attach Anyway*" principle proposed in [Fodor and Inoue, 1998].

This conception is, as presented in the previous section for classical grammatical generative grammars, syntacto-centric, consisting in organizing all information with respect to the syntactic structure. Moreover, it is essentially sequential, in the sense that lexical information is processed before syntax, starting from which interpretation becomes possible. This is a modular syntax-first conception, supported by several classical works [Fodor, 1983, Frazier and Fodor, 1978] and still at work in many recent psycholinguistics models.

In terms of interpretation or meaning access, these approaches are also basically compositional. Moreover, two different approaches have been proposed, explaining how interpretation can be built [Gibson, 2000]. In serial models, language processor initially computes only one of the possible interpretations [Fodor and Ferreira, 1998, Gorrell, 1995]. In case of this interpretation becomes difficult or impossible, another interpretation is built. For parallel models on the opposite, all possible interpretations are computed at once, the analysis with the greatest support being chosen over its competitors [MacDonald et al., 1994, Marslen-Wilson and Tyler, 1980, Spivey and Tanenhaus, 1998]. These two options both rely on an incremental view: interpretation is built at each new word, on the basis of a syntactic and semantic analysis. Many issues are raised with both of these modes. First, they consider incrementality in a strict manner: an interpretation covering all the words at a given position is to be built, even by building an ill-formed structure [Fodor and Ferreira, 1998]. Moreover, the question of memory remains also an issue in both cases: what elements are to be stored, under what form, requiring what capacity?

Brain basis of a modular architecture (a view from neurolinguistics): The study of the physiological and brain basis of language processing also leads to contributions about LP architecture in many different respects. Among the possible investigation techniques for the exploration of neural correlates of LP, electrophysiological studies are very frequently used. These experiments focuses on the study of ERP (*event-related potentials*) that are potential changes measurable from the scalp and that reflect the activity of a specific neural process [Luck, 2005]. LP modulates a number of ERP components, located between 100 and 600 ms after the stimulus (for example reading a word).

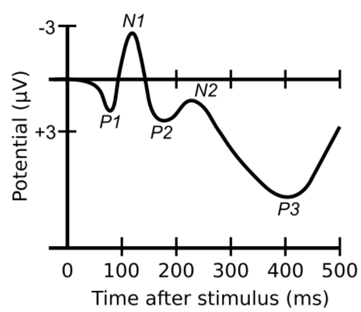


Figure 1: The main ERP components in language processing

Many effects have been explored, related to different linguistic domains (prosody, morphology, syntax, semantics in particular) [Kutas et al., 2006, Kaan, 2007]. Even though no electric component is strictly related with one domain, we can very roughly find in the literature some early effects reported for speech perception at 100ms after the stimulus, word production around 200ms, semantics around 400ms and syntax around 600 ms. Again, this in only a very rough picture, and all the observed effect depends on the linguistic material, in particular the amount of information coming from each domain.

Several works in neurolinguistics support a modular and serial view of language processing. In particular, the 3-phases model [Friederici, 2002, Friederici, 2011] proposes an organization into three different steps, after an initial phase of acoustic-phonological analysis:

1. Local phrase structure is built on the basis of word category information
2. Syntactic and semantic relations (verb/argument, thematic role assignment)
3. Integration of the different information types and interpretation

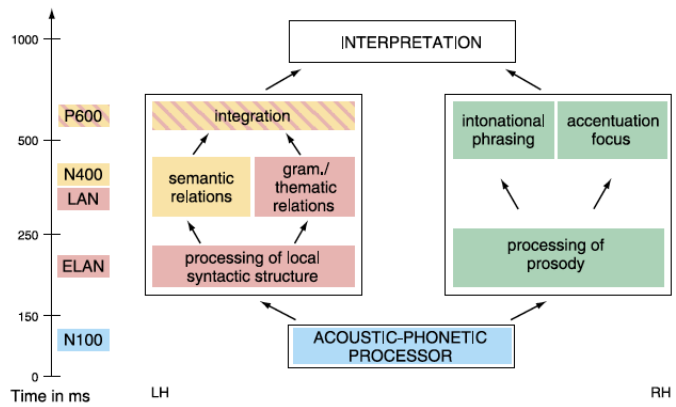


Figure 2: General organization of the 3-phases model [Friederici, 2011]

This organization can be completed, in an auditory comprehension model, by adding interaction of prosody at each of these stages. Different language ERP components are in relation with these phases, as shown in figure 2:

- Early left anterior negativity (ELAN, between 120-200 ms): initial syntactic structure building processes
- Centroparietal negativity (N400, 300-500ms): semantic processes
- Left anterior negativity (LAN, 300-500ms): grammatical relation between arguments and verb, assignment of thematic relations
- Late centroparietal positivity (P600): late syntactic processes

This model is syntax-first, and consists in building a syntactic structure starting from which interpretation can be done.

2.2 However, LP is more integrated

The classical view is now challenged by recent theories considering that no domain is at the center of the architecture, the processing being described in terms of interaction between them. Instead of begin serial, processes are considered as parallel, as described in [Jackendoff, 2007].

Construction Grammar [Fillmore, 1988]; [Goldberg, 1995] is one of those theories proposing an alternative organization. Here, no structure is predefined, and no domain need to be described before another: all linguistic phenomena are described thanks to a set of interacting properties. As presented in [Goldberg, 2003] constructions are form and meaning pairings: a set of properties makes it possible to characterize a construction, which meaning is accessed directly. Constructions can be of very different types, as presented in the following examples:

1. Ditransitive Construction

   `Subj V Obj1 Obj2`: "*She gave him a kiss*"

2. Covariational Conditional Construction

   `The Xer the Yer`: "*The more I read the less I understand*"

3. Idiomatic Constructions “*kick the bucket*”; “*to put all eggs in one basket*”

What is important with constructions is the fact that they are defined on the basis of different properties, coming from different linguistic domains, without any need of a precise analysis of each of these domains. For example, in the conditional or idiomatic constructions (see above), a syntactic tree (provided that it is possible to build one) is of no help to access to the meaning. In such cases, meaning cannot be built compositionally.

As a consequence, in the perspective of language processing, this means that two types of mechanisms cohabit: one based on a compositional architecture, and a second relying on direct access. In the first case, the architecture consists in analyzing all sources of information and their interactions. Each source or combination of sources contain a partial meaning, their composition leading to a complete interpretation of the message. In the direct access case, the different properties, more than bearing part of the meaning, are used as cues for the identification of a construction. The recognition of the pattern leads to a direct interpretation, without any composition. In some cases, only a few properties makes it possible to recognize a construction and then to interpret an entire message.

2.3 Moreover, LP is often shallow

A flexible model of LP, called *Good-Enough Theory* [Ferreira and Patson, 2007], has been proposed recently. It is based on the observation that representations formed from complex or difficult material are often shallow and incomplete. For example, [Swets et al., 2008] shows in a self-paced reading study that when participants expected superficial comprehension questions, ambiguous sentences are read faster, showing that no precise attachment resolution is done, leading to underspecified semantic representations. In this case, it is suggested that the ambiguity is not resolved at all, explaining the facilitation effect.

Several experiments confirm this observation that comprehension of sentences can be quite shallow. For example, thematic role assignment is subject to a simple heuristics: the first NP is the agent, the second the entity affected by the action. The use of such heuristic has been exhibited by simple experiments; the interpretation of sentences contradicting this heuristics lead more often to misinterpretations. These observations tends to show that in several cases, no compositional processing is at work. Instead, as explained in [Ferreira and Patson, 2007], “*the comprehension system tries to construct interpretations over small numbers of adjacent words whenever possible and can be lazy about computing a more global structure and meaning*”. The building of a complete and precise interpretation is then often delayed or even never done, leaving place to the identification of “*islands*”, starting from which a general interpretation can be approximated.

Note that this theory contradicts several classical language processing models. In this case, there is no systematic instantiation of thematic roles, at least in a first stage, contrarily to what is required in generative theories. This is also contradictory with some psycholinguistic difficulty models [Gibson, 2000] stipulating that NP without thematic roles as well as unassigned thematic role imposes a burden on working memory. On the opposite, good-enough theory proposes that shallow semantic processing, even involving such underspecification, can be an element of facilitation.

Several works in neurolinguistics focusing on semantic processing also suggest that some type of basic and shallow processing can be at play. In line with the *GE* theory, meaning integration can be switched of when the context renders it unnecessary. This is the case

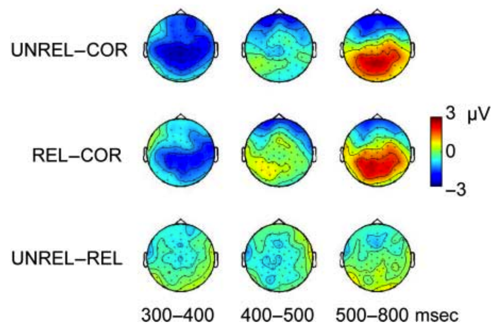


Figure 3: Illustration of the differences between the conditions in idiomatic context (COR=correct sentence, REL=soft violation, UNREL=hard violation). This schema shows a difference in the potentials when comparing the correct and both types of violation, but no between hard and soft violations (UNREL-REL).

when processing idioms, as shown by [Rommers et al., 2013]. In this case, semantic composition might be not fully engaged during comprehension. In particular, the activation of literal word meanings are only carried out when they are really necessary. When analyzing the cortical responses of semantic violations, this study observed no significant difference at N400 (the component related to semantic surprisal) between hard and soft violations for idiomatic context, where an important reduction in N400 amplitude appears for soft violation in comparison to hard one for literal contexts.

2.4 Finally, LP is not always based on words

One of the first question to address is that of the types of units that are recognized an processed during language processing. The classical option considers that LP is strictly incremental and compositional, consisting in recognizing atomic elements and aggregating them progressively (phonemes, morphemes, words, phrases, etc.) in order to build an interpretation made of the composition of the semantic information available at each level (in particular words and phrases). Another option considers that the input stream of data (being them heard or read) is stored in the working memory on the basis of larger units, identified thanks to a basic superficial processing, without actual analysis. The actual treatment leading to interpretation is made on the basis of such chunks. Several experimental observations support this idea of a more global non strictly incremental processing.

Many NLP applications only require a low level of precision, based on superficial information concerning the basic relations between word (cooccurrency, order). The technique is called shallow parsing [Uszkoreit, 2002, Balfourier et al., 2002, Baldwin et al., 2002] and rely on the identification of basic structures called *chunks* [Abney, 1991] that consists in groups of adjacent words, usually identified thanks to their boundaries markers instead of the syntactic relations between their constituents :

```
[When I read] [a sentence], [I read it] [a chunk] [at a time]
```

We find in the literature several works showing that this notion of chunk can be considered as a relevant unit when studying human language processing. For example, studying eye-movement when reading a text shows that fixation are only done from time to time. This is very well-known and comes from the capacity of a parafoveal vision that consists in a preview of adjacent words. As presented in the following figure [Schuett et al., 2008], during a fixation, readers extract visual information from the foveal visual field (central white oval) and the parafoveal visual field (grey ellipse).

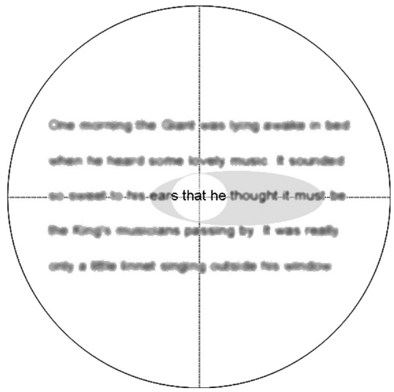


Figure 4: Parafoveal vision: extracting features from the surrounding words

Such capacity makes it possible to extract information about upcoming word, opening the capacity to deal with entire sequences, not only separate words. Moreover, [Rauzy and Blache, 2012] has shown that fixations can be done by chunks (defines here by a sequence of function word and content word).

This observation is an argument in favor of a more global treatment, including at the physiological level. It can be supported by other observations focusing on the neural correlates of LP: several works have specifically study the question of syntax and more precisely its role in the processing of basic properties. In particular, [Pulvermuller et al., 2008, Pulvermüller, 2010] have observed that some morpho-syntactic properties can be assessed automatically, at a low level, when studying differences between chunks with or without a *Det-N* or *Pro-V* agreement violations. When comparing the two conditions, one observe a difference in the cortical reaction at a very early stage (around 100ms after the stimulus, see figure 5). This effect, called *Mismatch Negativity* (MMN), occurs in a range of time and an experimental design in which there is no strictly conscious activity. This research suggests that syntax can function as a discrete combinatorial system, that can be implemented by discrete combinatorial neuronal assemblies in the brain [Pulvermüller, 2010], that can connect categories into larger constructions.

2.5 Intermediate Summary

Many of the observations presented above tend to indicate that there is not a simple, homogeneous and unique language processing architecture. The easiest and more natural way to explain LP is to distinguish different modules that are organized in a serial schema. Supporting this idea, many experiments (judgments, eye-movements, brain activity) have shown the respective effects and contribution to these modules in the global processing. They usually rely on a theoretical conception of language in which each module bears a subpart of the information, the global interpretation resulting from their composition.

However, this approach appears to be too simple when faced with the description of natural

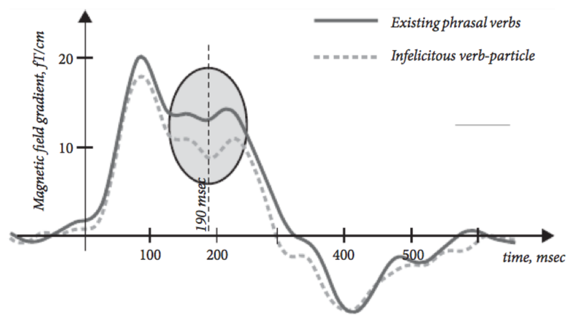


Figure 5: Early EEG effects of syntactic violation (MMN)

data. First, we know that language is intrinsically heterogeneous, made of different sources of information, different modalities, that interact at any time, producing a complex signal. In a natural environment (typically conversation), the linguistic signal is made of different sources, that cannot be strictly separated and analyzed independently from the others. It is made of multiple streams that are not strictly temporally aligned.

This view of language fits better with many observations presented above. First, language processing often stays at a shallow level, leading to incomplete processing: *chunks*, identified in terms of basic properties instead of complete analysis, can be considered as the basic processing unit, also giving access to a certain level of meaning and interpretation.

These objects can be at different levels and result from the convergence of different sources of information. In the following, we introduce a distinction between chunks and constructions:

- *Chunk*: group of words, gathered on the basis of low-level morpho-syntactic properties
- *Construction*: chunks (or set of chunks) that can be associated to a global meaning, sometimes figurative

Chunks and constructions are described as sets of interacting *properties* instead of structures that are built step-by-step from atomic to complex objects. These properties can be at a *low level* and *automatically assessed*.

At the interpretation level, in line with the notion of construction, we have seen that *meaning* can in some cases be accessed *directly* instead of compositionally (e.g. idioms, multi-words expressions, etc.). Moreover, interpretation can be often only incomplete, or *underspecified*. In particular, it has been shown that ambiguity can be left unresolved and interpretation *delayed* (or even never completely built).

We propose to take into consideration these different features, gathering them into a language processing architecture capable of accounting for different types of processing, at different levels, depending on the type of input to deal with. The objective is to describe any kind of situation, from the more controlled one (for example laboratory speech, isolated words, etc.) to the more natural one (e.g. conversations). The proposal relies on the idea that, according to the context and the sources of information, LP can be either serial, modular and compositional or on the opposite parallel, integrated and directly interpretable. This approach relies on an hybrid processing: one, at a low level, a shallow and partial which is supposed to be at work by default and the second, relying on deep, modular and compositional parsing, that is triggered when processing complex material (in other words when interpretation becomes difficult). This organization relies on several assumptions:

- The basic objects in LP are chunks, instead of words
- Chunks are specified in terms of low-level properties, automatically assessed
- Semantic interpretation can be delayed
- Chunks offer the possibility of a direct access to the meaning

We propose in the remaining of this paper to investigate these aspects by addressing specific questions, with elements of answer:

- What is the nature of basic properties: constraints
- How can they specify entire chunks: constraint interaction
- How to access directly from low-level properties to meaning: constructions

3 The theoretical framework: Property Grammars

We present in this section the main features of *Property Grammars* (noted PG) [Blache, 2000]. PG is a linguistic theory that proposes a constraint-based processing architecture; More precisely, all linguistic information in PG is represented by means of different properties (implemented as constraints). At the difference with the classical generative paradigm, there is no specific module: all properties are mutually independent, offering the possibility to represent separately the different types of information, whatever their domain (morphology, syntax, semantics, etc.) or their level (relations between features, categories, chunks, etc.) . These properties connect the different words of a sentence when processing an input. As a consequence, instead of building a structure, the processing mechanism consists here in describing the input by identifying its different properties. Focusing on syntax and semantics, the following list summarizes the possible relations between words:

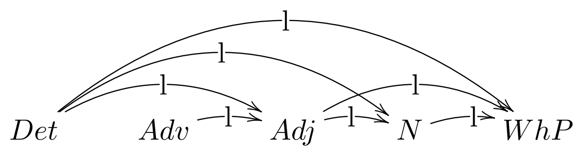
- Linearity: linear order that exists between two words
- Co-occurrence: mandatory co-occurrence between two words
- Exclusion; impossible co-occurrence between two words
- Uniqueness: impossible repetition of a same category
- Dependency: syntactic-semantic dependency between two words. Different types of dependencies are encoded: complement, subject, modification, specification, etc.

In PG , a grammar is a set of all the possible relations between categories, describing the different constructions. When parsing a given sentence S, evaluating a property of the grammar consists in verifying whether the relations between two categories corresponding to words of S are satisfied or not. We propose in the following an overview of each typeof property.

Linearity: In PG is the same kind of linear precedence relation as proposed in GPSG [Gazdar et al., 1985] is represented. For example, the nominal construction in English must follow the linearity properties:

$$Det \prec Adj; \quad Det \prec N; \quad Adj \prec N; \quad N \prec WhP; \quad N \prec Prep \tag{1}$$

Note that relations are expressed directly between the lexical categories. As such, the $N \prec Prep$ property indicates precedence between these two categories regardless of their dependencies. The following example illustrates the linearity relations in the nominal construction "*The very old reporter who the senator attacked*":



(2)

In general, properties are also used to control attribute values. For example, one can distinguish linearity properties between the noun and the verb, depending on whether N is *subject* or *object* by specifying this value in the property itself:

$$N[subj] \prec V; \quad V \prec N[obj] \tag{3}$$

Cooccurrence: This property typically represents subcategorization, implementing the situation where two categories must be realized together. As for verbal constructions, a classical example of co-occurrence concerns nominal and prepositional complements of ditransitive verbs, which are represented through the following properties:

$$V \Rightarrow N; \quad V \Rightarrow Prep \tag{4}$$

It should be noted that cooccurrence not only represents complement-type relations, it can also include co-occurrence properties directly between two categories independently form the head. For example, the indefinite determiner is not generally used with a superlative:

(1) a. The most interesting book of the library

*b. *A most interesting book of the library*

In this case, there is a co-occurrence relation between the determiner and the superlative, which is represented by the property:

$$Sup \Rightarrow Det[def] \tag{5}$$

Exclusion: In some cases, restrictions on the possibilities of co-occurrence between categories must be expressed. These include, for example, cases of lexical selection, concordance, etc. The following properties describe some co-occurrence restrictions in nominal constructions:

$$Pro \otimes N; \quad N[prop] \otimes N[com]; \quad N[prop] \otimes Prep[inf] \tag{6}$$

These properties stipulate for example that a pronoun and a noun, a proper noun and a common noun, a proper noun and an infinitive construction introduced by a preposition cannot be realized simultaneously.

Uniqueness: Certain categories cannot be repeated inside a rection domain. More specifically, categories of this kind cannot be instantiated more than once in a given domain. The following example describes the uniqueness properties for nominal constructions:

$$Uniq = \{Det, Rel, Prep_{[inf]}, Adv\} \tag{7}$$

These properties are classical for the determiner and the relative pronoun. They also specify here that it is impossible to duplicate a prepositional construction that introduces an infinitive (*the will to stop*) or a determinative adverbial phrase (*always more evaluation*).

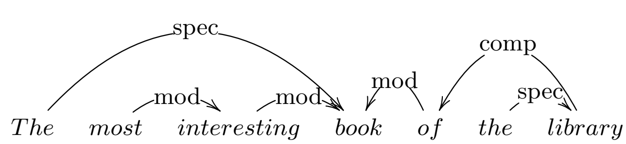
Dependency: This property describes syntactic-semantic relations between categories, indicating that the dependent category complements governor and contributes to its semantic structure. Dependency relations are type-based, following a type hierarchy, rendering possible to vary the level of precision of the relation, from the most general to the most specific. These types and sub-types corresponds to a classical syntactic relation:

dep	:	generic relation, indicating dependency between a constructed component and its governing component.
mod	:	modification relation (typically an adjunct).
spec	:	specification relation (typically *Det-N*).
comp	:	the most general relation between a head and an object (including the subject).
subj	:	dependency relation describing the subject.
obj	:	dependency relation describing the direct object.
iobj	:	dependency relation describing the indirect object.
xcomp	:	other types of complementation (for example between *N* and *Prep*).
aux	:	relation between the auxiliary and the verb.
conj	:	conjunction relation.

Dependency (noted $\rightsquigarrow$), possibly bearing the dependency sub-type as an index. The following properties indicate the dependency properties applied to nominal constructions:

$$Det \rightsquigarrow_{spec} N[com]; \quad Adj \rightsquigarrow_{mod} N; \quad WhP \rightsquigarrow_{mod} N \tag{8}$$

The following example illustrates some dependencies into a nominal construction:



(9)

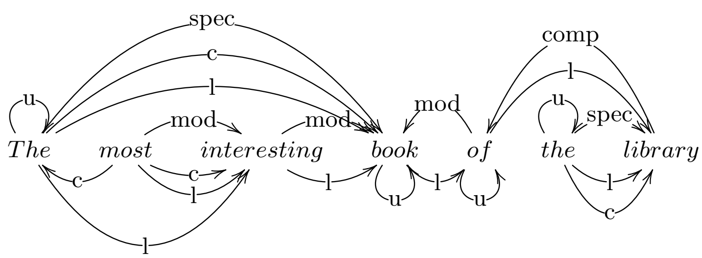
In this schema, we can see the specification relations between the determiner and the noun, and the modification relations between the adjectival and prepositional constructions as well as between the adverb and the adjective inside the adjectival construction.

Example: Each property as defined above corresponds to a certain type of syntactic information. In PG , the description of syntactic units or linguistic phenomena (chunks, constructions) in the grammar consists in gathering all the relevant properties into a set. The following table summarizes the properties describing the nominal construction:

$Det \prec \{Det, Adj, WhP, Prep, N\}$ $N \prec \{Prep, WhP\}$	$Det \leadsto_{spec} N$ $Adj \leadsto_{mod} N$
$Det \Rightarrow N[com]$ $\{Adj, WhP, Prep\} \Rightarrow N$	$WhP \leadsto_{mod} N$ $Prep \leadsto_{mod} N$
$Uniq = \{Pro, Det, N, WhP, Prep\}$	$Pro \otimes \{Det, Adj, WhP, Prep, N\}$ $N[prop] \otimes Det$

(10)

In this approach, a syntactic description, instead of being organized around a specific structure (for example a tree), consists in a set of independent (but interacting) properties together with their status (satisfied or violated). The graph in the figure below illustrates the PG description of the nominal construction: "*The most interesting book of the library*".



(11)

In PG, a syntactic description is therefore the graph containing all the properties of the grammar that can be evaluated for the sentence to be parsed. As illustrated in the example, this property graph represent explicitly all the syntactic characteristics associated to the input, each of them being represented independently from the others.

4 Chunks, constructions and properties

As underlined in the first section, many observations tend to show that chunks and constructions are to be taken into consideration in a LP architecture. More precisely, our hypothesis is that two different types of processing coexist: one serial, incremental, word-by-word and compositional (the classical LP organization in the literature) and another shallow, based on chunks or constructions, recognized as a whole, giving when possible a direct and global access to the meaning. This hypothesis is supported by several observations, showing the existence of such units in particular when studying the brain correlates of language processing. Moreover, some basic properties of such units (typically agreement) are identified at a very early stage, indicating an automatic and low-level process. It is then necessary to explain first what are these basic properties and second how can chunks or constructions be recognized starting from them.

4.1 Basic properties

The different properties presented above can be assessed directly when processing a sentence: for each set of categories, it is possible to verify whether or not some properties link them in the grammar and whether in the specific context of their realization in the sentence, they are verified or not. In this perspective, a property play exactly the role of a constraint, describing an input consists in assessing the properties, assigning them a truth value.

Two different types of properties can be distinguished, according to the way they can be evaluated and their dependence to the context [Blache and Rauzy, 2004]. More precisely, a property can be assessed as soon as the categories they apply to are realized in a sentence.

The difference between the two types of properties is that in one case, the truth value (i.e. the result of their evaluation) remains the same whatever the window of words taken into consideration, and in the other case, this value can vary depending on the window (being then sensitive to the context). These two types of properties can be precised as follows:

- *Success-monotonic properties*: when a property between two categories is true for a given sentence, it remains true when parsing the rest of the sentence. For example, the linearity between *most* and *interesting* in Fig. 11 holds as soon as it can be evaluated, and remains true until the end, whatever the span of words.

  In a more formal manner, the linearity relation $a \prec b$ is true in the sequence of words $s = [\gamma, a, b, \eta]$, whatever the composition of γ and η. Two types of properties are success-monotonic: *linearity* and *co-occurrence*.

- *Success-non monotonic properties*: A property can be true locally and false at a larger span: the evaluation of a property depends on the set of categories taken into account. For example an *exclusion* relation between the words a and d is true within the set of words $s1 = a, b, c$, but false when adding a new category d to this sequence $s2 = a, b, c, d$. In this case, it is then necessary to choose a partition into which evaluating the constraint.

Success-monotonic properties are then computationally simpler than non-monotonic because they do not need to be re-evaluated at each step. We consider in the following these types of properties as *basic*. They are the low-level properties, automatically assessed at an early stage. Moreover, it is interesting to note that they encode the two types of information used when evaluate transition probabilities between categories (linearity and cooccurrence), reinforcing the proposal to consider them at a first level.

4.2 Chunks from properties

We have seen how to represent and assess properties. The question is now how they can be used in order to recognize a chunk or a construction. As presented in the first section, several experiments have shown that some syntactic properties can be assessed very early, without any deep and precise analysis (see in particular [Pulvermuller et al., 2008]). In our hypothesis, they correspond to "basic properties", which evaluation can be done on the basis of the immediate context. However, the question is to know how, starting from such basic properties, it is possible to identify higher level organizations such as chunks and constructions.

Our proposal relies on the idea that in some cases, there exists a link between properties: the existence of some properties can also activate other properties. For example, the verification of a linearity property between a *Det* and a *N* activates a dependency relation between them. This is the same kind of relation that exists in lexical selection, collocations, etc.: the realization of a given word activates or predicts that of another one.

A se consequence, the description of a construction in the grammar consists in two types of information: the set of properties and the identification of those basic properties that can activate other ones. We propose to add to the representation of the properties this information by adding a new argument encoding the properties that can be linked to the current as follows:

```
<id, type, source, target, weight, linked_props>
```

The linked_props argument is a set of indexes, pointing towards other properties describing the same construction. For example, the dependency relation between a preposition and a noun depends on the linearity: if $Prep \prec N$, then *Prep* is the head and N depends on it. Reciprocally, when $N \prec Prep$, the *Prep* depends on N. These relations between properties are represented as follows:

```
<1, lin, Prep, N, H, {}>  <2, comp, N, Prep, S, {1}>
<3, lin, N, Prep, H, {}>  <4, mod, Prep, N, S, {3}>
```

The example of the ditransitive construction can be implemented in the same manner, specifying different dependency types according to the form (the first noun is the indirect object, the second the direct):

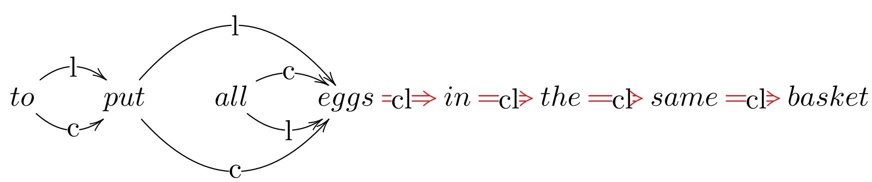
```
<1, lin, V[dit], N1, H, {}>  <4, iobj, N1, V, H, {1,2,3}>
<2, lin, V[dit], N2, H, {}>  <5, obj, N2, V, H, {1,2,3}>
<3, lin, N1, N2, H, {}>
```

4.3 Processing Idioms

This activation mechanism based on linked properties can be generalized to the processing of other types of constructions such as idioms. In this case, we know that the processing, including its brain correlates (see [Vespignani et al., 2010, Rommers et al., 2013]), is done in two steps: before and after reaching the word starting from which the idiom is recognized (called hereafter the recognition point, or *RP*). Before the *RP*, the processing consists in assessing the basic properties. At the *RP*, the idiom is recognized, its meaning is globally accessed, without any need to analyze the rest of the idiom. All the remaining words become fully predictable. In terms of properties, this means that a set of mandatory coccurrence as well as linearity between the list of words is activated.

This phenomenon can be implemented with the mechanism of *linked properties*: reaching the *RP* means having already assessed a certain amount of basic properties, relating the initial words of the idiom. Recognizing the *RP* consists in inferring a set of linked properties starting to the basic ones.

The following figure illustrates this mechanism for the idiomatic expression "*to put all eggs in one basket*" ;



In this idiom, the *RP* is at the word "*eggs*". Before the *RP*, the basic properties are assessed, linking the first words of the idioms. After this point, all the other properties can be automatically inferred, as well as the association of a global meaning. The property-based description of this idiom can be implemented around the following properties:

(1) $put \prec all$
(2) $all \prec eggs$
(3) $put \Rightarrow \{in, one, basket\}$
(4) $eggs \prec in \prec one \prec basket$
(5) $sem(put) = [[risk\_losing\_everything]]$

In this description, we only describe the basic linearity and cooccurrence properties. The RP is implemented by the factorized properties (3) and (4). The general mechanism is described to its turn by the following formula, indicating that properties (3), (4) and (5) can be inferred directly form the basic properties (1) and (2):

$$(1 \wedge 2) \Rightarrow (3 \wedge 4 \wedge 5)$$

Note that the semantics of the idiom is represented by a denotation attached, arbitrarily, to the verb (the idiomatic construction being verbal in this case).

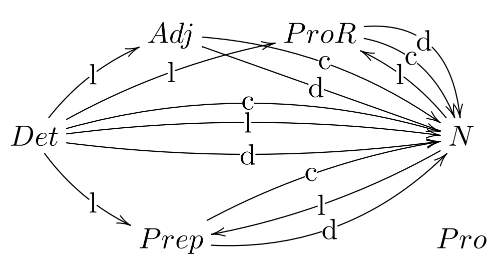
4.4 Cohesion

Besides inferred properties thanks to the linkage mechanism, constructions can be identified on the basis of the cohesion of theirs specific information. A construction being defined as a set of interacting properties, a PG based representation makes it possible to propose a measure based on the evaluation of their number and their importance.

As shown above, the linguistic description of a construction corresponds to a set of properties that can be represented as a graph in which nodes are words and edges represent the relations. The graph density constitute a first type of indication: a high density corresponds to a high number of properties, representing then a certain type of cohesion between the words. Moreover, the quality of these relations can also be evaluated: some properties can be more important than some others (in other words, they have different weights). A high density of heavy properties constitute then a second type of information. Finally, some sentences can be non-canonical, bearing certain properties that can be violated (for example agreement or linear precedence violation). Taking into consideration the number of violated properties in comparison with the satisfied ones is the last type of indication we propose to use in the evaluation of the cohesion.

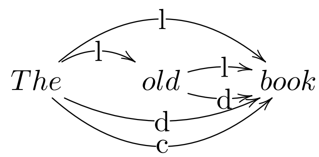
Our hypothesis is that a correlation exists between cohesion, defined on the basis of these three types of information and constructions. In other words, a construction correspond to a set of words linked with a high number of properties, of heavy weights, with few or no violations.

The first parameter of the cohesion measure relies on the number of properties that are assessed for a given construction, in comparison with the possible properties in the grammar. The following graph illustrates the set of properties in the grammar describing the nominal construction:



The number of possible relations in which a category is involved can be estimated by the number of incident relations of the corresponding vertex in the graph (called in graph theory the vertex degree). We propose then to define the degree of a category by this measure. In the previous example, we have the following degrees: $deg_{[gram]}(N) = 9; deg_{[gram]}(ProR) = 2; deg_{[gram]}(Adj) = 1$.

The same type of evaluation can also be applied to the constraint graph describing a construction or a sentence, as in the following example:



Each word is involved into a set of relations. The degree of a word is, similarly to the grammar, the set of incident edges of a word. In this example, we have: $deg_{[sent]}(N) = 5; deg_{[sent]}(Adj) = 1; deg_{[sent]}(Det) = 0$.

The first parameter of our estimation of the cohesion relies on a comparison of these two values: for a given word, we know from the grammar the number of properties in which it could theoretically be involved. We also know from the parsing of a given sentence how many of these properties are effectively assessed. We can define then a value, the *completeness ratio*, indicating the density of the category: the higher the number of relations in the grammar is verified, the higher the completeness value:

$$Comp(cat) = \frac{deg_{[sent]}(cat)}{deg_{[gram]}(cat)}$$

Besides this completeness ratio, it is also interesting to examine the density of the constraint graph itself. In graph theory, this value is calculated as a ratio between the number of edges and the number of vertices. It is more precisely defined as follows (S is the constraint graph of a sentence, E the set of edges, V the set of vertices):

$$Dens(S) = \frac{|E|}{5 * |V|(|V| - 1)}$$

In this formula, the numerator is the number of existing edges, the denominator is the total number of possible edges (each edge connecting two different vertices, multiplied by 5, the number of different types of properties). This value makes it possible to distinguish between *dense* vs. *sparse* graphs. In our hypothesis, a dense graph is correlated with a construction.

The last parameter taken into account is more qualitative and takes into account the weights of the properties. More precisely, we have seen that all properties can be either satisfied or violated. We define then a normalized satisfaction ratio as follows (where W^+ is the sum of the weights of the satisfied properties and W^- that of the violated ones):

$$Sat(S) = \frac{W^+ - W^-}{W^+ + W^-}$$

Finally, the cohesion value can be calculated as a function of the three previous parameter as follows (C being a construction, G_C its corresponding constraint graph):

$$Cohesion(C) = \sum_{i=1}^{|S|} Comp(w_i) * Dens(G_C) * Sat(G_C)$$

Note that the *density* ant *satisfaction* parameters can be evaluated directly, without depending on the context (as it is the case for success-monotonic properties) and without needing to know the type of the construction. On the contrary, evaluating the *completeness* parameter requires to know the construction in order to extract from the grammar all the possible properties that describe it. In a certain sense, the two first parameter are *basic*, in the same sense as described for properties, and can be assessed automatically.

5 The hybrid architecture

The language processing architecture we propose is an alternative to the classical incremental, modular and serial organization. We think that, instead of processing word-by-word, trying to integrate each new word into a partial structure, interpreting it compositionally (the meaning of the whole being a function of each component), it is preferable to propose a flexible architecture, more in line with what is observed in human LP.

5.1 General organization

The first general idea is that processing is not strictly incremental and meaning access not compositional. We have seen in the first section that there exists a basic level of processing gathering word into larger units thanks to a basic processing level. Moreover, there also exist some situations (typically constructions) in which meaning can be accessed directly. This means that two different types of processing are juxtaposed: one consisting in identifying directly large units (chunks) that offer the possibility of a global processing. This process relies on low-level mechanisms and is considered to be the default one. The second type is classical, word-by-word, serial and compositional, and is applied when the first is not possible.

Generally speaking, we consider that the first level of processing is superficial and delay as much as possible the interpretation. Whenever possible, larger units grouping several words are built. Such groups make it possible to gather different sources of information, preparing a first level of interpretation. In some cases, they even constitute entire construction offering the possibility to access directly to the meaning. This first level of processing is supposed to be done automatically, on the basis of low-level mechanisms.

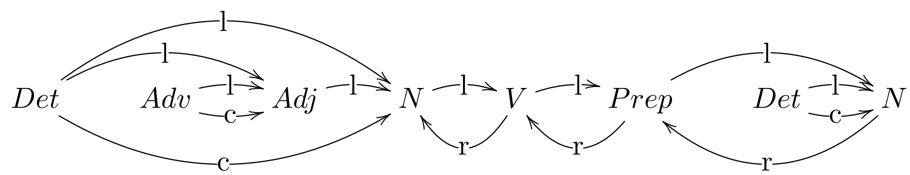
These units are stored in the working memory, together with their interpretation when it exists. The global interpretation consists in gathering the different local meanings. When no grouping is possible, then words instead of groups are stored in the memory. In both case, the interpretation process is only done after gathering a certain amount of information (or when reaching the maximal capacity of the working memory).

As a result, different types of objects coexist: words, chunks and constructions. The existence of units grouping words facilitates the processing: these objects are recognized by means of low-level mechanisms and offer the possibility to directly contribute to the meaning. We propose in the following the description of the identification of these units.

5.2 Recognizing chunks

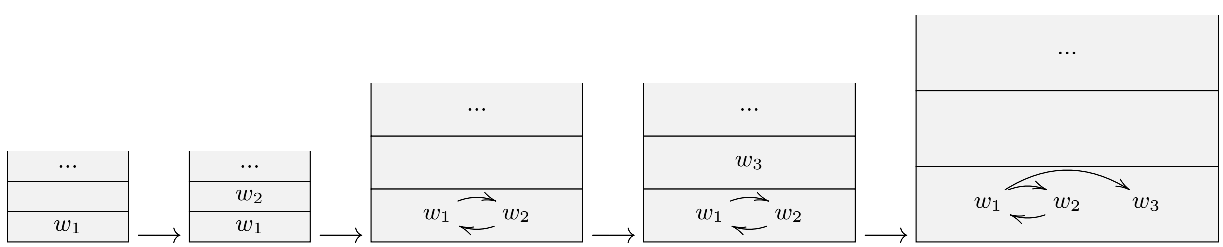
Chunks are set of words, usually adjacent (but not necessarily), linked by tight morpho-syntactic relations (typically *Det/N*). As presented in the previous section, such relations mainly corresponds to what has been called “basic properties”: linearity and cooccurrence. When parsing an input, the processing of a new word consists in checking such properties with adjacent words. The resulting graph makes it possible to identify subgroups, formed by the set of words that are connected by such properties. When looking at a constraint graph obtained from basic properties, such subgroups can be immediately identified: they correspond to the complete subgraphs (the set of nodes in a graph that are directly connected).

The following example illustrates a constraint graph, containing basic properties only:



In this graph, we can identify several subset of nodes that are all connected. For example, the subset *Adv-Adj* forms a complete subgraph, but not *Det-Adv-Adj*: in this last case, *Det* is not connected with *Adv*. A complete subgraph is then made of words with morphosyntactic relations which correspond to our definition of chunks. The list of complete subgraphs of the example above is : *Adv - Adj; Det - Adj - N; N - V; V - Prep; Prep - N; Det - N.* This correspond to the list of the chunks, identified only by means of basic properties.

Chunk recognition constitutes the first level of processing. In the following, we leave aside the question of word segmentation and recognition during human LP (even though it remains an open question) and consider the initial input as a set of words. The processing consists then in scanning the input word by word. The initial word of the input is simply stored in the working memory which is made of several buffers (represented here as a stack). After this initial step, the process consists in scanning a new word, pushing it into the stack of buffers and then looking whether it can be connected with the word of the lower buffer in the stack by basic properties. When the current word can be linked to the previous word, they form a chunk which is pushed onto the buffer. In the same way, a word can be linked to an already existing chunk, forming a new chunk replacing the previous one, ass illustrated in the following figure:



5.3 Recognizing constructions

The identification of constructions can be explained thanks to linked properties and cohesion evaluation. They constitutes distinct mechanisms making it possible, starting from a chunk, to identify the type of the construction and complete its description with new properties.

A chunk is made of words connected with basic properties. In some cases, this set of properties can be associated to linked properties, resulting in the inference of new relations completing the constraint graph.

As shown above, this mechanism explains the effect of the recognition point in idiom processing. Before this point, the preceding words are processed as explained in the previous section, building chunks when possible. When reaching the word corresponding to the RP, a set of linked properties is directly assessed. We obtain then a new constraint graph which also bears for idioms a complete interpretation: in this case, the entire set of properties describing this specific construction is formed by linked properties, making it possible to infer directly the description and its interpretation.

For other types of constructions, with a certain level of flexibility [Goldberg, 2006], only a subpart of the properties are linked and can be automatically inferred from the basic ones. In this situation, an evaluation of *cohesion* complete the mechanism. The constraint graph, completed by the possible linked properties is analyzed. If its cohesion value reach a certain threshold, then the construction is recognized and the set of properties that describes it is activated. Such activation can lead to a partial interpretation, to be completed when scanning the rest of the input.

5.4 Light parsing with chunks and constructions

The recognition of chunks and a fortiori constructions facilitates language processing in the sense that they give a direct access to a certain interpretation (that can be complete in some cases of constructions) thanks to basic properties. In our approach, this constitutes the first level of processing: based on the assessment of basic properties that can trigger the inference of other types of properties thanks to the mechanism of linked properties. In this hypothesis, constructions are encoded in the memory as recurrent networks, encoding directly linked properties. This view fits with the MUC model (memory, unification control) proposed in [Hagoort, 2005, Hagoort, 2013] in which memory contains lexical building blocks that encodes complex lexical entries, including syntactic and semantic relations. MUC is in line with modern linguistic theories such as HPSG [Pollard and Sag, 1994] or TAG [Joshi and Schabes, 1997] in which most of syntactic and semantic information is encoded in the lexicon. The memory stores such units when the unification component is in charge of integrating them. In our model, the linked properties are stored in the memory together with lexical units. The unification component can then directly assemble units thanks to a simple mechanism: basic properties assessment plus linked properties inference.

The light processing architecture distinguishes then two levels of unification during sentence processing:

- Light level: used as default, storing words in the working memory, assembling them into chunks, inferring linked properties and activating constructions when possible
- Deep level: used when the light level does not lead to interpretation. The processing is classical: strictly incremental and serial, interpretation being built compositionally, starting from a syntactic structure.

6 Conclusion

Sentence processing is fast, happens in real time in spite of the complexity of linguistic mechanisms. One classical explanation is that language make use of frequent structures or patterns, that can be learnt, rendering LP mainly probabilistic. However, several experiments have shown that even some types of rare structures can be processed automatically, in real time [Pulvermüller, 2010]. We describe in this paper a new way of representing linguistic information, based on properties and show how two types of such properties can be distinguished, according to the way they can be assessed. Some properties are assessable in a simple and direct manner: they constitute “basic properties”. Moreover, we have shown how properties, whatever their type, can be linked and directly inferred from each others. This mechanism (basic assessment+inference) is the basis of the recognition of chunks and even, in some cases, of constructions. This form then the first level of parsing in our model, based on “light parsing”. This processing mechanism is the default one, explaining why language processing can be often shallow but fast: interpretation can be directly accessed thanks to such basic mechanism. In some difficult and complex cases, light parsing does not lead to interpretation. In this case, the classical incremental and serial processing mechanism is used.

The light-and-deep parsing constitutes then a candidate for a new language processing architecture, explaining why human LP is efficient and opening the way to new types of experiments in neurolinguistics.

References

[Abney, 1991] Abney, S. (1991). Parsing by chunks. In *Principle-Based Parsing. Kluwer Academic Publishers*, pages 257–278.

[Adriaens and Hahn, 1994] Adriaens, G. and Hahn, U., editors (1994). *Parallel Natural Language Processing.* Ablex Publishing Corporation.

[Altmann and Mirković, 2009] Altmann, G. T. M. and Mirković, J. (2009). Incrementality and prediction in human sentence processing. *Cognitive Science*, 33(4):583–609.

[Baldwin et al., 2002] Baldwin, T., Dras, M., Hockenmaier, J., King, T. H., and van Noord, G. (2002). The impact of deep linguistic processing on parsing technology. In *Proceedings of IWPT-2007*.

[Balfourier et al., 2002] Balfourier, J.-M., Blache, P., and Rullen, T. V. (2002). From Shallow to Deep Parsing Using Constraint Satisfaction. In *Proc. of the 6th Int'l Conference on Computational Linguistics (COLING 2002)*.

[Blache, 2000] Blache, P. (2000). Property grammars and the problem of constraint satisfaction. In *Linguistic Theory and Grammar Implementation*, ESSLLI 2000 workshop.

[Blache and Rauzy, 2004] Blache, P. and Rauzy, S. (2004). Une plateforme de communication alternative. In *Actes des Entretiens Annuels de l'Institut Garches*, pages 82–93, Issy-Les-Moulineaux, France.

[Chomsky, 1981] Chomsky, N. (1981). *Lectures on Government and Binding.* Foris, Dordrecht.

[Copestake et al., 2001] Copestake, A., Flickinger, D., Pollard, C., and Sag, I. (2001). Minimal Recursion Semantics: An Introduction. In *Language and Computation (L&C)*, volume 1, pages pp. 1–47. Hermes Science Publishing LTD.

[de Marneffe and Manning, 2008] de Marneffe, M.-C. and Manning, C. D. (2008). Stanford typed dependencies manual stanford typed dependencies manual stanford typed dependencies manual. Technical report, Stanford Parser v. 3.5.2.

[Ferreira and Patson, 2007] Ferreira, F. and Patson, N. D. (2007). The 'good enough' approach to language comprehension. *Language and Linguistics Compass*, 1(1).

[Fillmore, 1988] Fillmore, C. J. (1988). The mechanisms of "construction grammar". In *Proceedings of the Fourteenth Annual Meeting of the Berkeley Linguistics Society*, pages 35–55.

[Fodor, 1983] Fodor, J. (1983). *The modularity of mind: an essay on faculty psychology.* MIT Press.

[Fodor and Inoue, 1998] Fodor, J. and Inoue, A. (1998). Attach anyway. In Fodor, J. and Ferreira, F., editors, *Reanalysis in Sentence Processing*. Kluwer.

[Fodor and Ferreira, 1998] Fodor, J. D. and Ferreira, F. (1998). *Reanalysis in sentence processing*. London: Kluwer Academic Publishers.

[Frazier and Fodor, 1978] Frazier, L. and Fodor, J. (1978). The sausage machine: a new two-stage model of the parser. *Cognition*, 6:291–325.

[Friederici, 2002] Friederici, A. D. (2002). Towards a neural basis of auditory sentence processing. *Trends in Cognitive Sciences*, 6(22):78–84.

[Friederici, 2011] Friederici, A. D. (2011). The brain basis of language processing: From structure to function. *Physiological Reviews*, 91(4):1357–1392.

[Gazdar et al., 1985] Gazdar, G., Klein, E., Pullum, G., and Sag, I. (1985). *Generalized Phrase Structure Grammars*. Blackwell.

[Gibson, 2000] Gibson, E. (2000). The Dependency Locality Theory: A Distance-Based Theory of Linguistic Complexity. In Marantz, A., Miyashita, Y., and O'Neil, W., editors, *Image, Language, Brain*, pages 95–126. Cambridge, Massachussetts, MIT Press.

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

[Goldberg, 2003] Goldberg, A. E. (2003). Constructions: a new theoretical approach to language. *Trends in Cognitive Sciences*, 7(5):219–224.

[Goldberg, 2006] Goldberg, A. E. (2006). *Constructions at Work: The Nature of Generalization in Language*. Oxford University Press.

[Gorrell, 1995] Gorrell, P. (1995). *Syntax and Parsing*. Cambridge University Press.

[Grodner and Gibson, 2005] Grodner, D. J. and Gibson, E. A. F. (2005). Consequences of the serial nature of linguistic input for sentenial complexity. *Cognitive Science*, 29:261–291.

[Hagoort, 2005] Hagoort, P. (2005). On broca, brain, and binding: a new framework. *Trends in Cognitive Sciences*, 9(9).

[Hagoort, 2013] Hagoort, P. (2013). Muc (memory, unification, control) and beyond. *Frontiers in Psychology*, 4.

[Jackendoff, 2007] Jackendoff, R. (2007). A parallel architecture perspective on language processing. *Brain Research*, 1146(2-22).

[Jindal et al., 2013] Jindal, P., Roth, D., and Kale, L. (2013). Efficient development of parallel nlp applications. Technical report, Tech. Report of IDEALS (Illinois Digital Environment for Access to Learning and Scholarship).

[Joshi and Schabes, 1997] Joshi, A. K. and Schabes, Y. (1997). Tree-adjoining grammars. In Rozenberg, G. and Salomaa, A., editors, *Handbook of Formal Languages, volume 3: Beyond Words*, pages 69–124. Springer.

[Kaan, 2007] Kaan, E. (2007). Event-related potentials and language processing: A brief overview event-related potentials and language processing: A brief overview event-related potentials and language processing: A brief overview. *Language and Linguistics Compass*, 1(6).

[Keller, 2010] Keller, F. (2010). Cognitively plausible models of human language processing. *Proceedings of the ACL 2010 Conference Short Papers*, pages 60–67.

[Kutas et al., 2006] Kutas, M., Petten, C. K. V., and Kluender, R. (2006). Psycholinguistics electrified ii: 1994–2005. In Gernsbacher, M. A. and Traxler, M., editors, *Handbook of Psycholinguistics*, pages 659–724. Elsevier.

[Luck, 2005] Luck, S. J. (2005). *An Introduction to the Event-Related Potential Technique*. MIT Press.

[MacDonald et al., 1994] MacDonald, M., Pearlmutter, N., and Seidenberg, M. (1994). Lexical nature of syntactic ambiguity resolution. *Psycholical Review*, 101:676 –703.

[Marslen-Wilson and Tyler, 1980] Marslen-Wilson, W. and Tyler, L. (1980). The temporal structure of spoken language understanding. *Cognition*, 8:1–71.

[Pollard and Sag, 1994] Pollard, C. and Sag, I. (1994). *Head-driven Phrase Structure Grammars*. Center for the Study of Language and Information Publication (CSLI), Chicago University Press.

[Pulvermüller, 2010] Pulvermüller, F. (2010). Brain embodiment of syntax and grammar: Discrete combinatorial mechanisms spelt out in neuronal circuits. *Brain and Language*, 112(3):167–179.

[Pulvermuller et al., 2008] Pulvermuller, F., Shtyrov, Y., Hasting, A. S., and Carlyon, R. P. (2008). Syntax as a reflex: Neurophysiological evidence for early automaticity of grammatical processing. *Brain and Language*, 104(3):244–253.

[Rauzy and Blache, 2012] Rauzy, S. and Blache, P. (2012). Robustness and processing difficulty models. a pilot study for eye-tracking data on the french treebank. In *Proceedings of the 1st Eye-Tracking and NLP workshop*.

[Rommers et al., 2013] Rommers, J., Dijkstra, T., and Bastiaansen, M. (2013). Context-dependent Semantic Processing in the Human Brain: Evidence from Idiom Comprehension. *Journal of Cognitive Neuroscience*, 25(5):762–776.

[Schuett et al., 2008] Schuett, S., Heywood, C. A., Kentridge, R. W., and Zihl, J. (2008). The significance of visual information processing in reading: Insights from hemianopic dyslexia. *Neuropsychologia*, 46(10):2445–2462.

[Spivey and Tanenhaus, 1998] Spivey, M. J. and Tanenhaus, M. K. (1998). Syntactic ambiguity resolution in discourse: Modeling the effects of referential context and lexical frequency. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 24:1521–1543.

[Steedman, 2000] Steedman, M. (2000). Information structure and the syntax-phonology interface. *Linguistic Inquiry*, 31:649–689.

[Sturt and Lombardo, 2005] Sturt, P. and Lombardo, V. (2005). Processing coordinated structures: Incrementality and connectedness. *Cognitive Science*, 29(2).

[Swets et al., 2008] Swets, B., Desmet, T., Charles Clifton, j., and Ferreira, F. (2008). Underspecification of syntactic ambiguities: Evidence from self-paced reading. *Memory and Cognition*, 36(1):201–216.

[Uszkoreit, 2002] Uszkoreit, H. (2002). New chances for deep linguistic processing. In *proceedings of COLING-02*.

[Vespignani et al., 2010] Vespignani, F., Canal, P., Molinaro, N., Fonda, S., and Cacciari, C. (2010). Predictive mechanisms in idiom comprehension. *Journal of Cognitive Neuroscience*, 22(8):1682–1700.

[Werning et al., 2012] Werning, M., Hinzen, W., and Machery, E. (2012). *The Oxford Handbook of Compositionality*. Oxford University Press.