

Chunks et activation : un modèle de facilitation du traitement linguistique

Philippe Blache

Aix-Marseille Université, CNRS, LPL
5 Avenue Pasteur, 13100 Aix-en-Provence
blache@lpl-aix.fr

RÉSUMÉ

Nous proposons dans cet article d'intégrer la notion de chunk au sein d'une architecture globale de traitement de la phrase. Les chunks jouent un rôle important dans les théories cognitives comme ACT-R (Anderson *et al.*, 2004) : il s'agit d'unités de traitement globales auxquelles il est possible d'accéder directement via des buffers en mémoire à court ou long terme. Ces chunks sont construits par une fonction d'activation (processus cognitif pouvant être quantifié) s'appuyant sur l'évaluation de leur relation au contexte. Nous proposons une interprétation de cette théorie appliquée à l'analyse syntaxique. Un mécanisme de construction des chunks est proposé. Nous développons pour cela une fonction d'activation tirant parti de la représentation de l'information linguistique sous forme de contraintes. Cette fonction permet de montrer en quoi les chunks sont faciles à construire et comment leur existence facilite le traitement de la phrase. Plusieurs exemples sont proposés, illustrant cette hypothèse de facilitation.

ABSTRACT

Chunks and the notion of activation : a facilitation model for sentence processing

We propose in this paper to integrate the notion of chunk within a global architecture for sentence processing. Chunks play an important role in cognitive theories such as ACT-R cite Anderson04 : they constitute global processing units which can be accessed directly via short or long term memory buffers. Chunks are built on the basis of an activation function evaluating their relationship to the context. We propose an interpretation of this theory applied to parsing. A construction mechanism is proposed, based on an adapted version of the activation function which takes advantage of the representation of linguistic information in terms of constraints. This feature allows to show how chunks are easy to build and how they can facilitate treatment. Several examples are given, illustrating this hypothesis of facilitation.

MOTS-CLÉS : Chunks, ACT-R, activation, mémoire, parsing, traitement de la phrase, expérimentation.

KEYWORDS: Chunks, ACT-R, activation, memory, parsing, sentence processing, experimentation.

1 Introduction

L'interprétation d'un énoncé, à commencer par son traitement syntaxique, peut être plus ou moins facile pour un sujet humain. Plusieurs travaux proposent des éléments d'explication de cette variabilité. Au niveau syntaxique, des travaux proposent par exemple des explications en termes de distance pour une relation à établir entre deux éléments, une grande distance étant plus complexe à traiter qu'une plus faible (Gibson, 1998) ; (Grodner et Gibson, 2005). D'autres travaux portent sur l'identification d'un niveau d'activation des items en s'appuyant notamment sur des relations avec le reste de la structure en cours de construction (Lewis et Vasishth, 2005). Dans tous les cas, ces modèles de difficulté abordent la question d'un point de vue global, en tentant d'identifier les paramètres pouvant complexifier le traitement. Nous proposons dans cet article d'aborder un point de vue complémentaire en tentant d'identifier des facteurs qui au contraire peuvent permettre de *faciliter* le traitement.

En se situant dans l'hypothèse d'un traitement incrémental du langage, dans laquelle les mots sont intégrés au fur et à mesure de leur décodage dans une structure en cours de construction, des travaux antérieurs ont montré la possibilité de mesurer la quantité d'information linguistique¹ disponible au moment de l'intégration d'un mot. Dans les cas où le niveau d'information est élevé, le traitement (la compréhension) s'en trouve facilité. En revanche, un déficit d'information entraîne une complexification du traitement. En termes computationnels, la quantité d'information disponible permet de contrôler *l'espace de recherche* requis pour l'interprétation d'un énoncé. Une construction associée à une faible quantité d'information est très ambiguë et donc difficile à traiter car le nombre d'interprétations possibles (donc l'espace de recherche) est très grand. En revanche, une construction pour laquelle une grande quantité d'information (éventuellement redondante) est disponible sera peu ou pas ambiguë, son espace de recherche plus restreint et son traitement (son interprétation) devient plus facile. Dans certains cas, il n'y a aucune ambiguïté, le traitement est alors purement déterministe. La quantité d'information est dans ce cas un facteur de *simplification* du traitement et non pas de complexification.

D'une façon générale, la quantité (ou densité) d'information disponible est variable selon les parties de l'énoncé ou de la phrase. L'hypothèse que nous formulons est que les zones comportant une densité d'information importante sont traitées plus facilement que les autres. Dans certains cas, ces zones de haute densité peuvent être traitées d'un bloc. Nous nous intéressons dans cet article à cette idée que le processus d'intégration syntaxique pourrait se faire au niveau de ces zones plutôt qu'au niveau des mots. Une présence plus importante de zones de haute densité d'information dans un énoncé ou une phrase faciliterait ainsi son traitement. Cette idée s'appuie sur le principe *Maximize On-Line Processing* (noté *MoP*) proposé dans (Hawkins, 2003) :

The human parser prefers to maximize the set of properties that are assignable to each item X as X is parsed. [...] The maximization difference between competing orders and structures will be a function of the number of properties that are misassigned or unassigned to X in a structure S, compared with the number in an alternative.

Ce principe comporte plusieurs éléments. Il intègre tout d'abord l'idée selon laquelle, dans un processus incrémental, l'intégration d'un mot repose sur la vérification d'un ensemble de propriétés. Il indique également que deux constructions peuvent se distinguer par le nombre de propriétés qu'elles vérifient. La notion de densité d'information recoupe donc ce principe de

1. On entend ici par information linguistique toute propriété morpho-syntaxique ou syntaxique caractérisant la structure en cours de construction.

maximisation : un mot sera plus ou moins facilement intégré à la structure selon que le nombre de propriétés qui lui sont associées est important ou pas.

Notre hypothèse est que ces unités, définies par maximisation, correspondent en termes de traitement à des *chunks* tels que décrits dans les théories cognitives de type ACT-R (*Adaptive Character of Thought–Rational* (Anderson *et al.*, 2004)) et peuvent à ce titre être stockés en mémoire à court terme et bénéficier d’un accès direct.

2 Chunks et activation

La notion de chunk est bien connue en TAL, et généralement définie comme une suite de catégories non récursive, formée d’une tête, à laquelle peuvent être adjoints mots fonctionnels et modificateurs adjacents (Abney, 1991) ; (Bird *et al.*, 2009). Nous nous intéressons dans cet article à la façon dont ces chunks peuvent être construits, dans le cadre d’un processus incrémental, par un parseur humain.

2.1 Les chunks dans les théories cognitives

Le traitement du langage, comme celui des activités cognitives de haut niveau, repose sur la capacité d’identifier des unités de traitement pouvant être de taille et de nature variable. Cette idée est plus particulièrement développée par la théorie ACT-R et son adaptation au langage (Lewis et Vasishth, 2005), (Reitter *et al.*, 2011) dans laquelle les mécanismes de traitement s’organisent autour de buffers (jouant comme en informatique le rôle de mémoire tampon) pouvant mémoriser des chunks. Un chunk est dans cette approche décrit comme un ensemble de propriétés caractérisant une catégorie (ou une unité de plus haut niveau), pouvant par exemple contenir une structure syntaxique partielle (Lewis et Vasishth, 2005). Les chunks sont représentés en ACT-R par des structures de traits et peuvent représenter des objets atomiques ou complexes, offrant la possibilité pour un chunk de faire référence à un autre chunk et exprimer ainsi des relations. La définition d’un chunk est donc très générale et permet de référencer des structures incomplètes ou sous-spécifiées.

La théorie ACT-R s’intéresse d’une part aux processus de base et d’autre part aux structures de mémoire sur lesquelles ils s’appuient. Elle distingue notamment entre mémoire *procédurale* et *déclarative*, cette dernière permettant de stocker à la fois des informations lexicales (à long terme) mais également les structures nouvelles (à court terme). La mémoire déclarative repose sur un petit nombre de buffers, chacun contenant un chunk. L’élément important de cette organisation réside dans le fait que ces chunks forment une unité et sont utilisables (ou accessibles) directement en mémoire. Cette accessibilité est soumise à un niveau d’activation dépendant de plusieurs paramètres : degré de latence depuis le dernier accès, poids des éléments associés au chunk et qui peuvent l’activer (les sources), mais également force des relations associant les sources au chunk considéré. Il est ainsi possible de proposer une formule permettant de quantifier l’activation d’un chunk i :

$$A_i = B_i + \sum_j W_j S_{ji} \quad (1)$$

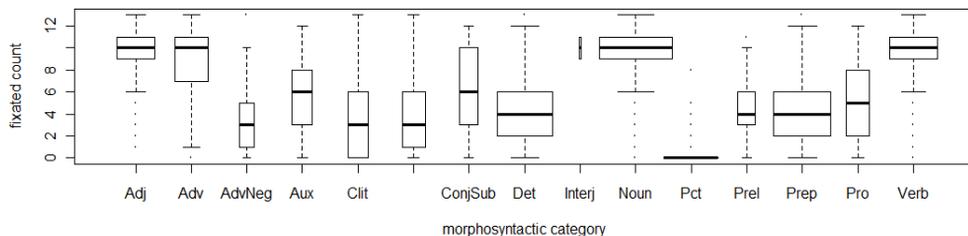


FIGURE 1 – Nombre de fixations par catégorie

Dans cette formule, B représente l’activation de base (fréquence et historique de l’accès au chunk), W correspond aux poids des termes en relation avec le chunk et S la force des relations reliant ces termes au chunk. Il est donc possible de caractériser un chunk en fonction de son niveau d’activation. Le point important qui nous intéresse ici réside dans le fait que cette activation est en partie dépendante des relations avec le contexte. En d’autres termes, la force des relations permettra d’activer de façon plus ou moins importante un chunk (et donc la catégorie correspondante). Or, l’activation d’un chunk contrôle à la fois sa probabilité et la vitesse de son accès : un chunk fortement activé sera ainsi accessible très rapidement.

On remarquera que cette approche est compatible avec le principe *MoP* de Hawkins (cf. section précédente) : les relations activant un chunk peuvent être vues comme des propriétés dont on recherche la maximisation.

Dans le cadre du traitement du langage et plus particulièrement de l’analyse syntaxique, notre hypothèse est que les chunks facilitent l’analyse d’un énoncé. Plus précisément, les énoncés comportant des chunks hautement activés sont traités plus facilement que les autres.

2.2 Une observation expérimentale des chunks dans le traitement de la phrase

Dans le cadre d’une expérience récente, consistant à acquérir des données de mouvement oculaire de sujets lisant le French Treebank (Rauzy et Blache, 2012), nous avons observé un phénomène intéressant en relation avec les chunks. Le nombre de fixations du regard par mot diffère en effet fortement en fonction de la taille du mot, mais également de sa catégorie. La figure 1 représente le nombre moyen de fixations par catégorie. On observe ainsi que les catégories à contenu lexical (N , V , Adj , Adv) ont un nombre de fixations du regard nettement plus élevé que les mots grammaticaux (Det , $Prep$, $Clit$, etc.).

Ce phénomène peut être mis en relation avec l’étude de l’évolution de l’*indice de surprise* (Hale, 2001) dans une phrase. Cet indice reflète une probabilité d’intégration de chaque mot dans la structures syntaxique en cours de construction (calculé comme une fonction de la différence de probabilité entre les structures précédant et celle intégrant le mot courant). Plusieurs expériences ont montré qu’il était un bon prédicteur du temps de lecture, pouvant donc être utilisé comme

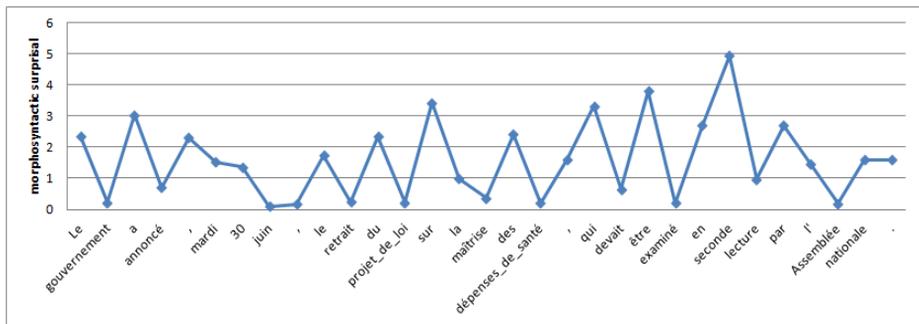


FIGURE 2 – Evolution de l'indice de surprise dans une phrase

mesure de *difficulté* (voir (Demberg et Keller, 2008) pour l'anglais et (Rauzy et Blache, 2012) pour le français). Un indice de surprise peut donc être associé à chaque mot de la phrase. La figure 2 illustre l'évolution de la valeur de cet indice (calculé selon la méthode décrite dans (Blache et Rauzy, 2011)) sur une phrase. On remarque là aussi un phénomène intéressant, soulignant la succession d'indices élevés et faibles en fonction de la catégorie : les mots grammaticaux correspondent systématiquement à un indice de surprise plus élevé que les mots lexicaux auxquels ils sont associés.

Ces deux observations sont convergentes : la fixation du regard en lecture englobe en un seul mouvement le token lexicalisé et les mots grammaticaux qui lui sont associés, ce qui peut être prédit au niveau de l'évolution de l'indice de surprise. Elles confortent donc l'hypothèse d'un traitement non pas au niveau du mot, mais directement par chunk, chaque fois que c'est possible.

2.3 Hypothèse

La théorie ACT-R appliquée au langage fait l'hypothèse que le traitement linguistique d'intégration repose sur des chunks. Ceux-ci sont des structures partielles, pouvant être à la fois stockées dans la mémoire à long terme, mais également construites en temps réel, en mémoire à court terme. Ces chunks reposent sur une notion d'activation, elle-même correspondant au principe *Maximize Online Processing* : l'intégration d'un mot à une structure (par exemple l'association de deux catégories pour construire un chunk) repose sur la vérification d'un maximum de propriétés. La force des relations unissant un objet avec des éléments qui le précèdent permet d'activer fortement cet objet.

Nous émettons l'hypothèse que les chunks facilitent le traitement linguistique. Nous nous appuyons pour cela sur trois aspects :

1. Les chunks sont construits en mémoire sur la base du processus d'activation, qui ne correspond pas à une véritable analyse syntaxique. Leur construction peut reposer sur des mécanismes de bas niveau (comme la fréquence de cooccurrence) ou sur l'accumulation de propriétés ou relations entre deux catégories. Lorsqu'une catégorie est fortement activée par une ou plusieurs catégories précédentes, elle formera un chunk avec elles. Dans la plupart des cas, ces chunks sont formés d'une suite [*mot grammatical + mot lexical*].

2. Les chunks sont stockés en mémoire déclarative et accessibles directement. Certains chunks peuvent être très fréquents voire correspondre à des suites plus ou moins figées (par exemple dans des collocations). Dans ce cas, ils sont stockés en mémoire à long terme. Les chunks construits dynamiquement sur la base d'une activation sont quant à eux disponibles dans des buffers de traitement à court terme.
3. La présence de chunks dans une phrase facilite son traitement : ils sont accessibles d'un bloc et ne nécessitent pas d'analyse. Une phrase contenant des chunks sera plus facile à traiter qu'une autre n'en contenant pas.

La question qui se pose est celle de la notion d'*activation*, son évaluation et sa mise en œuvre dans le processus de construction des chunks. Nous proposons pour cela d'utiliser la description des propriétés syntaxiques sous la forme de contraintes. Maximiser les propriétés (et donc activer une catégorie) correspond ainsi à la maximisation de l'ensemble des contraintes à satisfaire. Nous utilisons pour cela la représentation proposée dans le cadre des *Grammaires de Propriétés* (Blache, 2001).

3 Propriétés et activation

Nous présentons dans cette section les principales caractéristiques de l'approche des Grammaires de Propriétés (Blache, 2001) utilisées pour définir la notion d'activation. Elle repose sur la représentation des informations syntaxiques sous la forme d'un ensemble de propriétés pouvant être décrites, suivant la proposition de (Duchier *et al.*, 2009), comme des relations caractérisant un syntagme (ici noté A) et mettant en relation des constituants (notés B, C ou S) :

Obligation	$A : \Delta B$	au moins un B
Unicité	$A : B!$	au plus un B
Linéarité	$A : B \prec C$	B précède C
Implication	$A : B \Rightarrow C$	si $\exists B$, alors $\exists C$
Exclusion	$A : B \not\Rightarrow C$	pas de B et C simultanément
Constituance	$A : S?$	les descendants $\in S$
Dépendance	$A : B \rightsquigarrow C$	B dépend de C

Une Grammaire de Propriétés associe à chaque syntagme un ensemble de contraintes. Le tableau suivant illustre la grammaire du syntagme adjectival (noté SA) (extraite du French Treebank, cf. (Abeillé *et al.*, 2003)). Soulignons au passage la compacité de la représentation : 22 contraintes sont utilisées pour décrire les constructions possibles du SA^2 .

<i>Constituance</i>	$AP : \{AdP, A, VPinf, PP, Ssub, AP, NP\} ?$
<i>Lin</i>	$AP : A \prec \{VPinf, Ssub, PP, NP, AP\}$ $AP : AdP \prec \{A, Ssub, PP\}$ $AP : AP \prec \{A, AdP\}$ $AP : PP \prec \{Ssub\}$
<i>Dépendance</i>	$AP : \{AdP, VPinf, PP, Ssub, NP\} \rightsquigarrow A$
<i>Unicité</i>	$AP : \{A, VPinf, Ssub\} !$
<i>Obligation</i>	$AP : \Delta A$
<i>Exclusion</i>	$AP : VPinf \not\Rightarrow \{PP, Ssub\}$

2. Le jeu d'étiquettes utilisé est celui du FTB, notant AP pour syntagme adjectival, AdP pour syntagme adverbial, etc.

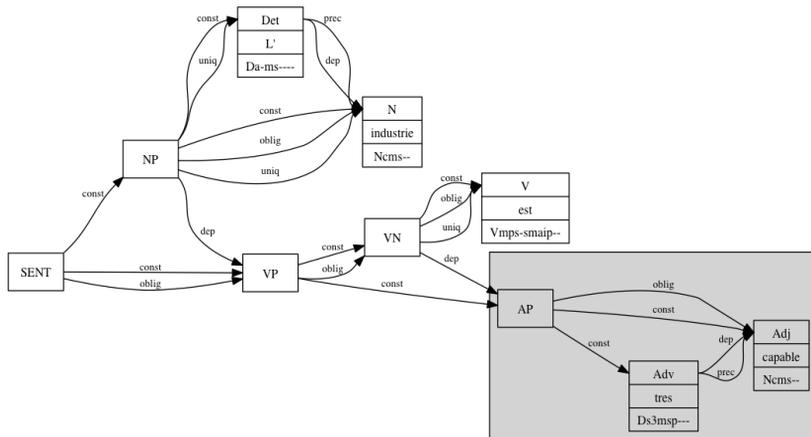


FIGURE 3 – Graphe des propriétés satisfaites pour “L’industrie est très capable.”

Une analyse dans le cadre de GP consiste, pour une suite de catégories donnée, à évaluer l’ensemble des propriétés correspondantes. Une propriété correspondant à une relation entre une ou plusieurs catégories, le résultat de l’analyse est donc un graphe comme représenté dans la figure suivante illustrant l’analyse de la phrase “L’industrie est très capable.”, extraite du FTB. Ce graphe indique les propriétés satisfaites entre les différentes catégories composant la structure syntaxique. Par exemple, la contrainte de linéarité entre le déterminant et le nom est représentée par un arc reliant les deux nœuds correspondants) :

Construire une analyse syntaxique dans ce type d’approche consiste donc à chaque étape à parcourir le systèmes de contraintes en évaluant celles qui correspondent aux catégories concernées. Dans une perspective incrémentale, il est donc possible à chaque étape de connaître les relations qui concernent le mot ou la catégorie à analyser. Cette caractéristique constituera la base de la définition de la notion d’activation utilisée ici.

Par ailleurs, il est possible de distinguer deux constructions en fonction du nombre de relations permettant de les caractériser. Dans l’exemple précédent, le SA est formé d’un adjectif accompagné d’un modifieur adverbial. L’exemple suivant illustre une construction légèrement différente d’un SA, correspondant à la phrase “L’industrie est capable d’investir.” dans laquelle une infinitive est complément de l’adjectif. Dans ce cas, conformément à la grammaire du SA décrite plus haut, un plus grand nombre de contraintes sera vérifiée, la densité du graphe est donc plus importante. Le nombre de propriétés vérifiées joue un rôle important en offrant la possibilité de quantifier l’information syntaxique. Dans la perspective du principe *MoP*, la maximisation reposera précisément sur cette capacité.

Un des avantages de cette approche réside dans sa souplesse : il est toujours possible d’évaluer les relations existant entre deux catégories, sans qu’il ne soit nécessaire de construire de structure syntaxique. Cette caractéristique répond au besoin d’évaluation de la notion d’activation d’une catégorie : celle-ci sera dépendante du nombre et de la force des relations existant entre un mot et les catégories qui la précèdent. Nous disposons ainsi d’un cadre théorique d’implantation des notions proposées par ACT-R appliquée au langage.

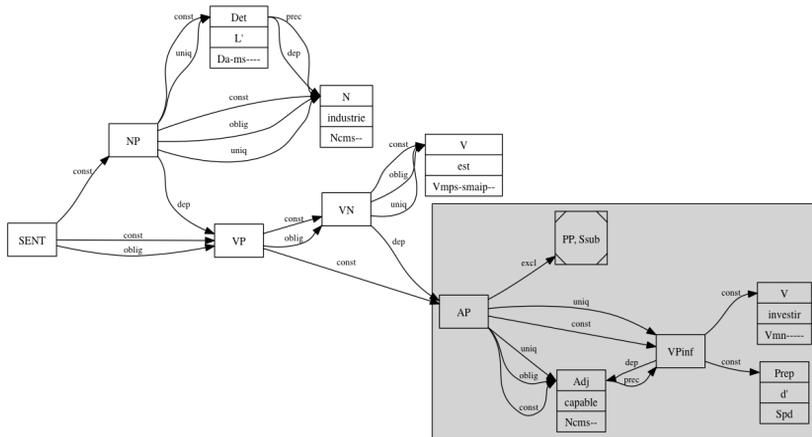


FIGURE 4 – Graphe des propriétés satisfaites pour “L’industrie est capable d’investir.”

4 Activation et création de chunks

Nous proposons de définir la notion d’activation sur la base des caractérisations syntaxiques construites à l’aide des contraintes présentées dans la section précédente. Nous avons vu qu’il était possible en *Grammaire de Propriétés* d’évaluer, pour tout sous-ensemble de catégories, les contraintes qui leur sont attachées. Il s’agit pour cela d’identifier les *contraintes pertinentes*, à savoir celles qui permettent de mettre en relation les catégories concernées. Le principe est simple et consiste à parcourir la grammaire (l’ensemble des contraintes) et sélectionner celles qui concernent les catégories. En reprenant l’exemple de la grammaire du syntagme adjectival décrite plus haut, le sous-ensemble de catégories $\{Adj, A\}$ permettra d’identifier comme pertinentes les contraintes suivantes :

AP : $\{Adj, A\}$?
AP : $Adj \prec A$
AP : $Adj \rightsquigarrow A$
AP : $A !$
AP : ΔA

En généralisant ce mécanisme, il est également possible d’identifier les contraintes qui sont *potentiellement pertinentes* : soit une contrainte $A \mathcal{R} B$ reliant deux catégories A et B , la connaissance de A permet de dire que $A \mathcal{R} B$ pourra devenir pertinente, à la condition que B soit réalisé. Dans le cas de la grammaire du SA, la réalisation de la catégorie Adj permet d’identifier comme contrainte potentiellement pertinente l’ensemble suivant :

AP : $\{Adj\}$?
AP : $Adj \prec A$
AP : $Adj \prec Adj$
AP : $Adj \rightsquigarrow A$

Nous proposons d’utiliser cette caractéristique pour décrire et évaluer la notion d’activation. Dans la perspective d’un traitement incrémental de la langue, le principe consiste à associer à chaque catégorie les contraintes potentiellement pertinentes qui peuvent lui être associées. Remarquons que du point de vue du traitement automatique, cette information n’a pas besoin d’être calculée *online*, mais peut être compilée. L’ensemble des contraintes ainsi identifiées permet de définir les catégories activées : il s’agit de toutes les catégories appartenant à cet ensemble et pouvant être réalisées après la catégorie en question. Cette dernière information est obtenue en vérifiant les contraintes de linéarité. Dans l’exemple précédent, seule la catégorie *A* se retrouve activée par *AdP* (la catégorie *AP* ne pouvant suivre *AdP* comme stipulé par la contrainte $AP : AP \prec AdP$).

4.1 Calcul du degré d’activation

Le niveau d’activation d’une catégorie dans un contexte donné dépend de sa densité ou, en d’autres termes, du nombre de contraintes dont elle est la cible (et dont la source la précède) et de leur poids. Il s’agit donc exactement de la notion d’activation telle que décrite dans la théorie ACT-R. Nous proposons d’évaluer cette activation en tirant parti de la représentation par contraintes. Pour chaque catégorie *c* de la grammaire, nous établissons une *liste de transition* formée par toutes les catégories présentes dans au moins une contrainte contenant *c* et respectant les contraintes de linéarité (i.e. pouvant suivre *c*). L’activation est alors évaluée comme suit :

- Soit la catégorie courante c_i . Notons $Trans(c_i)$ l’ensemble des catégories faisant partie de la liste de transition de c_i . Notons $PP(c_i)$ l’ensemble des propriétés potentiellement pertinentes déclenchées par la catégorie c_i . Notons N le nombre de ces propriétés ($N = |PP(c_i)|$).
- Notons $PP_{c_j}(c_i)$ le sous ensemble de $PP(c_i)$ formé des propriétés contenant une catégorie c_j , avec n son cardinal. Chacune des propriétés de PP est associée dans la grammaire à un poids. Notons $\sum W_{c_j}^{c_i}$ la somme des poids de ces propriétés.
- Pour toute catégorie de transition de c_i tq $c_j \in Trans(c_i)$, son degré d’activation est donné par la formule suivante :

$$A(c_j) = \frac{n}{N} * \sum W_{c_j}^{c_i} \quad (2)$$

Le premier terme de l’activation correspond à une évaluation de la densité du réseau de contraintes en rapportant le nombre de contraintes n qui permet d’activer la catégorie étudiée par rapport au nombre total de contraintes potentiellement pertinentes pour la catégorie source. Le second terme correspond quant à lui à la force des relations qui unissent la catégorie courante (ou catégorie activante) à la catégorie activée.

Concrètement, en cours d’analyse, cette mesure permettra d’identifier le type de catégorie activée par la catégorie courante ainsi que le niveau de son activation. Lorsque qu’une catégorie est activée et réalisée, elle formera un chunk avec la catégorie qui l’active. Ce chunk pourra avoir un niveau d’activation plus ou moins élevé, identifié par cette fonction d’activation. Notons que cette définition de l’activation permet également de rendre compte des relations lexicales du type collocationnelles. La sélection lexicale entre les termes sera dans ce cas représentée par une contrainte d’implication avec un poids élevée. Il sera ainsi possible de former un chunk doté d’un niveau d’activation fort.

L’exemple qui suit illustre l’utilisation de la fonction d’activation pour la construction d’un chunk à l’intérieur du *SN* entre les catégories *Det* et *N* en nous appuyant sur la grammaire extraite

du *French Treebank*. Les contraintes dont la catégorie *Det* est source sont répertoriées dans le tableau suivant, comportant également l'indication de leurs poids (calculé en suivant la méthode proposée dans (Blache, 2012)).

Dépendance		Linéarité	
Det \rightsquigarrow N	7,080586081	Det \prec N	12,18569885
Exclusion		Det \prec Np	0,718659942
Pro $\not\Leftarrow$ Det	4,358766626	Det \prec AdP	0,178675795
Clit $\not\Leftarrow$ Det	0,003417994	Det \prec AP	0,135447163
Unicité		Det \prec VPpart	0,077399536
Det	3,253068199	Det \prec VPinf	0,03891139
Exigence		Det \prec Ssub	0,025216138
Det \Rightarrow N	2,461019161	Det \prec Srel	0,021433718
		Det \prec PP	0,016570605
		Det \prec NP	0,016030259

L'ensemble de transition de *Det* extrait de ces contraintes est le suivant :

$$\text{Trans}(\text{Det}) = \{N, Np, AdP, AP, VPpart, VPinf, Ssub, Srel, PP, NP\} \quad (3)$$

L'évaluation du *degré d'activation* des catégories de l'ensemble de transition est récapitulée dans le tableau suivant :

Catégorie activée	Contraintes	Densité	Poids	Activation
N	3	0,2	21,72730409	4,345460818
Np	1	0,066666667	0,718659942	0,047910663
AdP	1	0,066666667	0,178675795	0,01191172
AP	1	0,066666667	0,135447163	0,009029811
VPpart	1	0,066666667	0,077399536	0,005159969
VPinf	1	0,066666667	0,03891139	0,002594093
Ssub	1	0,066666667	0,025216138	0,001681076
Srel	1	0,066666667	0,021433718	0,001428915
PP	1	0,066666667	0,016570605	0,001104707
NP	1	0,066666667	0,016030259	0,001068684

Cet ensemble de résultats indique, comme attendu, une forte activation de la catégorie *N* provenant d'une part du nombre de propriétés potentielles qui l'activent et d'autre part de leur importance (i.e. un poids élevé). Cette forte activation conduit à la constitution d'un chunk [Det, N] qui sera stocké dans un buffer de la mémoire déclarative. Ce processus d'identification de chunk repose donc sur des mécanismes de bas niveau, effectués en temps réel ce qui se manifeste concrètement par un traitement global notamment au niveau du mouvement oculaire dans le cas de la lecture. L'exemple de la figure 5 illustre ce mécanisme. La réalisation de la catégorie *Det* permet d'identifier trois propriétés activant le *N* conduisant à la création du chunk.

L'exemple de la figure 6 décrit le même mécanisme, appliqué ici à la constitution d'un chunk formé, dans le cas d'une relative sujet, par le pronom relatif et le verbe qui suit. Les catégories activées les plus importantes (celles correspondant à des contraintes de plus fort poids) sont *V* et *N*, représentées dans le cadre associé au pronom relatif. La catégorie *V* dispose cependant d'un niveau d'activation très supérieur au *N*. Le *V* étant réalisé immédiatement après l'activation, ceci conduit à la construction du chunk [ProR, V].

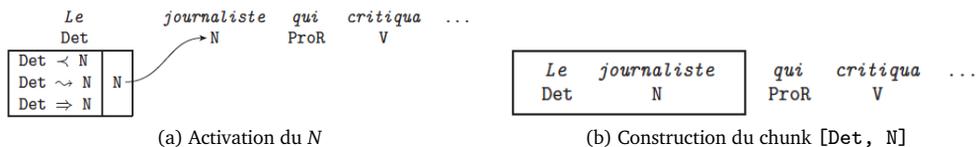


FIGURE 5 – Activation et construction de chunk

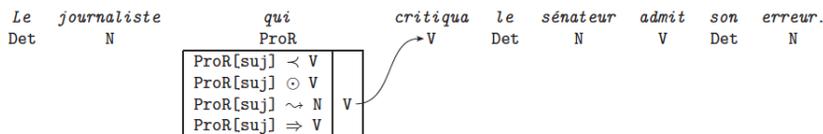


FIGURE 6 – Activation et construction de chunk, suite

Ce processus appliqué à la suite des catégories de la phrase permet de construire la suite de chunks illustrée par la figure 7.

4.2 Les chunks, mécanisme de facilitation

L'hypothèse que nous défendons repose tout d'abord sur l'idée que les chunks sont construits directement, sur la base de mécanismes tirant parti à la fois de critères de fréquence et de densité de relation. Les mécanismes conduisant à la construction de chunks ne sont donc pas les mécanismes classiques de l'analyse syntaxique : le problème posé consiste à mesurer les relations unissant deux catégories adjacentes alors que l'analyse syntaxique consiste à intégrer une catégorie à une structure syntaxique globale. Il s'agit donc de mécanismes de bas niveau, effectués très rapidement.

Une fois construits, ces chunks sont stockés en mémoire et accessibles directement, comme indiqué dans la théorie ACT-R. Notre hypothèse consiste donc à dire que les chunks facilitent le traitement. Leur accès se faisant en bloc, il revient du point de vue cognitif à un accès lexical. De plus, leur intégration se fait également de façon globale. Par conséquent, la présence de chunks dans un énoncé ou une phrase en facilitera le traitement par rapport à d'autres situation où l'intégration devra se faire mot par mot. Autrement dit, une phrase contenant un grand nombre de chunks sera plus facile à traiter qu'une phrase qui en contiendra moins.

Illustrons cette hypothèse en revenant sur le cas des phrases relatives. Les travaux en psycholinguistique (Gibson, 2000), confirmés par plusieurs études expérimentales (Fedorenko *et al.*, 2006), (Demberg et Keller, 2009) ont montré que les relatives objet sont plus difficiles à traiter que les



FIGURE 7 – Construction des chunks pour la phrase complète

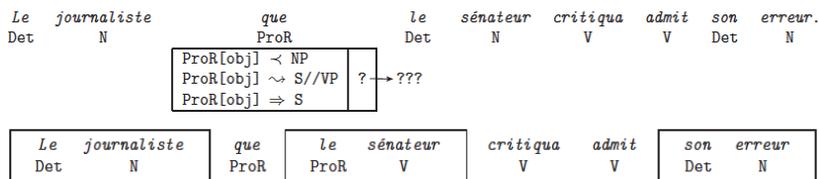


FIGURE 8 – Cas de la relative objet

relatives sujet. Ce phénomène se retrouve au niveau de la construction des chunks. Nous avons vu en effet dans l'exemple de la figure 7 que la relative sujet conduisait à la construction d'un chunk entre le pronom relatif et le verbe. La phrase correspondante contient ainsi 4 chunks au total. La figure 6 illustre ce phénomène par l'impossibilité de construire un chunk contenant le relatif. Celui-ci active bien un certain nombre de catégories, mais aucune d'entre elle ne correspond directement à la catégorie adjacente. Au total, la phrase contenant la relative objet ne contient que 3 chunks. Cet exemple ne prétend bien entendu pas ériger le rôle des chunks en théorie de la difficulté syntaxique comme proposé par (Gibson, 2000). Elle illustre cependant des différences de fonctionnement pouvant accompagner ou compléter ces modèles.

5 Conclusion

Nous avons présenté dans cet article une approche proposant de donner une place centrale à la notion de chunk dans le processus de traitement de la phrase par des sujets humains. Nous utilisons pour cela l'architecture de traitement des processus cognitifs élaborée dans le cadre de la théorie ACT-R. Cette approche précise le rôle joué par les chunks en mémoire. Elle introduit de plus une notion d'activation permettant d'expliquer la rapidité de traitement de ces objets. Appliquée à la question de l'analyse syntaxique (ou du traitement de la phrase si l'on se situe dans une perspective psycholinguistique), cette théorie offre un cadre permettant de décrire la construction et le rôle joué par ces chunks.

En tirant parti d'une description des informations syntaxiques basée sur les contraintes (dans le cadre des *Grammaires de Propriétés*), nous avons proposé une évaluation de la notion d'activation servant de base à la construction des chunks. Il s'agit d'un mécanisme de bas niveau, n'ayant pas recours à l'analyse syntaxique à proprement parler et qui permet la construction d'unités de niveau supra-lexical facilitant le processus car accessibles directement en mémoire. L'utilisation de telles unités correspond à des observations expérimentales, notamment de mouvement oculaire, montrant que les chunks correspondent à des unités de traitement pertinentes.

Il reste à évaluer la validité de l'hypothèse de facilitation des chunks de façon expérimentale. Il s'agira notamment de vérifier que la construction des chunks est un processus de bas niveau et que leur accès correspond à un accès lexical en complétant les observations de mouvement oculaire par des expériences à l'aide de potentiels évoqués et de localisation de source. L'étape suivante consistera à vérifier la facilitation induite par les chunks en termes de temps de traitement.

Remerciements

Ce travail réalisé dans le cadre du Labex BLRI (<http://www.blri.fr>) portant la référence ANR-11-LABX-0036 a bénéficié d'une aide de l'Etat gérée par l'ANR au titre du projet Investissements d'Avenir A*MIDEX portant la référence ANR-11-IDEX-0001-02.

Références

- ABEILLÉ, A., CLÉMENT, L. et TOUSSENEL, F. (2003). Building a treebank for french. In ABEILLÉ, A., éditeur : *Treebanks*, Kluwer, Dordrecht.
- ABNEY, S. (1991). Parsing by chunks. In *Principle-Based Parsing*. Kluwer Academic Publishers, pages 257–278.
- ANDERSON, J. R., BOTHELL, D., BYRNE, M. D., DOUGLASS, S., LEBIERE, C. et QIN, Y. (2004). An integrated theory of the mind. *Psychological Review*, 111(4):1036–1060.
- BIRD, S., KLEIN, E. et LOPER, E. (2009). *Natural Language Processing with Python*. O'Reilly Media.
- BLACHE, P. (2001). *Les Grammaires de Propriétés : Des contraintes pour le traitement automatique des langues naturelles*. Hermès.
- BLACHE, P. (2012). Estimating constraint weights from treebanks. In *Proceedings of CSLP*.
- BLACHE, P. et RAUZY, S. (2011). Predicting linguistic difficulty by means of a morpho-syntactic probabilistic model. In *Proceedings of PACLIC 2011, december 2011*, Singapour.
- DEMBERG, V. et KELLER, F. (2008). Data from eye-tracking corpora as evidence for theories of syntactic processing complexity. In *Cognition*, volume 109, Issue 2, pages 193–210.
- DEMBERG, V. et KELLER, F. (2009). A computational model of prediction in human parsing : Unifying locality and surprisal effects. In *Proceedings of the 31st Annual Conference of the Cognitive Science Society*, pages 1888– 1893.
- DUCHIER, D., PROST, J.-P. et DAO, T.-B.-H. (2009). A model-theoretic framework for grammaticality judgements. In *Conference on Formal Grammar (FG'09)*.
- FEDORENKO, E., GIBSON, E. et ROHDE, D. (2006). The nature of working memory capacity in sentence comprehension : Evidence against domain-specific working memory resources. *Journal of Memory and Language*, 54(4):541–553.
- GIBSON, E. (1998). Linguistic complexity : locality of syntactic dependencies. *Cognition*, 68:1–76.
- GIBSON, E. (2000). The dependency locality theory : A distance-based theory of linguistic complexity. In *Image*. A. Marantz, Y. Miyashita, W. O'Neil (Edts).
- GRODNER, D. J. et GIBSON, E. A. F. (2005). Consequences of the serial nature of linguistic input for sentential complexity. *Cognitive Science*, 29:261–291.
- HALE, J. (2001). A probabilistic earley parser as a psycholinguistic model. In *Proceeding of 2nd Conference of the North American Chapter of the Association for Computational Linguistics*, Pittsburgh, PA.
- HAWKINS, J. (2003). Efficiency and complexity in grammars : Three general principles. In MOORE, J. et POLINSKY, M., éditeurs : *The Nature of Explanation in Linguistic Theory*, pages 95–126. CSLI Publications.

LEWIS, R. L. et VASISHTH, S. (2005). An activation-based model of sentence processing as skilled memory retrieval. *Cognitive Science*, 29:375–419.

RAUZY, S. et 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*.

REITTER, D., KELLER, F. et MOORE, J. D. (2011). A computational cognitive model of syntactic priming. *Cognitive Science*, 35(4):587–637.