

# Résumé public : projet IDIOME

---

## ***Interaction entre complexification et facilitation dans le traitement du langage***

*Etude de la violation de contraintes dans les idiomes*

### **Participants**

Philippe Blache, Sophie Dufour, Chotiga Pattamadilok, Carlos Ramisch, Stéphane Rauzy

### **Présentation**

L'étude des facteurs de complexité du traitement du langage consiste à identifier les paramètres, les phénomènes qui induisent une difficulté accrue. De nombreux travaux ont ainsi proposé des modèles de difficulté, en particulier au niveau de la phrase (Gibson, 2000; Warren & Gibson, 2002; Lewis & Vasishth, 2005). Ces travaux mettent à jour des informations variées, relevant de caractéristiques lexicales et morphosyntaxiques (fréquence, longueur des mots, structure morphologique, etc.), syntaxiques (niveaux de profondeur de la structure, éloignement des dépendances), sémantiques ou encore discursives (identification des référents et leur accessibilité). Parallèlement aux effets de difficulté liés à des constructions particulières, d'autres paramètres peuvent également jouer un rôle important dans l'évaluation de la complexité. Il s'agit en particulier des phénomènes de *violation* : une phrase ou un énoncé violant une règle ou une propriété linguistique induira une complexité de traitement accrue (Sorace & Keller, 2005). Plusieurs approches proposent une estimation quantifiée de la difficulté en exploitant notamment les informations fournies par des analyseurs automatiques. C'est le cas notamment de *l'indice de surprise* mesurant la prédictibilité d'une forme (Hale, 2001) ou encore de la complexité de la structure d'analyse estimée par la profondeur des piles de stockage d'un analyseur (Schuler et al., 2008).

Par ailleurs, plusieurs études ont montré que ces facteurs de difficulté étaient sensibles à un effet de *cumulativité* (Keller, 2006) : le nombre de propriétés violées est corrélé avec le niveau de complexité. Le phénomène de cumulativité négative (dans le sens où les facteurs de difficulté s'additionnent pour rendre le processus de traitement plus complexe) est d'ailleurs dans une certaine mesure pris en compte par les estimations probabilistes issues des analyseurs. Dans le même temps, mais de façon plus limitée, des études ont montré de surcroît l'effet *facilitateur* que certains paramètres peuvent conférer au traitement. Typiquement, les phénomènes de *niveau d'activation* permettent de moduler certains facteurs de complexité : le traitement d'une construction peut être facilité en fonction de son contexte (Vasishth, 2003).

La question qui se pose alors est celle de l'estimation de ***l'interaction existant entre effets de complexification et de facilitation***. Des travaux ont ainsi montré que le niveau de grammaticalité d'une phrase pouvait être décrit et évalué en prenant en compte à la fois les propriétés violées et

satisfaites. Ce niveau de grammaticalité ainsi estimé est corrélé au jugement d'acceptabilité (Blache et al., 2006). Il y aurait donc, parallèlement à une *cumulativité négative*, un effet de *cumulativité positive* qu'il serait possible d'estimer en fonction des caractéristiques linguistiques exprimées en termes de violation et satisfaction de contraintes.

Le but de cette étude est de rechercher une validation expérimentale de ce modèle de complexité. Nous proposons pour cela d'étudier des phrases comportant une construction idiomatique. Ce type de phénomène est en effet connu pour introduire un effet de facilitation puissant (Vespignani et al., 2010; Rommers et al., 2013). Cet effet provient, selon notre modèle, à la fois du nombre et du poids des propriétés satisfaites et non satisfaites. Notre hypothèse est que l'introduction d'une violation syntaxique au sein d'un idiomme se trouve compensée par le niveau élevé de facilitation produit par l'idiome. Pour étudier ce phénomène, nous proposons de comparer l'effet d'une même violation syntaxique sur la phrase comportant un idiomme par rapport à une phrase contrôle de même type ne comportant pas d'idiome. Selon notre hypothèse, la violation syntaxique sera peu ou faiblement détectée, ce qui devrait se traduire par des effets moindres de détection de violation en termes de potentiels évoqués.

## Méthode

Nous proposons de conduire une expérience permettant d'examiner les potentiels évoqués de sujets confrontés à la lecture de phrase contenant des idiomes avec violation syntaxique, par comparaison avec des phrases contrôles.

120 idiomes de type « *avoir une idée derrière la tête* » ont été sélectionnés. Pour chaque idiomme, 4 phrases ont été construites donnant lieu à 4 conditions expérimentales :

- 1) Une phrase idiomatique sans violation telle que « *Paul a une idée derrière la tête depuis ce matin* »
- 2) Une phrase idiomatique avec violation syntaxique introduite après le point de reconnaissance de l'idiome. « *Paul a une idée derrière **le** tête depuis ce matin* »
- 3) Une phrase non idiomatique sans violation ayant la même structure syntaxique que l'idiome : « *Paul a une douleur derrière la nuque depuis ce matin* »
- 4) Une phrase non idiomatique avec violation syntaxique introduite au même endroit que dans l'idiome « *Paul a une douleur derrière **le** nuque depuis ce matin* »

A noter que les violations syntaxiques portent toujours sur un mot de fonction.

Pour s'assurer que les sujets prêtent attention aux phrases, des essais fillers contenant une phrase suivie d'une image ont été ajoutés (10%). A l'apparition des images, les participants doivent décider si l'image correspond à la phrase qui la précède.

Les phrases seront présentées visuellement au rythme d'un mot toutes les 600 ms (ms (le mot est affiché pendant 300 ms suivi par l'écran noir pendant 300 ms). Les réponses électrophysiologiques seront mesurées au moment de la violation et sur le mot contenu qui suit la violation.

Pour chaque point de mesure, 3 composantes seront analysées

- 1) ELAN (100-300ms) : reflète la détection automatique d'une violation syntaxique
- 2) LAN et N400 (300-500 ms) : reflète l'intégration morpho-syntaxique ou lexicosemantique
- 3) P600 (500-1000 ms) : reflète un processus de réparation ou ré-analyse de la phrase suite à la détection de la violation

## *Participants*

30 participants d'origine francophone, droitiers sans déficits visuels, auditifs, langagiers et neurophysiologiques, âges compris entre 18 et 35 ans seront recrutés.

## **Situation par rapport aux objectifs du BLRI**

Ce projet répond aux objectifs du BLRI de plusieurs façons :

- Il s'agit d'un projet **interdisciplinaire**, réunissant des linguistes, des informaticiens et des psycholinguistes de deux laboratoires du BLRI (le LPL et le LIF). L'objectif de ce projet est de valider une hypothèse formulée dans le cadre d'une théorie linguistique selon laquelle la complexité du traitement syntaxique dépend de la quantité et de la qualité des informations disponibles (les contraintes satisfaites ou violées ainsi que leur importance). On utilise pour cela une modélisation computationnelle permettant de quantifier cette information. Une estimation plus fine, adaptée au type de construction particulier utilisé dans cette étude (les idiomes), utilisera les techniques développées par les partenaires du projet.
- Ce projet offre la possibilité d'avancer à partir d'un point de vue théorique et formel en linguistique vers une expérimentation permettant de poser la question de la modélisation du traitement syntaxique en adoptant une perspective jusqu'ici peu explorée du point de vue expérimental. Il répond à l'un des objectifs du BLRI de mettre en place de nouvelles expérimentations, partant de questionnements n'ayant pas encore conduit à des études en direction du fonctionnement cérébral. En d'autres termes, il s'agit d'une problématique provenant de questions posées par des linguistes et des informaticiens, sans que ces questions n'aient fait jusqu'à présent l'objet de recherches en direction des bases cérébrales.
- Ce projet permet d'envisager, à partir du même jeu de données, des investigations variées, par exemple en mouvement oculaire ou en MEG.

## **Demande de soutien auprès du BLRI**

1. Aide à la préparation et la passation de l'expérience EEG
  - a. Convocation des sujets
  - b. Passation des expériences
2. Traitement des données
3. Indemnisation des sujets :  $30 \times 20\text{€} = 600\text{€}$