

Prédire c'est comprendre : un modèle neuro-cognitif du langage fondé sur la prédiction

Philippe Blache

Laboratoire Parole et Langage - CNRS



Des langues au langage

Benoît SAGOT
CHAIRE ANNUELLE INFORMATIQUE
ET SCIENCES NUMÉRIQUES
En partenariat avec Inria

**Apprendre les langues
aux machines**



Qu'est-ce que les machines peuvent-elles nous apprendre sur le langage ?

Langage, compréhension, cognition: la controverse des LLMs

■ Chomsky (NY Times, EduKitchen 2023)

- ▶ *ChatGPT is basically high-tech plagiarism*
- ▶ *The human mind is an efficient system that operates with small amounts of information*
- ▶ *The human mind seeks not to infer brute correlations among data points but to create explanations*
- ▶ *LLMs have absolutely no value with regard to understanding anything about language or cognition*



**Modern language models refute
Chomsky's approach to language**

Steven T. Piantadosi^{a,b}
^aUC Berkeley, Psychology ^bHelen Wills Neuroscience Institute

■ Piantadosi (2023)

- ▶ *There is nothing comparable in all of linguistic theory to the power of large language models in both syntax and semantics*
- ▶ *LLMs allow us to finally develop compelling theories of the interplay of structure and statistics*

■ Les positions

- ▶ Les LLMs sont capables de prédire mais pas d'expliquer
- ▶ Les théories linguistiques sont capables d'expliquer mais pas de prédire

Langage, compréhension, cognition: la controverse des LLMs

■ Chomsky (NY Times, EduKitchen 2023)

- ▶ *ChatGPT is basically high-tech plagiarism*
- ▶ *The human mind is an efficient system that operates with small amounts of information*
- ▶ *The human mind seeks not to infer brute correlations among data points but to create explanations*
- ▶ *LLMs have absolutely no value with regard to understanding anything about language or cognition*



Modern language models refute
Chomsky's approach to language

Steven T. Piantadosi^{1,2}
¹UC Berkeley, Psychology ²Helen Wills Neuroscience Institute

■ Piantadosi (2023)

- ▶ *There is nothing comparable in all of linguistic theory to the power of large language models in both syntax and semantics*
- ▶ *LLMs allow us to finally develop compelling theories of the interplay of structure and statistics*

■ Les positions

- ▶ Les LLMs sont capables de prédire mais pas d'expliquer
- ▶ Les théories linguistiques sont capables d'expliquer mais pas de prédire

Langage, compréhension, cognition: la controverse des LLMs



■ Chomsky (NY Times, EduKitchen 2023)

- ▶ *ChatGPT is basically high-tech plagiarism*
- ▶ *The human mind is an efficient system that operates with small amounts of information*
- ▶ *The human mind seeks not to infer brute correlations among data points but to create explanations*
- ▶ *LLMs have absolutely no value with regard to understanding anything about language or cognition*

Modern language models refute
Chomsky's approach to language

Steven T. Piantadosi^{1,2}
¹UC Berkeley, Psychology ²Helen Wills Neuroscience Institute

■ Piantadosi (2023)

- ▶ *There is nothing comparable in all of linguistic theory to the power of large language models in both syntax and semantics*
- ▶ *LLMs allow us to finally develop compelling theories of the interplay of structure and statistics*

■ Les positions

- ▶ Les LLMs sont capables de prédire mais pas d'expliquer
- ▶ Les théories linguistiques sont capables d'expliquer mais pas de prédire

Définir la compréhension

■ Qu'est-ce que la compréhension ?

- ▶ Capacité à transmettre, échanger des informations
- ▶ Capacité à interpréter, utiliser des informations **explicites** (*compréhension littérale*)
- ▶ Capacité à produire, utiliser des informations **implicites** (*compréhension inférentielle*)

■ Décrire la compréhension

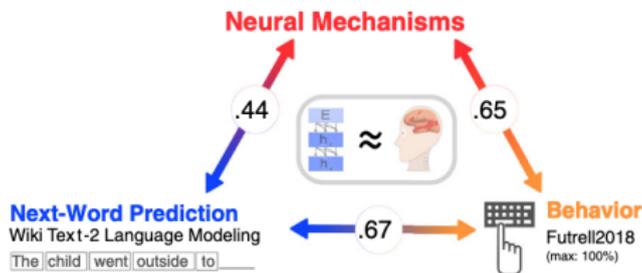
1. Comment représenter le sens ?
2. Comment construire une représentation du sens en conversation ?
3. Comment utiliser cette représentation du sens pour :
 - ▶ *atteindre un but* : résoudre un problème, raisonner, etc.
 - ▶ *construire de nouvelles informations* : inférer, déduire, etc.
 - ▶ *interagir socialement* : émotion, engagement, ToM, etc.
 - ▶ ...



CC BY-SA

Il existe pourtant des similarités : corrélation LLMs / signal cérébral

- ▶ Corrélation entre les prédictions du modèle et les enregistrements cérébraux
- ▶ Les modèles ayant le meilleur résultat de prédiction du mot suivant sont aussi ceux qui prédisent le mieux l'activité cérébrale



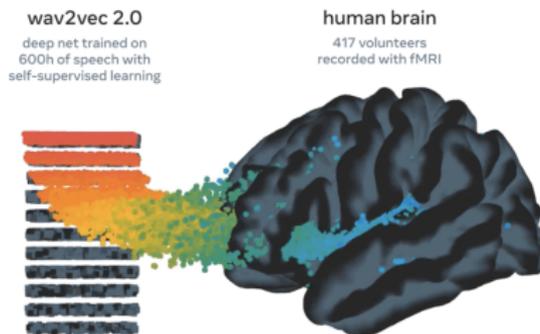
PNAS, CC BY-NC-ND

Schirmpf M. et al. (2021) "The neural architecture of language: Integrative modeling converges on predictive processing", PNAS, 118(45)

LLMs et cerveau : architecture

■ Peut-on trouver un fonctionnement comparable ?

- ▶ Les LLMs et le cerveau sont organisés de façon hiérarchique
- ▶ Le cerveau fonctionne de façon comparable aux LLMs pour prédire le prochain événement
- ▶ Il existe une application linéaire (*mapping*) entre l'activation des modèles et les réponses cérébrales à la parole



©J.R. King, autorisation de l'auteur

Caucheteux, Gramfort, & King (2023) "Evidence of a predictive coding hierarchy in the human brain listening to speech", Nature Hum. Behav., 7

Point intermédiaire

■ Les points essentiels de la compréhension

- ▶ Rôle central de la **prédiction**
- ▶ Le contexte implicite joue aussi un rôle essentiel : **activation** d'information
- ▶ Les interlocuteurs convergent pendant une interaction : **alignement**

■ Suite de la présentation

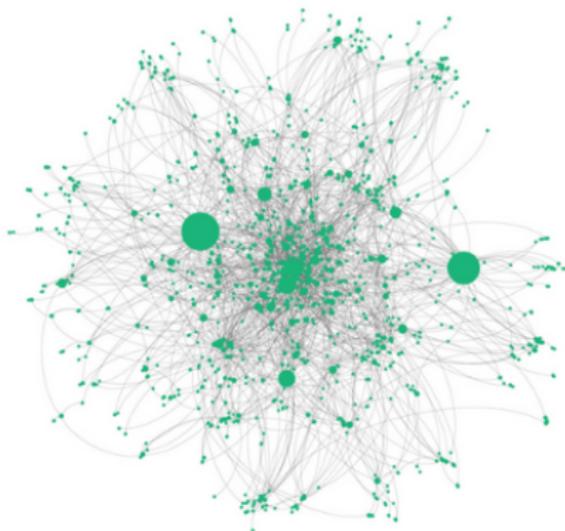
- ▶ Comment représenter le sens ?
- ▶ La prédiction
- ▶ L'activation
- ▶ Le modèle

Part I

Représentation du sens

Quelles sont les unités de sens ?

- ▶ La représentation de l'information sémantique se fonde sur des **entités** et des **relations**

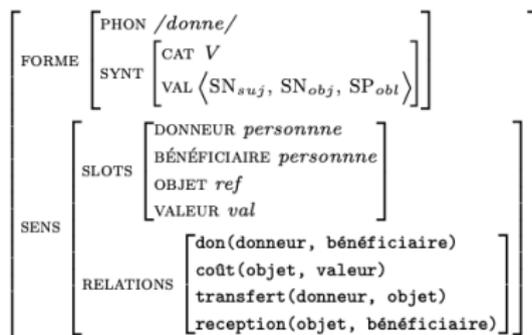


Graphes de connaissance
(Lynn et al., PNAS 117, CC BY-NC-ND)

Unités de base : les signes

► Signes : paires forme-sens

- **Forme** : représentation des caractéristiques linguistiques (*phonétique, syntaxe, gestes, etc.*)
- **Sens** : frames
 - *Slots* (éléments d'information)
 - *Relations* (entre les slots)



Signe associé au verbe "donner"

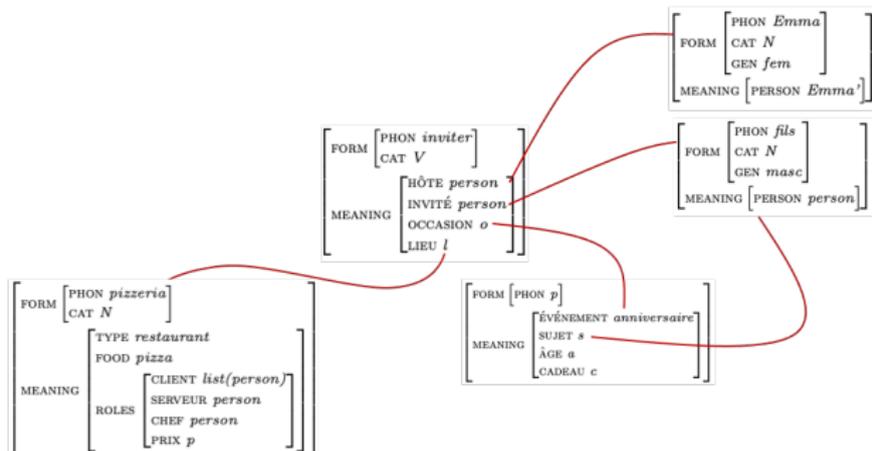
Sag, Ivan (2012) "*Sign-Based Construction Grammar: An Informal Synopsis*", CSLI Publications

Modèle de Situation : un graphe de signes

■ Représentation formelle du sens en contexte

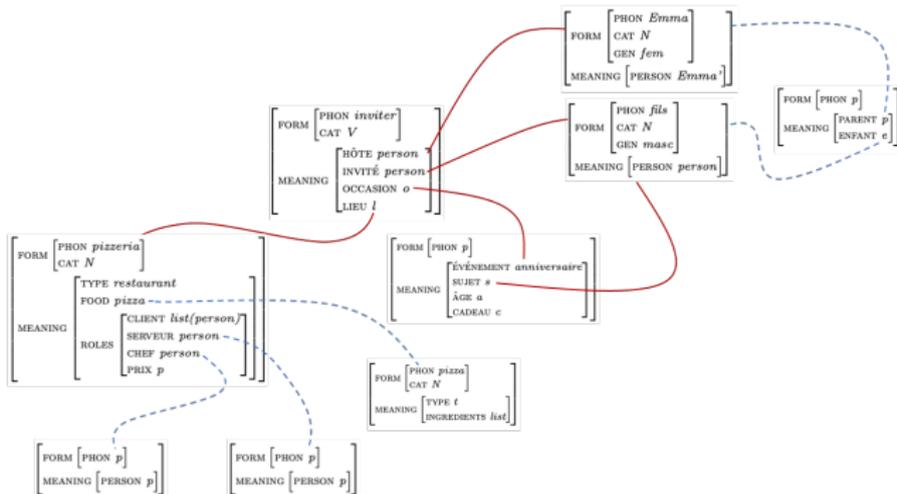
- ▶ Les **nœuds** du graphe sont des signes
- ▶ Les **arcs** du graphe sont les relations sémantiques entre signes

■ Exemple : “Emma a invité son fils à la pizzeria pour son anniversaire.”



Modèle de Situation : inférence

- ▶ Informations implicites
 - ▶ Unités d'information participant à un signe indirectement (e.g. *pizza/pizzeria*)
 - ▶ Déduites de la connaissance du monde (*filz/parentalité*)
- ▶ Les informations implicites sont **activées**



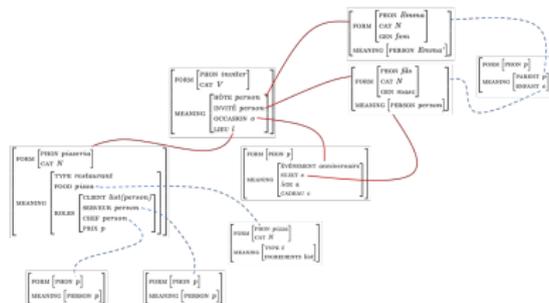
Niveau d'activation des nœuds

► Activation

- On intègre au *modèle de situation* tous les signes activés
- Les signes activés ne seront pas nécessairement réalisés, mais sont disponibles pour la compréhension

► Niveau d'activation

- Tous les éléments (signes, slots) ont un niveau d'activation
- Le niveau d'activation dépend du contexte
 - Nombre et le poids des relations entrantes sur le nœud
 - Historique (activation plus ou moins récente)
- Les signes les plus activés sont **prioritairement utilisés**

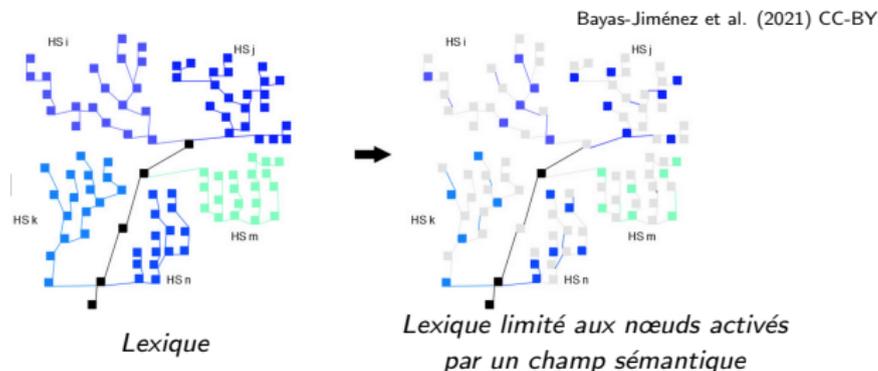


Mécanisme de l'activation : réduction de l'espace de recherche

■ Espace de recherche et graphe de connaissance

- ▶ Certains nœuds du graphe peuvent être plus **activés** que d'autres en fonction du contexte
- ▶ Limitation de la recherche aux nœuds les plus activés

■ Réduction de l'espace de recherche lexical (mémoire à long terme)

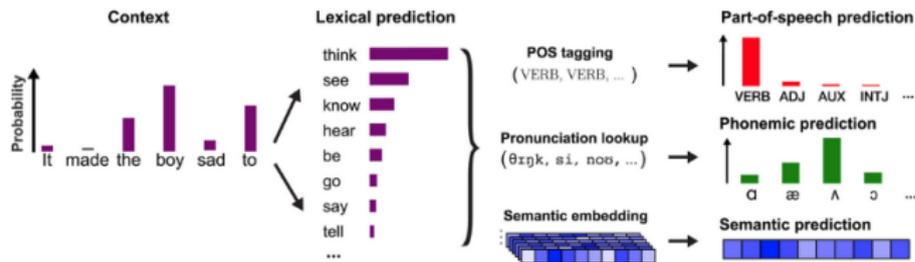


Part II

Prédiction

Les LLMs sont optimisés pour prédire le prochain événement linguistique

- ▶ Prédire un mot associé à des informations



Heilbron M. et al. (2022), PNAS, CC-BY-4.0

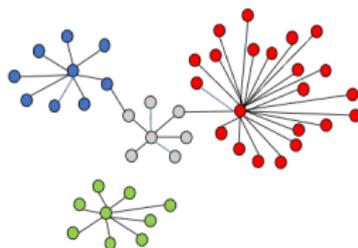
- ▶ Prédire des informations moins spécifiées

| Contexte | Type de prédiction | Caractéristiques prédites |
|---------------------|-----------------------|------------------------------------|
| je pense la ... | <i>morpho-syntaxe</i> | <i>[N/Adj, féminin, singulier]</i> |
| sur l' ... | <i>ms, phonologie</i> | <i>[N/Adj, voyelle initiale]</i> |
| je bois dans un ... | <i>ms, sémantique</i> | <i>[N, masc, sing, +contenant]</i> |

Prédiction de structures complexes

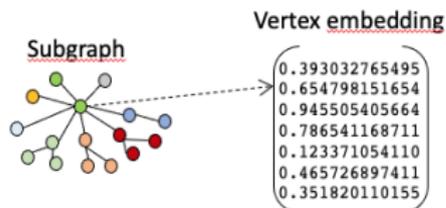
■ Graphes de connaissance

- ▶ Sous-graphes : sous-ensemble d'information

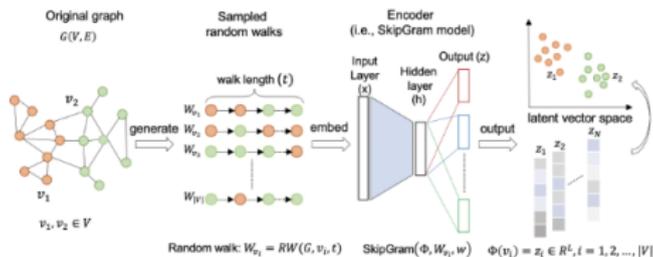


■ Knowledge Graph Embedding

- ▶ Encodage distributionnel



■ Knowledge Graph Reasoning



Xu M., 2020, arXiv2012.08019, CC-BY-4.0

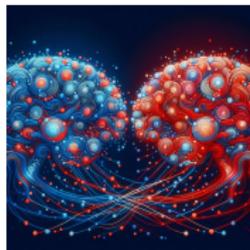
Comment fonctionne la prédiction pendant une conversation

■ *Prediction-by-production*

- ▶ Le locuteur prédit (prépare) la production du mot suivant
- ▶ L'interlocuteur prédit la production du locuteur

Comprehenders:

- *predict using their production system*
- *run the same system for production and comprehension but stop the production just before*
- *imitate the linguistic form of the speaker's utterance (alignment)*

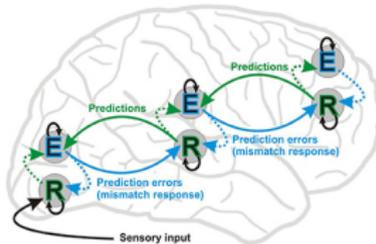


Généré par Bing

Pickering M. et al. (2018), "Predicting while comprehending language: A theory and review", *Psy.Bul* 144:10

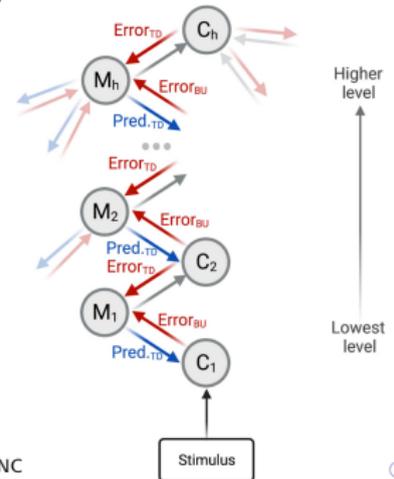
Au niveau cérébral : *predictive coding*

- ▶ Le cerveau prédit en permanence les futurs inputs sensoriels
- ▶ Flux de l'information hiérarchique : des inputs sensoriels (bas niveau) aux informations abstraites (haut niveau)



Stefanics G. et al., 2014, CC-BY-4.0

- ▶ **Prédictions descendantes** : chaque niveau prédit les réponses du niveau inférieur (*feedback connections*)
- ▶ **Erreur de prédiction** : mesure des écarts entre la prédiction et le signal effectif
- ▶ **Propagation ascendante** des erreurs de prédiction (*feedforward connections*)



Ryskin R. et al., 2020, CC-BY-NC

Comment mesurer l'erreur de prédiction : l'unification

■ Unification: évaluer la compatibilité de structures complexes

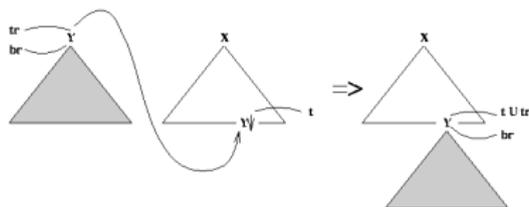
- ▶ Mécanisme : comparaison de deux structures
- ▶ Succès si elles sont compatibles
- ▶ Résultat : fusion des 2 structures

■ Exemple: unification de matrices

$$\text{ACCORD} \begin{bmatrix} \text{PERS } 1 \\ \text{NB } plu \end{bmatrix} \sqcup \text{ACCORD} \begin{bmatrix} \text{PERS } x \\ \text{NB } plu \end{bmatrix} \mapsto \text{ACCORD} \begin{bmatrix} \text{PERS } 1 \\ \text{NB } plu \end{bmatrix} / \{x=1\}$$

$$\text{ACCORD} \begin{bmatrix} \text{PERS } 1 \\ \text{NB } plu \end{bmatrix} \sqcup \text{ACCORD} \begin{bmatrix} \text{PERS } x \\ \text{NB } sing \end{bmatrix} \mapsto \perp$$

$$\text{ACCORD} \begin{bmatrix} \text{PERS } 1 \\ \text{GEN } masc \end{bmatrix} \sqcup \text{ACCORD} \begin{bmatrix} \text{PERS } x \\ \text{NB } plu \end{bmatrix} \mapsto \text{ACCORD} \begin{bmatrix} \text{PERS } 1 \\ \text{GEN } masc \\ \text{NB } plu \end{bmatrix} / \{x=1\}$$



Quantifier l'erreur de prédiction

- ▶ Comparaison entre ce qui est **prédit** et ce qui est **produit**

Contexte : L'étudiant lit le ...

Prédiction

$$\left[\begin{array}{l} \text{PHON} \left[\begin{array}{l} \text{FORM } x \\ \text{SYLL} \mid \text{ONSET } \textit{consonne} \end{array} \right] \\ \text{SYNT} \mid \text{CAT } N \\ \text{SEM} \left[\begin{array}{l} \text{FEATURE } +\textit{objet_lisible} \\ \text{ROLE } \textit{patient} \end{array} \right] \end{array} \right]$$
$$\left[\begin{array}{l} \text{PHON} \left[\begin{array}{l} \text{FORM } x \\ \text{SYLL} \mid \text{ONSET } \textit{consonant} \end{array} \right] \\ \text{SYNT} \mid \text{CAT } N \\ \text{SEM} \left[\begin{array}{l} \text{FEATURE } +\textit{objet_lisible} \\ \text{ROLE } \textit{patient} \end{array} \right] \end{array} \right]$$
$$\left[\begin{array}{l} \text{PHON} \left[\begin{array}{l} \text{FORM } x \\ \text{SYLL} \mid \text{ONSET } \textit{consonne} \end{array} \right] \\ \text{SYNT} \mid \text{CAT } N \\ \text{SEM} \left[\begin{array}{l} \text{FEATURE } +\textit{objet_lisible} \\ \text{ROLE } \textit{patient} \end{array} \right] \end{array} \right]$$

Production

$$\left[\begin{array}{l} \text{PHON} \left[\begin{array}{l} \text{FORM } / \textit{livre} / \\ \text{SYLL} \mid \text{ONSET } \textit{consonne} \end{array} \right] \\ \text{SYNT} \mid \text{CAT } N \\ \text{SEM} \left[\begin{array}{l} \text{FEATURE } +\textit{objet_lisible} \\ \text{ROLE } y \end{array} \right] \end{array} \right]$$
$$\left[\begin{array}{l} \text{PHON} \left[\begin{array}{l} \text{FORM } / \textit{étagère} / \\ \text{SYLL} \mid \text{ONSET } \textit{voyelle} \end{array} \right] \\ \text{SYNT} \mid \text{CAT } N \\ \text{SEM} \left[\text{FEATURE } -\textit{objet_lisible} \right] \end{array} \right]$$
$$\left[\begin{array}{l} \text{PHON} \left[\begin{array}{l} \text{FORM } / \textit{article} / \\ \text{SYLL} \mid \text{ONSET } \textit{voyelle} \end{array} \right] \\ \text{SYNT} \mid \text{CAT } N \\ \text{SEM} \left[\begin{array}{l} \text{FEATURE } +\textit{objet_lisible} \\ \text{ROLE } y \end{array} \right] \end{array} \right]$$

Unification

$$\left[\begin{array}{l} \text{PHON} \left[\begin{array}{l} \text{FORM } / \textit{livre} / \\ \text{SYLL} \mid \text{ONSET } \textit{consonne} \end{array} \right] \\ \text{SYNT} \mid \text{CAT } N \\ \text{SEM} \left[\begin{array}{l} \text{FEATURE } +\textit{objet_lisible} \\ \text{ROLE } \textit{patient} \end{array} \right] \end{array} \right]$$

→

→

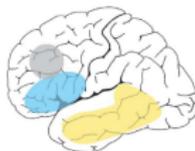
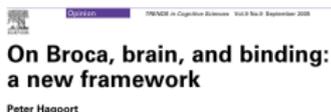
→

⊥

Part III

MUC+ : un modèle neuro-cognitif intégrant prédiction et activation

Memory-Unification-Control : un modèle neuro-cognitif



■ Trois composants du traitement du langage

▶ Memory

- ▶ *"The brain tries to do as much processing as possible with what is stored in memory"*
- ▶ **Unités de base** : objets complexes, en mémoire à long terme
- ▶ Dans notre approche : les signes

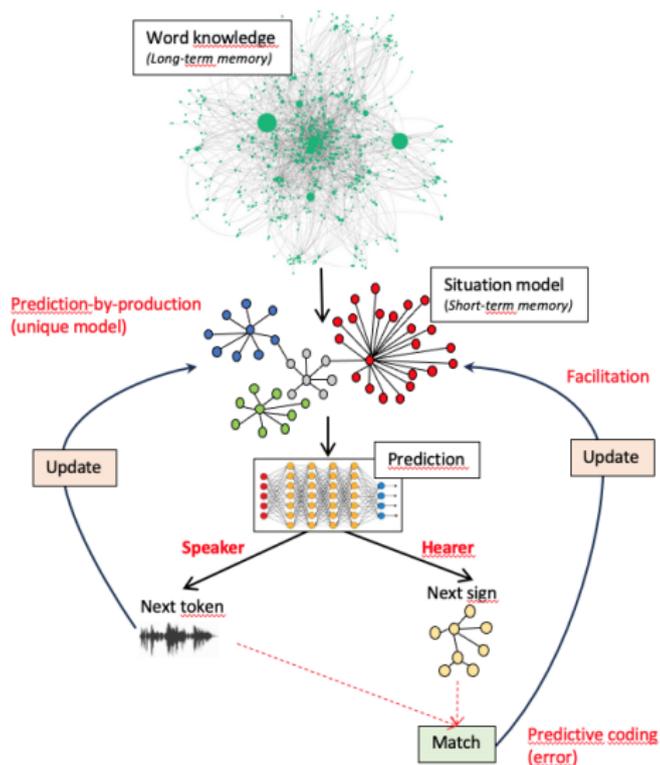
▶ Unification

- ▶ Assemblage des structures
- ▶ Dans notre approche : assemblage + calcul de l'erreur de prédiction

▶ Control

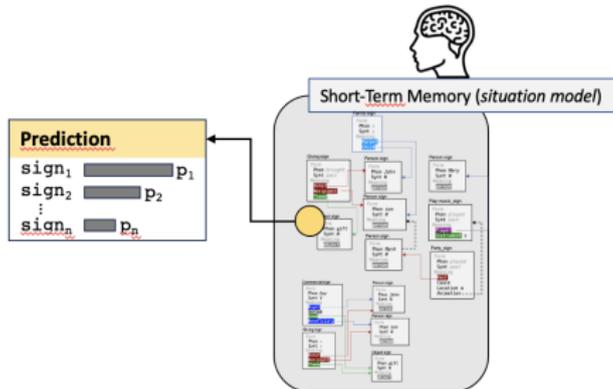
- ▶ Intègre les informations pragmatiques, liées à la connaissance du monde
- ▶ Dans notre approche : activation

Schéma général du modèle



Composant #1 Prédiction à partir du *Modèle de Situation*

- ▶ **Contexte**: formé par l'ensemble des informations du *Modèle de Situation*
- ▶ Deux types d'informations :
 - Explicites : ensemble des signes effectivement *produits* par le locuteur
 - Implicites : ensemble d'informations activées (inférées)
- ▶ La prédiction retourne une **liste de signes possibles**
 - Prédiction de signes de granularité variable
 - Liste probabilisée



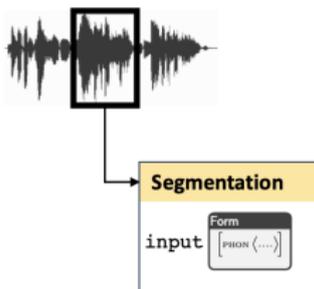
Composant #2 Extraction du stimulus (à partir de l'input audio)

■ Approche classique

- ▶ Décodage acoustico-phonétique + accès lexical

■ Segmentation en mots fondée sur les **indices acoustiques**

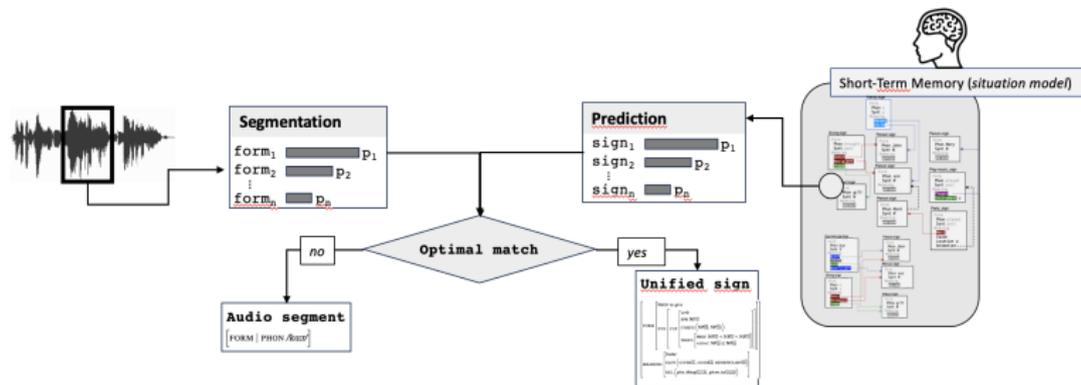
- ▶ Indices *prosodiques* (pauses, allongements, changement d'intensité)
- ▶ Informations *phonotactiques* (probabilités transitionnelles entre phonèmes, syllabes)



■ Segmentation en mots sans accès au lexique

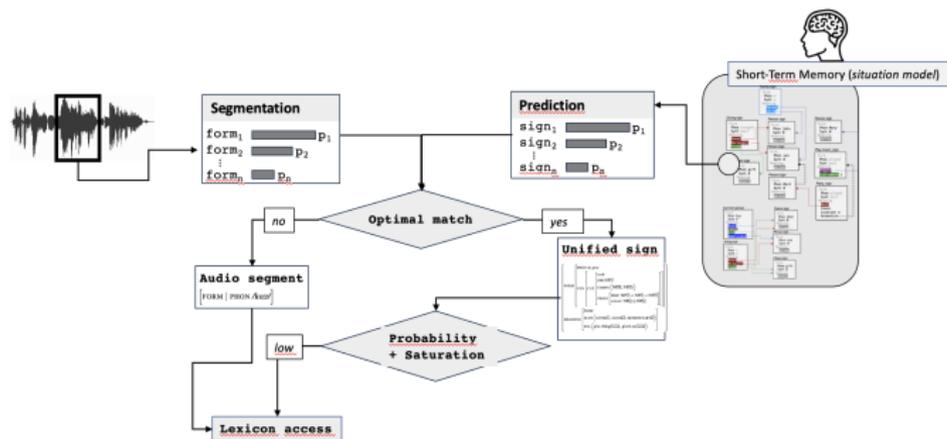
- ▶ Segmentation plutôt que reconnaissance de mot
- ▶ La segmentation retourne une *forme phonétique* (sans autre information)

Composant #3 Vérification de la prédiction par unification



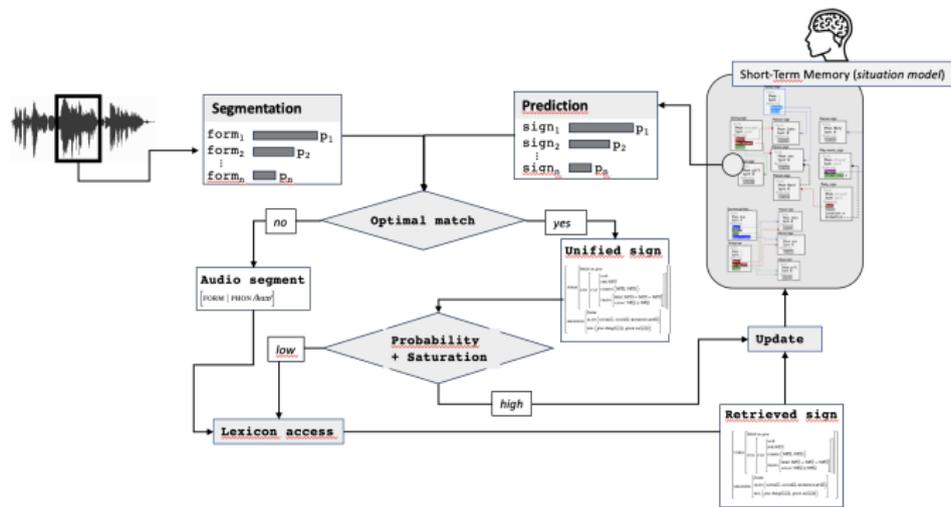
- ▶ Si l'unification entre [pred] et [input] **réussit**
 - Le signe prédit [pred] correspond au stimulus produit [input]
 - Le résultat de l'unification est un signe complétant [pred] avec les informations (phonétiques) de [input]
- ▶ Si l'unification entre [pred] et [input] **échoue**
 - Le signe suivant de la liste de prédiction est choisi
 - Si aucun signe ne convient, [pred] est laissé libre : $[pred] \sqcup [input] = [input]$

Composant #4 Accès au lexique facultatif



- ▶ Appelons **[unif]** le signe résultant de $[pred] \sqcup [input]$
- ▶ Contenu informationnel de **[unif]**
 - Elevé si forte probabilité et contenu détaillé
 - Si contenu informationnel **élevé** : mise à jour directe du *Modèle de Situation*
 - Si contenu informationnel **faible** : accès au lexique puis mise à jour

Composant #5 Mise à jour du *Modèle de Situation*



■ Rappel: contenu du *Modèle de Situation*

- Graphe de signes
- Chaque signe contient un ensemble de 'slots'

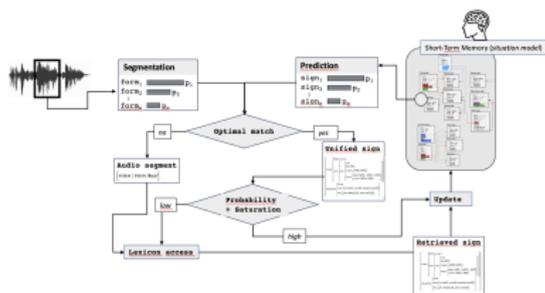
■ Mise à jour = 'slot filling'

- Trouver le slot avec la plus haute **activation** et la meilleure **unification**

Mise à jour du *Modèle de Situation*

Etape 1 A partir du slot cible le plus activé

- ▶ **Cas 1:** [unif] \sqcup [cible]
 - [unif] devient la nouvelle valeur du 'slot cible'
- ▶ **Cas 2:** [unif] $\not\sqcup$ [cible]
 - Evaluation de l'**erreur d'unification**
 - Si **erreur faible** : unification relâchée
 - Recherche d'une réparation
 - Mise à jour est effectuée avec [unif]
 - Si **erreur forte**
 - [unif] est ajouté au *Modèle de Situation* sans connexion au graphe



Etape 2 Inférer les informations implicites

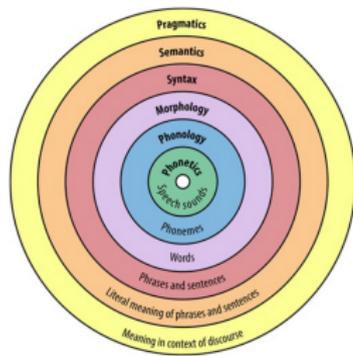
- ▶ Les niveaux d'activation de tous les signes du modèle de situation sont réévalués
- ▶ Les nouveaux signes sont activés

Part IV

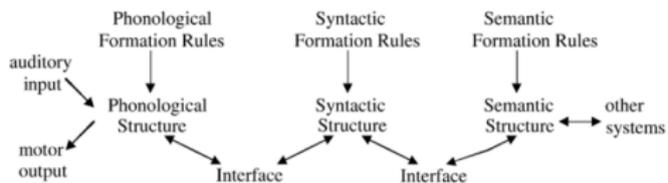
Conclusion : apport du modèle MUC+

Organisations classiques du traitement du langage

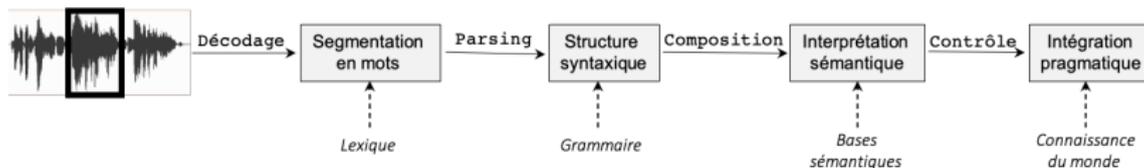
■ Organisation hiérarchique



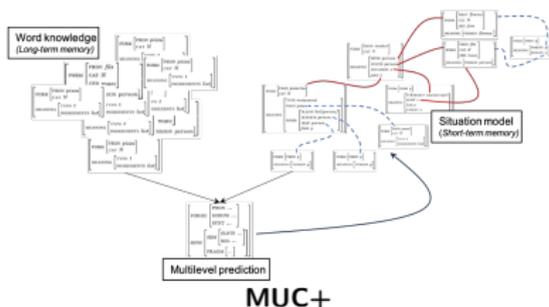
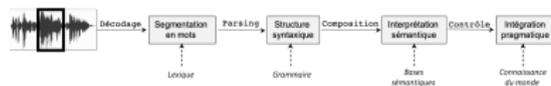
■ Organisation parallèle



■ Organisation séquentielle



MUC+ : un modèle dynamique du langage



Composants

Théories classiques

MUC+

Contexte

- Les unités et les mécanismes sont décrits indépendamment du contexte

- En langue, on a toujours un contexte

Unités de base

- Ensemble de catégories liées à un domaine (lexique, LTM)
- Décrivant un objet (plutôt qu'un phénomène)

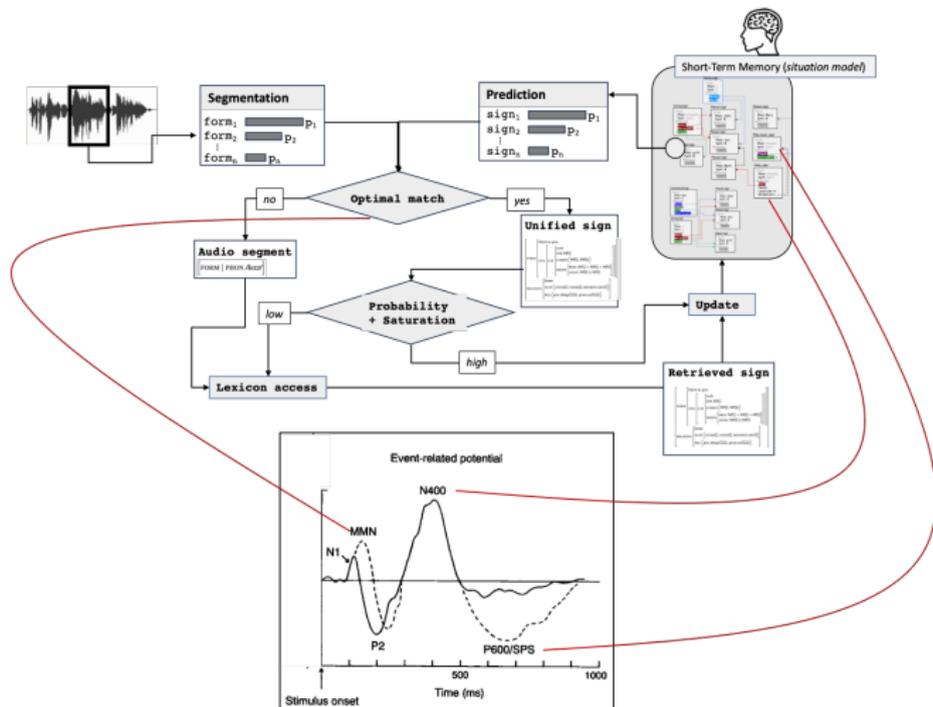
- Structure créée dynamiquement, par prédiction
- A partir du contexte et des catégories (LTM + STM)
- Multimodale, granularité variable

Mécanisme d'analyse

- Assemblage de catégories pour construire une représentation par niveau
- Interaction entre niveaux : interaction entre structures (late fusion)

- Assemblage de structures par unification
- Interaction entre niveaux : dans les signes

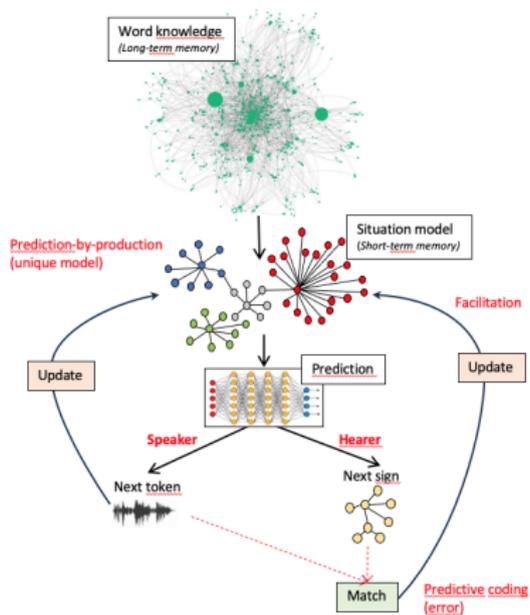
MUC+ et cerveau



Conclusion

- ▶ La **prédiction** joue un rôle central
 - ▶ Construction des structures dynamiques (unités de base du traitement)
 - ▶ Intégration du contexte dans le traitement du langage
- ▶ **Unités dynamiques**, construites par prédiction
 - ▶ Fondées sur le contexte
 - ▶ De granularité variable
 - ▶ Multimodales, intégrant tous les domaines
- ▶ La **facilitation** ne vient pas de la prédiction à proprement parler mais de la richesse de la structure prédite
- ▶ Un mécanisme central : **l'unification**
 - ▶ Mise à jour
 - ▶ Contrôle de l'accès en mémoire
 - ▶ Calcul de l'erreur de prédiction
- ▶ **L'activation** est distincte de la prédiction : contrôle, réduction de l'espace de recherche
- ▶ Le compréhension (i.e. construction du modèle de situation) repose sur la prédiction et l'activation
- ▶ Même modèle pour la **production et la perception**

Merci pour votre attention



Part V

ANNEXES

MUC+ : au niveau cérébral

latence du signal cérébral ERP : non pas par séquence/hiérarchie modulaire de traitement mais par coût/complexité des traitements

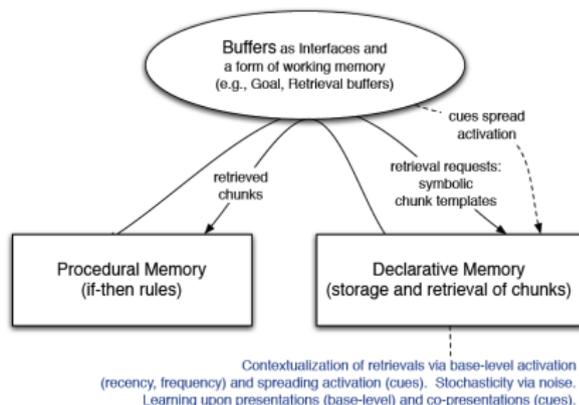
Comment déterminer l'activation : théorie *ACT-R* (Adaptive control of thought-rational)

John Anderson et al. (2004), An Integrated Theory of the Mind, *Psychological Review*, 111:4

Connaissance: unités d'information complexes (*chunks*), stockées en mémoire à long terme, et pouvant être activées selon le contexte

Organisation:

- ▶ **Mémoire déclarative:** contient les **chunks**
- ▶ **Mémoire procédurale:** règles de production
- ▶ **Niveau d'activation** des chunks
 - L'activation est fonction de la *fréquence d'usage* et du *délat*
 - Le chunk avec la plus haute activation est accédé prioritairement



ACT-R: mesure de l'activation

- ▶ **Activation de base (B):** fréquence et historique de l'accès au chunk
- ▶ **Activation de propagation (S_i):**
 - ▶ L'activation dépend de :
 - ▶ La puissance de la relation S_{ci} entre un indice c et un item i
 - ▶ Le poids de l'indice W_c
 - ▶ Un chunk reçoit une activation de tous les indices c qui lui sont reliés
 - ▶ Activation de propagation: $S_i = \sum_c W_c S_{ci}$
- ▶ **Activation globale (A_i):** $A_i = B_i + S_i + \epsilon_i$

Lewis, R. & Vasishth, S. (2005) An activation-based model of sentence processing as skilled memory retrieval, *Cognitive Science*, 29(3)