

Les modèles de parsing de la phrase

Activité

L'objectif de l'exercice est de réviser la matière, et de consolider vos connaissances. Effectuez le test par groupes de 3 ou 4. Répondez aux questions de manière précise et synthétique. Faites attention à gérer votre temps (vous pouvez voir où vous vous situez dans l'exercice en haut à gauche). Si vous ne trouvez pas la réponse, ne restez pas coincés dessus : encodez une lettre dans l'espace de réponse et cliquez sur « Vérifier » pour obtenir la réponse. Vos notes ne sont pas enregistrées !

Q1. Considérez la phrase suivante:

(1) La femme a vu l'homme *avec une longue-vue*

- a. Présentez les 2 structures en constituants qui peuvent la sous-tendre.
- b. L'ambiguïté de cette phrase est-elle globale ou temporaire ?
- c. Les recherches montrent que le parser préfère attacher le groupe prépositionnel *avec une longue-vue* au verbe. Cette observation a été interprétée comme reflétant le fait que le parser choisit un arbre contenant le moins de nœuds possibles. Quel est le principe sous-jacent à cette préférence selon la théorie du garden-path?
- d. On sait toutefois que le parser est également sensible à toutes sortes d'informations, y compris des informations sémantiques et distributionnelles/fréquentielles. Quelle variable faudrait-il contrôler dans ce matériel afin de pouvoir conclure que la préférence d'attachement au verbe est bien due au à ce principe d'attachement minimal et non à une influence sémantique sur le parser ?

Réponses :

- a. [La femme [a vu [l'homme [avec [une longue-vue]]]]]
[La femme [a vu [l'homme] [avec [une longue-vue]]]]
- b. Globale
- c. Le principe d'attachement minimal (on voit le nombre de nœuds dans les parenthèses !).
- d. Il faudrait contrôler que la fréquence ainsi que la force de l'association sémantique entre 'voir' et 'avec une longue-vue' est la même qu'entre 'homme' et 'avec une longue-vue'. Si le premier lien est plus fréquent/fort sémantiquement que le second, alors la préférence observée pourrait s'expliquer par ces autres facteurs plutôt que par le moindre nombre de nœuds (car tous ces facteurs co-varient).

Q2. Considérez les phrases suivantes:

- (1) Pendant qu'Anna lisait le livre le chien aboyait dehors.
- (2) Pendant qu'Anna lisait le livre était retiré du commerce.

- a. Quelle est la région ambiguë dans la phrase (1)?
- b. Quelle est la région ambiguë dans la phrase (2)?

- c. Selon la théorie du garden-path, laquelle de ces phrases devrait générer un effet garden-path, et pourquoi ?
- d. L'ambiguïté de cette phrase est-elle globale ou temporaire ?
- e. Quelle est la région désambiguïsante dans cette structure ?

Réponses :

- a. Le livre
- b. Le livre
- c. La (2), en vertu du principe de Late closure selon lequel le parser attache un nouvel élément de la phrase à un constituant qui est en cours de construction plutôt que de créer un nouveau constituant (il tend donc à attacher 'le livre' à 'lisait')
- d. Temporaire
- e. *était* (et possiblement le ou les mots qui suivent) ; c'est à ce moment que se produit l'effet garden-path (le ralentissement lié au fait que le parser doit réviser son analyse et donc 'retourner en arrière' sur son chemin)

Q3. Considérez les phrases suivantes:

- (1) Pendant qu'Anna lisait le livre était retiré du commerce.
- (2) Pendant qu'Anna lisait le lait était retiré du commerce.

- a. Selon la théorie modulaire, quelle(s) phrase devrai(en)t générer un effet garden-path, et pourquoi ?
- b. Que prédit le modèle de satisfaction de contraintes quant à la difficulté de chacune de ces deux phrases ?

Réponses :

- a. Les deux phrases, car selon cette théorie les contraintes sémantiques de sélection du verbe ne sont pas prises en compte lors de l'analyse initiale qui opère strictement sur une base syntaxique. Dans les deux cas, le parser va tenter d'attacher le NP au verbe 'lire' qui est transitif.
- b. Il prédit un effet garden-path uniquement dans la phrase (1), car les contraintes sémantiques de sélection sont directement prises en compte et informent le parser que 'le lait' n'est pas un objet plausible du verbe.

Q4. Une étude a comparé les temps de lecture dans 4 conditions expérimentales, illustrées dans le tableau ci-dessous. Considérez les exemples attentivement.

- a. Quelles sont les conditions structurellement ambiguës (en supposant que le parser n'utilise pas d'information sémantique pour désambiguïser la phrase) ?
- b. Quelle est la région ambiguë dans ces phrases et sur quoi repose l'ambiguïté ?
- c. Quelle est la région désambiguïsante ?
- d. Que prédit le modèle modulaire quant à l'effet du caractère animé ou non du sujet de la phrase (*defendant* vs. *evidence*) ?
- e. Les observations valident le modèle de satisfaction de contraintes. Que prédisait ce modèle quant à l'effet du caractère animé ou non du sujet de la phrase ?

Noun type	Clause type	Example
Animate	Reduced	The defendant examined by the lawyer turned out to be unreliable.
	Unreduced	The defendant that was examined by the lawyer turned out to be unreliable.
Inanimate	Reduced	The evidence examined by the lawyer turned out to be unreliable.
	Unreduced	The evidence that was examined by the lawyer turned out to be unreliable.

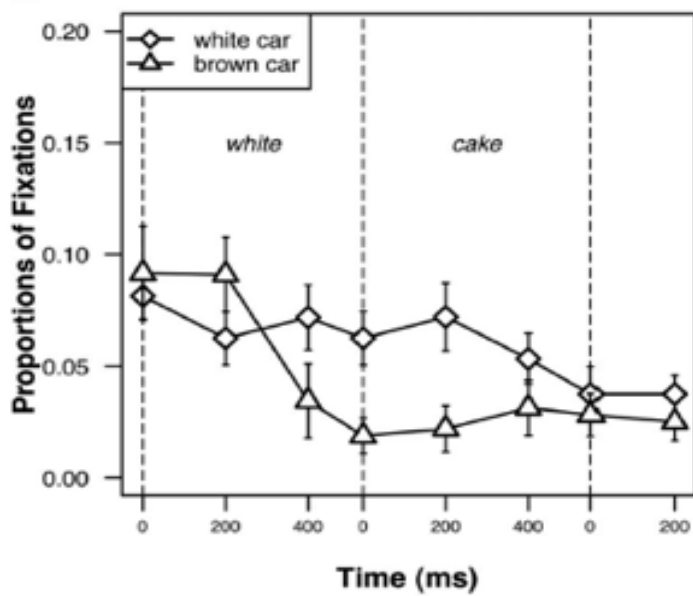
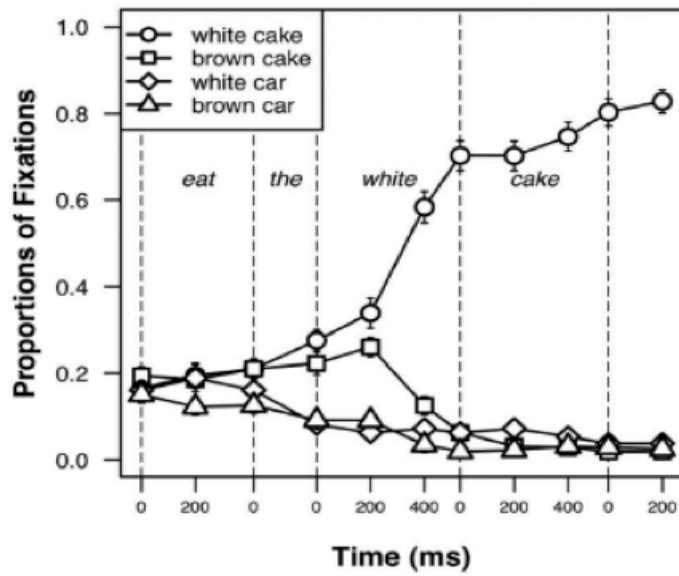
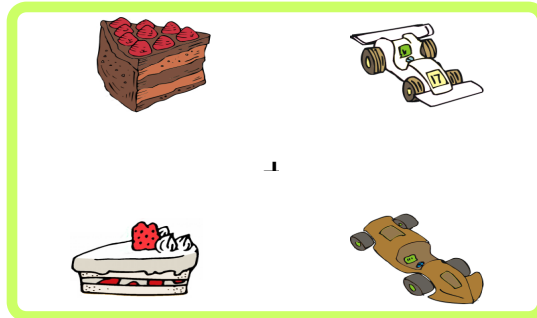
Réponses :

- Les conditions reduced relative, sans *that*
- Examined*. L'ambiguïté repose sur le fait que *examined* peut être soit un verbe principal (auquel cas le sujet est son agent) soit un participe passé (auquel cas le sujet est son patient ; il s'agit d'une structure passive)
- By the lawyer
- Aucun effet puisque la sémantique n'influence pas le choix initial du parser : un effet garden-path est attendu tant avec les sujets animés qu'avec les sujets inanimés.
- Le modèle de satisfaction de contraintes prédit que l'effet garden-path (càd le ralentissement dans la condition 'reduced' comparée à la condition 'unreduced') ne devrait s'observer que lorsque le sujet est animé car lorsqu'il est inanimé il n'est pas un sujet sémantiquement plausible du verbe *examined* et donc le parser ne va pas s'engager dans l'analyse incorrecte.

Q5. Dans le module préparatoire à ce cours, nous avons vu que les deux significations d'un mot ambigu sont initialement activées, même si l'une des deux n'est pas pertinente dans le contexte (travaux de Swinney, dia 12). Cette observation va dans le sens du modèle modulaire algorithmique selon lequel l'activation initiale est guidée exclusivement par les informations portées par le mot lui-même, aveugle aux informations du contexte. Dans une étude, Kukona et collaborateurs (2014) ont étudié le suivi du regard de participants face à un écran représentant 4 objets pendant qu'ils entendent une phrase, comme dans la figure ci-dessous. Les participants entendent des phrases comme « The boy will eat the white cake ». Le suivi du regard est mesuré au moyen d'un eye-tracker.

- Les temps de regard (fixations) pour chacun des 4 objets en fonction du déroulement temporel de la phrase sont illustrés dans le graphe ci-dessous. Qu'observe-t-on juste après l'apparition du mot 'eat' ?
- Quel(s) modèle(s) prédisent cette observation ?
- Examinez à présent le graphe ci-dessous, qui est un zoom des temps de regard vers les deux voitures aux niveaux de 'white' puis de 'cake'. Qu'observez-vous ?
- Quel modèle prédit des temps de regard plus longs vers la voiture blanche que vers la voiture brune ?

e. En quoi ce résultat s'apparente-t-il à celui de Swinney?



Réponses :

- a. Juste après le mot 'eat', le regard augmente vers les deux objets mangeables (cake blanc et cake brun), c'est-à-dire vers les deux entités respectant les contraintes sémantiques pesant sur l'objet du verbe 'manger' (on ne peut pas manger une voiture). Le système prend donc rapidement en compte les informations sémantiques pour anticiper ce qui va arriver ensuite : les regards se tournent vers les cakes avant même d'entendre le mot 'cake'.
- b. Les modèles rationnel (satisfaction de contraintes) et heuristique (good enough)
- c. Les participants regardent plus vers la voiture blanche que vers la voiture brune.
- d. Le modèle modulaire (algorithmique)
- e. Dans les deux cas, le système active une représentation qui pourtant n'est pas pertinente dans le contexte, comme le prédit le modèle modulaire.