

Génération symbolique

Une approche basique pour la génération de texte est d'utiliser la manipulation d'une grammaire et d'un lexique.

Grammaire et lexique. La grammaire donne des schémas à suivre pour construire une phrase. Chaque ligne correspond à une règle, écrite sous la forme « type 1 = type 2 + type 3 ». Les noms de types peuvent contenir des espaces. La type 1 de la première règle doit être « phrase ». Le lexique est une liste de mot avec leur type associé.

Génération. Cliquer sur « Générer une phrase ». Le modèle se sert de la grammaire et du lexique pour générer une phrase en prenant des mots au hasard dans le lexique et en les plaçant aux endroits prévus dans les règles.

Accord de genre. Cliquer sur « Générer 10 phrases ». Parmi les productions, beaucoup comporteront des fautes grammaticales, comme « le chat », « le voiture » ou « la chien ». C'est normal, on ne fait aucune distinction entre les noms féminins et les mots masculins. Pour prendre cela en compte, il faut changer les règles. Cliquer sur « Exemple avec accord de genre » et observer la nouvelle grammaire et le nouveau lexique : les cas masculins et féminins sont distingués. Mais désormais, il n'y a plus de problème d'accord.

Le problème du sens. Toutefois, certaines phrases ne sont toujours pas correctes, notamment car elles n'ont pas de sens. Par exemple, le modèle peut générer « un chat aboie » ou « la voiture pense », qui n'ont pas vraiment de sens. En fait, il faudrait distinguer les sujets selon leurs propriétés : ceux qui peuvent aboyer, ceux qui peuvent penser, ceux qui peuvent marcher, etc. On peut se retrouver très vite avec beaucoup de règles, de longues règles, qui ne pourront jamais couvrir toutes les possibilités.