

Tableau 1

Caractéristiques linguistiques du DS qui seront extraites

Caractéristique/famille de caractéristique	Définition	Opérationnalisation	Nombre de caractéristiques	Informations additionnelles
Mécanique de production de la parole				
Longueur de l'échantillon	Nombre total de mots dans l'échantillon.	Nombre total de lemmas qui ne sont pas des marqueurs de ponctuation, incluant les pauses remplies (ex., hmmm).	1	
Fragments de mots	Production de seulement une partie des phonèmes d'un mot, sans remplacements de sons ou erreurs d'articulation. Peut être suivi de la production complète du mot ou non.	Nombre total de fragments de mots (un ou plusieurs des phonèmes d'un mot, mais pas l'entièreté de ceux-ci) dans l'échantillon.	1	Exemple : « I her kitchkitchen »
Fluence				
Pauses silencieuses	Segments de l'échantillon au cours desquels aucun son n'est produit après que le participant ait commencé à parler.	Nombre total de fois où « [pause] » apparaît dans l'échantillon.	1	Pourraient indiquer : difficultés d'accès lexical, difficultés syntaxiques, difficultés de planification du discours (Boschi et al., 2017).
Pauses remplies	Pause dans le discours, marquée par « uhm » ou une variante de ce son (« hmmm », « hum », « er », « ah », etc.).	Nombre total d'occurrences des mots « uhm », « hmmm », « hum », « uh », « er » et « ah » dans l'échantillon.	1	Pourraient indiquer : difficultés d'accès lexical, difficultés syntaxiques, difficultés de planification du

Caractéristique/famille de caractéristique	Définition	Opérationnalisation	Nombre de caractéristiques	Informations additionnelles
Répétitions de mots, de groupes de mots ou d'idées	Mots ou informations de contenu qui sont présentes plus d'une fois dans l'échantillon (répétitions directement collées les unes sur les autres ou plus éloignées dans l'échantillon).	Nombre total de mots, de groupes de mots (combinaisons de 2 à 6 mots) ou d'informations de contenu qui sont présents plus d'une fois dans l'échantillon.	1 variable x 3 calculs = 3	discours (Boschi et al., 2017). Pourrait indiquer: déficits lexico-sémantiques, difficultés de planification du discours (Boschi et al., 2017)
Caractéristiques lexicales				
Parts-of-Speech*	Classe grammaticale d'un mot.	Nombre d'occurrences de chaque classe grammaticale (noms, pronoms, verbes, adverbess, adjectifs, prépositions, déterminants, conjonctions) dans l'échantillon. Calculé des 2 façons suivantes : en nombre absolu et en relation au nombre total de mots dans l'échantillon.	8 variables x 2 calculs = 16	
Mots de classe ouverte et fermée	<ul style="list-style-type: none"> – Mots de classes ouvertes : mots de contenu (ex., noms, verbes, adjectifs). – Mots de classes fermées : mots servants une fonction grammaticale (ex., conjonctions, pronoms, déterminants). 	<ul style="list-style-type: none"> Nombre total de mots de classe ouverte (noms, verbes, adjectifs, adverbess) et de classe fermée (conjonctions, pronoms, déterminants, prépositions) dans l'échantillon. 	1	

Caractéristique/famille de caractéristique	Définition	Opérationnalisation	Nombre de caractéristiques	Informations additionnelles
Ratio de différentes Parts-of-Speech et types de mots	Proportion de mots de différentes classes grammaticales ou de différents types de mots sur le nombre total de mots dans l'échantillon ou sur le nombre total de mots d'une ou plusieurs autre(s) classe(s) grammaticale(s).	<p>Nombre total d'occurrences d'une classe grammaticale dans l'échantillon divisé par le nombre total de mots dans l'échantillon ou par le nombre d'occurrences d'une ou plusieurs autre(s) classe(s) grammaticale(s).</p> <p>Les ratios suivants seront calculés :</p> <ul style="list-style-type: none"> -Pronoms/Noms + Pronoms -Noms/Noms + Pronoms -Noms/Noms + Verbes -Verbes/Noms + Verbes - Verbes avec inflexions/Nombre total de verbes -Nombre de mots de classe ouverte/Nombre total de mots -Nombre de mots de classe fermée/Nombre total de mots -Gérondifs/Nombre total de verbes -Gérondifs/Nombre total de mots 	10	
Verbes légers*	Verbe qui a peu de contenu sémantique par lui-même, mais qui a une signification	Nombre d'occurrences des verbes suivants (à l'infinitif ou conjugués) dans	1 variable x 2 calculs = 2	

Caractéristique/famille de caractéristique	Définition	Opérationnalisation	Nombre de caractéristiques	Informations additionnelles
	plus complexe lorsque combiné avec un autre mot.	l'échantillon : be, have, come, go, give, take, make, do, get, move, put. Calculé des deux façons suivantes: en nombre absolu et en relation au nombre total de verbes dans l'échantillon.		
Pronoms déictiques*	Pronoms utilisés pour faire référence directement aux caractéristiques personnelles, temporelles ou de localisation de l'image à décrire. La signification spécifique de ces pronoms dépend du contexte dans lequel ils sont utilisés (Crystal, 2011).	Nombre total d'occurrences des mots des quatre catégories suivantes dans l'échantillon: Deixis spatiale : « this », « that », « here », « there », Deixis personnelle : « he », « she », « her », « herself », « him », « himself » Deixis temporelle : « then », « now », « soon », « recently » Pronoms déictiques : somme des pronoms déictiques des trois catégories précédentes Calculé des 2 façons suivantes : en nombre absolu et en relation au nombre total de mots dans l'échantillon.	4 variables x 2 calculs = 6	

Caractéristique/famille de caractéristique	Définition	Opérationnalisation	Nombre de caractéristiques	Informations additionnelles
Termes indéfinis*	Mots qui font référence à un item sur l'image à décrire de façon non-spécifique.	Nombre total d'occurrences des termes suivants dans l'échantillon : « thing », « stuff », « anything », « nothing », « anyone », « one », « either », « neither », « everyone », « no one », « someone », « anybody », « everybody », « nobody », « somebody », « another », « the other », « each », « little », « less », « much », « both », « few », « fewer », « many », « other », « others », « several ». Calculé des 2 façons suivantes : en nombre absolu et en relation au nombre total de mots dans l'échantillon.	1 variable x 2 calculs = 2	
Moving Average Type-Token Ratio (MATTR)	Mesure de diversité lexicale reliant le nombre de mots différents utilisés à la longueur totale de l'échantillon.	Calculé en déplaçant une fenêtre de grandeur « x » à travers le texte. Pour chaque fenêtre, un Type-Token Ratio est obtenu en divisant le nombre de mots uniques par le nombre total de mots dans la fenêtre. Pour obtenir le MATTR global d'un échantillon, la	1	Un MATTR plus élevé indique une plus grande diversité lexicale (Covington & McFall, 2010)

Caractéristique/famille de caractéristique	Définition	Opérationnalisation	Nombre de caractéristiques	Informations additionnelles
		<p>moyenne des TTR de chaque fenêtre est calculée.</p> <p>La longueur de chaque fenêtre sera déterminée en calculant le nombre de mots moyen dans tous les échantillons de DS. Dans ce projet, il est attendu que les échantillons contiennent en moyenne de 100 à 150 mots. Trois groupes de fenêtre seront donc obtenus de façon à ce que chaque fenêtre contienne 10, 25 et 40 mots (Covington & McFall, 2010).</p>		
Statistique R de Honoré	Mesure de diversité lexicale reliant la longueur de l'échantillon au nombre de mots différents utilisés et au nombre de mots mentionnées une seule fois dans l'échantillon.	$R = 100 * \log N / (1 - V1/V)$, où N représente le nombre total de mots de l'échantillon, V représente le nombre de mots différents dans l'échantillon et V1 représente le nombre de mots mentionnés une seule fois dans l'échantillon.	1	Une statistique d'Honoré plus élevée indique une plus grande diversité lexicale.
Index W de Brunet	Mesure de diversité lexicale reliant la longueur de l'échantillon au nombre de	$W = N V^{-.165}$ où N représente le nombre total de mots dans	1	Un index W de Brunet plus élevé indique une moins grande diversité

Caractéristique/famille de caractéristique	Définition	Opérationnalisation	Nombre de caractéristiques	Informations additionnelles
	mots différents utilisés dans celui-ci.	l'échantillon et V représente le nombre de mots différents utilisés dans celui-ci.		lexicale (échelle inversée). Relativement peu affecté par les variations dans la longueur de l'échantillon.
Familiarité†	Degré avec lequel un mot est familier pour les locuteurs d'une langue.	Évaluations subjectives de la familiarité obtenues des normes de Glasgow (Scott et al., 2019) La familiarité moyenne sera calculée pour tous les mots/noms/verbes/adjectifs de l'échantillon.	1 variable x 4 calculs = 4	Des déficits sémantiques et/ou d'accès lexical pourraient se manifester par une utilisation accrue de mots évalués comme étant très familiers (Fraser et al., 2016).
Imageabilité†	Niveau d'effort impliqué dans la génération d'une image mentale du concept représenté par un mot.	Évaluations subjectives de l'imageabilité obtenues des normes de Glasgow (Scott et al., 2019). L'imageabilité moyenne sera calculée pour tous les mots/noms/verbes/adjectifs de l'échantillon.	1 variable x 4 calculs = 4	
Concrétude†	Degré avec lequel le concept dénoté par un mot fait référence à une entité perceptible/tangible.	Évaluations subjectives de la concrétude de Brysbaert et al., 2014. La concrétude moyenne sera calculée pour tous les mots/noms/verbes/adjectifs de l'échantillon.	1 variable x 4 calculs = 4	
Fréquence des mots dans le langage courant†	Évaluation de la fréquence avec laquelle un mot est utilisé dans le langage	Mesure objective de la fréquence des mots tirées du corpus SUBTLEX-US	1 variable x 4 calculs = 4	Les difficultés à accéder à des mots spécifiques résultent généralement en

Caractéristique/famille de caractéristique	Définition	Opérationnalisation	Nombre de caractéristiques	Informations additionnelles
Valence†	courant par les locuteurs d’une langue.	(Brysbaert & New, 2009). La fréquence moyenne sera calculée pour tous les mots/noms/verbes/adjectifs de l’échantillon.	1 variable x 4 calculs = 4	une surutilisation de mots avec une fréquence élevée (Wang et al., 2021).
	Degré d’agréabilité des émotions invoquées par un mot.	Évaluations subjectives de la valence de Warriner et al., 2013. La valence moyenne sera calculée pour tous les mots/noms/verbes/adjectifs de l’échantillon.		
Caractéristiques sémantiques				
25 informations de contenu (ICUs)	Sujets, lieux, objets et actions séparés qui sont représentés dans l’image <i>Cookie Theft</i> .	Une étiquette « VRAI » ou « FAUX » sera attribuée en fonction de la présence ou l’absence de chaque ICU dans l’échantillon. L’étiquette « VRAI » sera attribuée lorsque l’ICU est présente dans l’échantillon et l’étiquette « FAUX » sera attribuée lorsque l’ICU est absente de l’échantillon.	25	La liste d’ICUs utilisée sera celle de Kavé & Levy, 2003, qui se veut une mise à jour de la liste de Croisile et al., 1996.
Nombre total d’ICUs	Nombre total d’ICUs qui apparaissent dans l’échantillon.	Nombre total d’ICUs étiquetées comme « VRAI ».	1	
Efficacité	Ratio de la longueur totale de l’échantillon sur le nombre total d’ICUs présentes dans l’échantillon.	Nombre total de mots dans l’échantillon divisé par le nombre total d’ICUs	1	

Caractéristique/famille de caractéristique	Définition	Opérationnalisation	Nombre de caractéristiques	Informations additionnelles
Densité d'idées	Similarité sémantique moyenne entre les idées (conceptuellement distinctes) transmises à l'intérieur d'une fenêtre de mots déplacée à-travers le texte.	<p>étiquetées comme « VRAI ».</p> <p>Distance cosine (similarité sémantique) moyenne entre toutes les paires de « word embeddings » à l'intérieur d'une fenêtre déplacée à-travers le texte. Les « word embeddings » seront extraits à partir du modèle spaCy « en_core_web_lg » qui a supporté l'identification des dépendances syntaxiques et le Part-of-Speech tagging. À l'intérieur d'une fenêtre, la moyenne de toutes les distances cosines sera calculée. Des fenêtres de 3, 10, 25 et 40 mots avec un incrément de la moitié de la longueur de la fenêtre seront implémentées.</p>	4	<p>Une procédure conceptuellement similaire avec une implémentation différente est présentée dans Ivensky, 2019.</p> <p>Word embeddings : représentation vectorielle (numérique) d'un mot. Lors de la création de word embeddings, les mots sont représentés dans un espace sémantique. Les mots avec une plus grande similarité sémantique et régulièrement utilisés dans des contextes similaires auront des vecteurs (chiffres) plus près les uns des autres.</p>
Caractéristiques syntaxiques				
Dépendances syntaxiques universelles*	Relation de dépendance directionnelle entre une unité syntaxique et les entités qui constituent la structure relationnelle de cette unité (Crystal, 2011).	<p>Nombre total de chaque dépendance syntaxique. Calculé avec spaCy dependencies (DEP) des 2 façons suivantes : en nombre absolu et en</p>	42 variables x 2 calculs = 84	42 dépendances différentes sont retrouvées dans les échantillons qui seront analysés.

Caractéristique/famille de caractéristique	Définition	Opérationnalisation	Nombre de caractéristiques	Informations additionnelles
		relation au nombre total de mots dans l'échantillon.		https://universaldependencies.org/ https://spacy.io/usage/linguistic-features#dependency-parse
Longueur des dépendances syntaxiques	Longueur moyenne et maximale des dépendances syntaxiques.	Nombre moyen et maximal de mots dans les dépendances syntaxiques d'un échantillon.	1 variable x 2 calculs = 2	
Enfants gauches et droits*	Dépendants directs d'un mot qui sont connectés à celui-ci par un seul arc à sa gauche ou à sa droite dans l'arbre de dépendance.	Nombre moyen d'enfants gauches/droits pour chaque mot dans l'échantillon. Calculé à l'aide des commandes spaCy « n_left » et « n_right » des deux façons suivantes : en nombre absolu et en relation au nombre total de mots dans l'échantillon.	2 variables x 2 calculs = 4	https://spacy.io/usage/linguistic-features#navigating
Verbes avec inflexions*	Verbes conjugués.	Verbes dans l'échantillon qui ne correspondent pas à leur lemme tel qu'extraît par spaCy. Calculé des deux façons suivantes: en nombre absolu et en relation au nombre total de mots dans l'échantillon.	1 variable x 2 calculs = 2	
Clauses subordonnées*	Groupe de mots qui n'exprime pas une pensée	Nombre total des 4 types de dépendances universelles de base calculé à l'aide du	1 variable x 2 calculs = 2	https://universaldependencies.org/u/overview/complex-

Caractéristique/famille de caractéristique	Définition	Opérationnalisation	Nombre de caractéristiques	Informations additionnelles
	complète, ne constitue pas une phrase complète. Les clauses complexes impliquant la subordination surviennent lorsqu'un dépendant syntaxique (principal ou non) est utilisé comme structure clausale.	dependency parse par défaut de spaCy : Sujets clausaux (csubj) Compléments clausaux, divisés en ceux dont le sujet doit être contrôlé (sujet à l'extérieur de la clause; xcomp) et ceux dont le sujet n'est pas contrôlé (sujet à l'intérieur de la clause; ccomp) Modificateurs de clauses adverbiaux (advcl) Modificateurs de clauses adnominaux (acl) Calculé des deux façons suivantes : en nombre absolu et en relation au nombre total de mots l'échantillon.		syntax.html#subordination
Longueur moyenne des phrases	Nombre moyen de mots par phrase.	Le nombre moyen de mots par phrase dans l'échantillon sera calculé. Les limites des phrases seront déterminées par le « dependency parse » par défaut de spaCy.	1	https://spacy.io/usage/linguistic-features#sbd
Phrases incomplètes*	Phrases qui ne contiennent pas un minimum d'un verbe et son sujet.	Nombre total de phrases dans l'échantillon qui ne contiennent aucun verbe accompagné de son sujet.	1 variable x 2 calculs = 2	Pourraient indiquer : déficits lexico-sémantiques, déficits syntaxiques, difficultés à

Caractéristique/famille de caractéristique	Définition	Opérationnalisation	Nombre de caractéristiques	Informations additionnelles
		Calculé des deux façons suivantes : en nombre absolu et en relation au nombre total de mots dans l'échantillon.		planifier le discours (Boschi et al., 2017).
Nombre de phrases prépositionnelles* (Boschi et al., 2017)	Phrases qui contiennent une préposition, son objet (nom ou pronom) et n'importe quel modificateur de l'objet.	Nombre total de phrases dans l'échantillon qui contiennent une préposition, son objet (nom ou pronom) et n'importe quel modificateur de l'objet. Calculé des deux façons suivantes : en nombre absolu et en relation au nombre total de mots dans l'échantillon.	1 variable x 2 calculs = 2	
Nombre de phrases verbales*	Phrases de bases contenant au moins un verbe et ces dépendants.	Nombre total de phrases verbales dans l'échantillon. Calculé à l'aide des implémentations de base de spaCy. Calculé des deux façons suivantes : en nombre absolu et en relation au nombre total de mots dans l'échantillon	1 variable x 2 calculs = 2	
Longueur et nombre de phrases nominales*	Phrases de bases qui ont un nom en tête de phrase.	Nombre total et longueur moyenne des phrases nominales dans l'échantillon. Calculé à	2 variables x 2 calculs = 4	https://spacy.io/usage/linguistic-features#noun-chunks

Caractéristique/famille de caractéristique	Définition	Opérationnalisation	Nombre de caractéristiques	Informations additionnelles
		l'aide des implémentations de base de spaCy. Calculé des deux façons suivantes : en nombre absolu et en relation au nombre total de mots dans l'échantillon.		
Temps de verbes utilisés*	Formes que prennent les verbes pour indiquer à quel moment l'action se situe dans le temps.	Nombre total de verbes conjugués au présent, au passé et au futur dans l'échantillon. Calculé des deux façons suivantes : en nombre absolu et en relation au nombre total de mots dans l'échantillon.	3 variables x 2 calculs = 6	
Clauses par phrase	Groupe de mots comprenant un sujet et un verbe, normalement utilisés pour ajouter davantage de détails concernant un nom dans une phrase.	Nombre moyen de clauses par phrase calculé à l'aide des implémentations de base de spaCy.	1	
Proportion de noms accompagnés de déterminants	Proportion de noms pour lesquels un déterminant est présent.	Nombre de noms dans l'échantillon rattachés à un déterminant sur le nombre total de noms dans l'échantillon. Calculé à l'aide du dependency parse de spaCy.	1	

Caractéristique/famille de caractéristique	Définition	Opérationnalisation	Nombre de caractéristiques	Informations additionnelles
Phrases coordonnées* (Boschi et al., 2017)	Phrases unies par une ou plusieurs conjonctions de coordination.	Nombre total de phrases dans l'échantillon contenant les conjonctions de coordination suivantes : « and », « but », « for », « nor », « or », « yet », « so ». Calculé des deux façons suivantes : en nombre absolu et en relation au nombre total de mots dans l'échantillon	1 variable x 2 calculs = 2	
Caractéristiques pragmatiques				
Cohérence locale	Similarité sémantique d'une phrase avec la précédente.	Score de similarité sémantique moyen (distance cosin) entre les phrases, calculé à l'aide de la fonction cosine_similarity de la librairie scikit learn.	1	Des valeurs plus élevées indiquent une plus grande similarité et une moins grande distance sémantique. Des valeurs plus basses indiquent une moins grande similarité et une plus grande distance sémantique.
Mots dénotant l'incertitude*	Mots dénotant une incertitude à-propos de la nature d'un élément de l'image à décrire.	Nombre d'occurrences des mots suivants dans l'échantillon : « think », « look », « like », « kind », « seem », « maybe », « can », « something ». Calculé des deux façons suivantes : en nombre absolu et en relation au	1 variable x 2 calculs = 2	Liste de mots inspirée par (Garrard et al., 2014)

Caractéristique/famille de caractéristique	Définition	Opérationnalisation	Nombre de caractéristiques	Informations additionnelles
Difficultés à trouver les bons mots*	Utilisation de mots indiquant des difficultés d'accès lexical.	<p>nombre total de mots dans l'échantillon.</p> <p>Nombre d'instances des mots suivants dans l'échantillon : « know », « remember », « unable ».</p> <p>Calculé des deux façons suivantes : en nombre absolu et en relation au nombre total de mots dans l'échantillon.</p>	1 variable x 2 calculs = 2	Liste de mots inspirée par Garrard et al. 2014 et Rentoumi et al. 2014.
Connotation du discours	Émotions générées par le discours. Dépend de la valence moyenne des mots du discours.	<p>Le score de valence moyen de tous les mots de l'échantillon sera obtenu lors de l'extraction des variables psycholinguistiques. Pour chaque mot, les scores possibles se situent entre 1 et 9. Un score plus élevé indique qu'un mot a une connotation davantage positive alors qu'un score plus bas indique une connotation davantage négative.</p> <p>Si la valence moyenne est supérieure ou égale à 5,5, l'étiquette « connotation</p>	1	

Caractéristique/famille de caractéristique	Définition	Opérationnalisation	Nombre de caractéristiques	Informations additionnelles
		positive » sera donnée au discours.		
		Si la valence moyenne est supérieure ou égale à 4,5 et inférieure à 5,5, l'étiquette « connotation neutre » sera donnée au discours.		
		Si la valence moyenne est inférieure à 4,5, l'étiquette « connotation négative » sera donnée au discours.		
Expressions formulaïques* (Van Lancker Sidtis et al., 2015)	Expressions ayant une forme fixe et une signification non- littérale avec des nuances attitudinales.	Nombre total d'occurrences des expressions formulaïques suivantes dans l'échantillon : « well », « so », « I guess », « you know », « as it is », « as it were ». Calculé des deux façons suivantes : en nombre absolu et en relation au nombre total de mots dans l'échantillon.	1 variable x 2 calculs = 2	
Modalisations* (Boschi et al., 2017, Boyé et al., 2014)	Opinions d'un individu concernant le contenu de sa description (ou ce qui se passe sur l'image à décrire) incluant les doutes et les	Nombre total d'occurrences des expressions suivantes dans l'échantillon : « I think », « In my opinion », « of course », « naturally », « unsure », « likely », it «	1 variable x 2 calculs = 2	

Caractéristique/famille de caractéristique	Définition	Opérationnalisation	Nombre de caractéristiques	Informations additionnelles
	inquiétudes par rapport à sa production.	could be that », « unfortunately », « surely ». Calculé des deux façons suivantes : en nombre absolu et en relation au nombre total de mots dans l'échantillon.		
Mots de remplissage*	Sons, mots ou groupes de mots utilisés pour mettre l'accent sur ce qui sera dit ou a été dit ou qui signalent qu'un individu réfléchi à ce qu'il dira ensuite.	Nombre total de fois où les expressions « you know » et « I mean » sont mentionnées dans l'échantillon. Calculé des deux façons suivantes : en nombre absolu et en relation au nombre total de mots dans l'échantillon.	1 variable x 2 calculs = 2	Pourraient donner de l'information sur la capacité d'accès lexical d'un individu.

Note. Ce tableau se veut une version modifiée de celui qui est retrouvé dans Slegers et al., 2021.

* Caractéristiques qui seront calculées en nombre absolu et en relation au nombre total de mots dans l'échantillon.

† Caractéristiques psycholinguistiques. Seront calculées pour l'ensemble des mots, mais également de façon séparée pour les noms, les adjectifs et les verbes.