

ANALYSE LEXICO-GRAMMATICALE D'ÉCHANTILLONS DE LANGAGE PAR ORDINATEUR

Objectif :

Mettre à la disposition des cliniciens du langage des logiciels compatibles avec les micro-ordinateurs Apple II et un manuel explicatif de façon à procéder à une analyse lexicale et grammaticale largement automatisée d'échantillons de langage, analyse susceptible de mettre immédiatement en évidence les principales caractéristiques du langage produit.

Références :

Le projet est inspiré d'une réalisation similaire pour l'anglais mise au point par College-Hill, San Diego, California, vendue auprès des cliniciens américains au prix de \$975 (total package).

Procédure :

1) Recueil des échantillons de langage

Un échantillon de langage contenant entre 50 et 100 énoncés comportant deux mots ou davantage est recueilli dans une situation de jeu ou de conversation libre.

2) Transcription et protocolisation

- a - Les énoncés sont retranscrits intégralement sur un protocole spécialement préparé à cet effet. Diverses indications contextuelles peuvent également figurer sur le protocole en regard des énoncés.

- b - Les énoncés produits (éventuellement incorrects grammaticalement parlant) par le client sont pairés sur le protocole avec les énoncés corrects correspondants réécrits par le clinicien.

- c - Les noms propres, les noms communs (singuliers et pluriels), et les verbes (y compris les participes) de même que les morphèmes grammaticaux doivent être marqués dans chaque énoncé préalablement à l'entrée des données dans l'ordinateur. L'annexe 1 fournit un exemple de protocole.

3) Entrée des données dans le micro-ordinateur

Quelques règles simples doivent être respectées notamment celle qui concernent l'espacement obligatoire entre les mots.

4) Analyse des données

a - Analyse lexicale

— Les mots produits par le client ainsi que le lexique corrigé sont testés. Le relevé reprend également le nombre de fois que la version corrigée du mot est identique ou différente par rapport à la version produite par le client. Le nombre de fois qu'un item lexical obligatoire dans un contexte linguistique donné est omis par le client figure également au relevé.

— A. Relevé du nombre total de mots différents produits par le client

B. Nombre de total de mots produits

C. Indice de diversité lexicale : $C = \frac{A}{B}$

Des données normatives seront fournies permettant d'interpréter l'ICL.

b - Analyse grammaticale

— Longueur moyenne d'énoncé (LME)

Client Clinicien

— Nombre total d'énoncés

	Client	Clinicien
— Nombre total des mots	—	—
— Nombre total de morphèmes grammaticaux	—	—
— LME en nombre de mots	—	—
— LME en nombre de morphèmes grammaticaux	—	—

Des données normatives seront fournies permettant d'interpréter la LME.

— Analyse distributionnelle en classes grammaticales

ALGELO fournit un listing reprenant le nombre d'occasions d'apparitions, le nombre de cas d'usage correct, le pourcentage d'usage correct, le nombre d'erreurs et le type d'erreur (omission ou substitution) selon une dizaine de classes grammaticales (noms, verbes, adjectifs, adverbes, pronoms, prépositions, négations, etc.), et une centaine de sous-classes grammaticales. Les substitutions sont tabulées automatiquement révélant ce qui est substitué à quoi et selon quelle fréquence.

— Analyse distributionnelle de la structure des énoncés

ALGELO fournit un listing reprenant le nombre d'occasions d'apparitions, le nombre de cas d'usage correct, le pourcentage d'usage correct, et le nombre d'erreurs pour trois types de syntagmes (syntagme nominal, syntagme verbal et syntagme propositionnel), une dizaine de types de phrases simples, et une dizaine de types de phrases complexes. Un pourcentage global reprenant le nombre de structures énonciatives correctement formulées par rapport au nombre d'occasions d'apparitions est également fourni.

— Analyse distributionnelle portant sur le temps et l'aspect au niveau des verbes

Cette analyse identifie le nombre d'occasions d'apparitions, le nombre de cas d'usage correct, le pourcentage d'usage correct et le nombre d'erreurs pour une dizaine de temps du français. Un pourcentage global reprenant le nombre d'usages temporels corrects par rapport au nombre d'occasions d'apparitions est également fourni.

Annexe 1 : Exemple de protocole ALGELO

Analyse d'échantillons de langage

Client	Date de naissance	Age chronologique		
Adresse				
Niveau scolaire ou profession				
Clinicien	Date de l'examen	Commentaire contextuel		
	Enoncés			
	PP	PRFL	V	
I (Enoncé corrigé)	Ils	se	cachent	
I (Enoncé client)	Ils		cachent	
	PP		V	
	PP	PRFL	V	ADV
2	Je	me	cache	aussi
2	Il		cache	aussi
	PP		V	ADV
	A	N	AUX	PARTICIPE
3	La	voiture	est	partie
3		voiture		partir
		N		V