

# MANULEX

## SIGNIFICATION DES ABRÉVIATIONS

Le lexique des lemmes et des formes orthographiques sont inclus dans deux feuilles Excel séparées. Le lexique des lemmes reprend tous les mots rencontrés à chaque niveau mais sous une forme non fléchi (au singulier pour les noms et adjectifs, à l'infinitif pour les verbes). Les mots trouvés dans chaque niveau sont en colonnes. La cellule est vide quand le mot n'a pas été trouvé dans les manuels de ce niveau.

**CP** : mots trouvés dans les manuels de CP

**CE1** : mots trouvés dans les manuels de CE1

**CE2-CM2** : mots trouvés dans les manuels du cycle 3 (CE2 – CM1 – CM2)

**CP-CM2** : mots trouvés dans l'ensemble des 54 manuels du CP au CM2

**NLET** : Nombre de lettres

**SYNT** : **Catégorie morphosyntaxique** : NC-nom, NP-nom propre, VER-verbe, ADJ-adjectif, ADV-adverbe, PRO-pronom, PRE-préposition, CON-conjonction, INT-interjection, DET-déterminant, ABR-abréviation, and UEUPH-marque euphonique

**F** : **Fréquence brute** : le nombre d'occurrences du mot dans le niveau considéré. Attention: F n'est pas la fréquence par million; utilisez U pour une correcte estimation de la fréquence par million

**D** : **Indice de Dispersion** (de .00 à 1.00) : il reflète la dispersion des occurrences des mots à travers les manuels. D est égal à 0 lorsque toutes les occurrences du mot ont été trouvées dans un seul manuel, quelle que soit la fréquence (F). Il est égal à 1 si les fréquences sont réparties dans des proportions exactement égales à travers les manuels. Les valeurs comprises entre 0 et 1 indiquent un degré de dispersion entre ces deux extrêmes. Pour les mots à très haute fréquence, D modifie très peu le classement qui serait obtenu avec F, alors qu'il reclasse fortement les mots à fréquence moyenne ou faible.

**U** : **Fréquence estimée d'Usage pour 1 million de mots** : U est dérivée de F avec un ajustement avec D. Lorsque D est égal à 1, U est calculé simplement comme la fréquence par million de mots. Mais quand D est inférieur à 1, la valeur de U est ajustée à la baisse. On considère que U est un meilleur estimateur de la "vraie" fréquence par million que l'on trouverait dans un corpus théorique de taille infinie. L'indice permet ainsi des comparaisons directes des valeurs de fréquence issues de corpus de tailles différentes.

**SFI** : **Standard Frequency Index** : un indice de fréquence courant calculé à partir de U par transformation logarithmique. L'utilisateur devrait trouver cet indice pratique une fois maîtrisé. Un SFI de 90 signifie une rencontre (une occurrence) tous les 10 mots lus. Un SFI de 80 traduit une rencontre tous les 100 mots. Un SFI de 70, tous les 1000 mots, etc... Un SFI de 40 signifie une seule rencontre dans un million de mots. Le SFI est calculé à partir de U en utilisant la formule:

$$\text{SFI} = 10 * (\log_{10}(U) + 4)$$