

# Nouvelle mesure de complexité de l'information pour l'analyse du comportement de mobilité éco-responsable

CHAMBON Thomas<sup>1</sup>, LALLEMENT Jeanne<sup>2</sup>, REVEL Arnaud<sup>1</sup>

<sup>1</sup>Laboratoire Informatique, Image et Interaction (L3I), <sup>2</sup>Centre de Recherche en Gestion (CEREGE), La Rochelle Université, La Rochelle, France

[thomas.chambon1@univ-lr.fr](mailto:thomas.chambon1@univ-lr.fr), [jeanne.lallement@univ-lr.fr](mailto:jeanne.lallement@univ-lr.fr), [arnaud.revel@univ-lr.fr](mailto:arnaud.revel@univ-lr.fr)

Le changement climatique est l'un des principaux enjeux du siècle : les scénarios les plus optimistes prédisent une augmentation des températures de plusieurs degrés en quelques décennies, avec des conséquences catastrophiques. Pour endiguer ce fléau, chacun d'entre nous doit adapter son comportement pour réduire ses émissions de gaz à effet de serre. Parmi les différents facteurs contribuant à ces émissions, la mobilité joue un rôle central. Pour essayer de changer progressivement les comportements, nous utilisons le concept de nudging (coup de coude) [1]. Notre but est de rendre plus efficient ce concept de « coup de pouce », à l'aide des technologies mobiles. Dans cette perspective, nous souhaitons analyser et modéliser le comportement des utilisateurs en leur transmettant des informations, et ainsi comprendre si elles sont susceptibles de modifier leurs habitudes de transport. La première étape de ce processus est la quantification de l'information, pour cela, nous proposons une nouvelle méthode de classification de l'information en fonction de leur complexité. Cette méthode est appelée Information Complexity Ranking (ICR).

Pour mesurer l'impact de l'information, nous devons la quantifier, pour cela nous proposons une nouvelle méthode de classement (ICR) des objets numériques (en fonction de leur complexité) :

$$ICR(x,y) = \frac{NCD(x,y) \times C(y)}{C(x)}$$

Cette méthode repose sur deux notions:

- La complexité intrinsèque d'un objet (complexité de Kolmogorov [2]), notée  $C(x)$  ou  $C(y)$ . Le résultat est obtenu à l'aide d'algorithmes de compression.

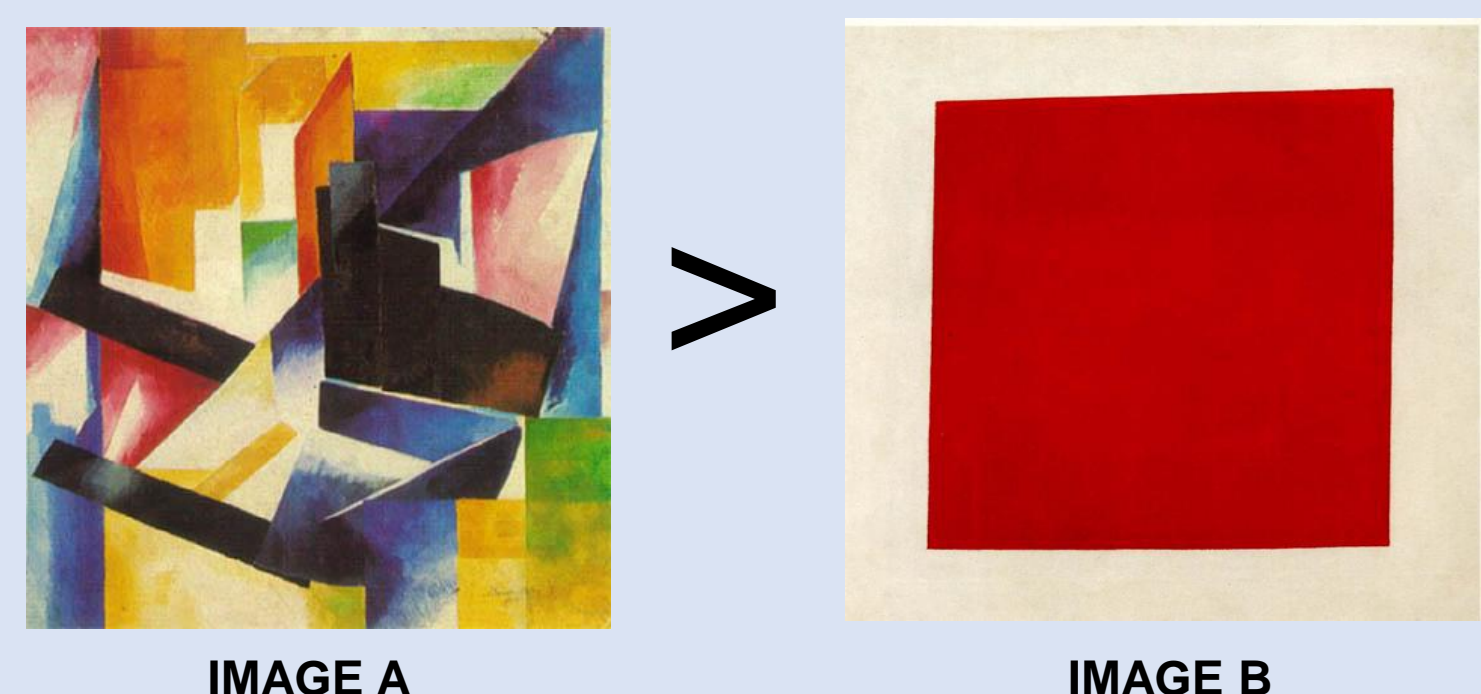
$$C(x) = \frac{\text{Taille originale de l'objet } x}{\text{Taille compressée de l'objet } x}$$

- La similarité entre deux objets (Distance de compression normalisée [3]), notée  $NCD(x,y)$

Cette relation d'ordre respecte les propriétés de **réflexivité**, de **transitivité** et de **antisymétrie**.

Figure 1 :

Exemple du résultat de la méthode ICR sur les images A et B.  $ICR(A,B) = 4,032$   
 $ICR(B,A) = 0,229$



## RESULTATS

Pour pouvoir mettre en œuvre notre méthode (ICR), les algorithmes de compression doivent respecter les propriétés précédemment citées. Le sens de concaténation de l'information est aussi pris en compte (horizontale ou verticale) (ordre de l'information). Tous les algorithmes testés respectent les propriétés requises.

Algorithme	Monotonie horizontale	Monotonie verticale	Distributivité horizontale	Distributivité verticale
PNG	0.318	0.344	0.313	0.329
PPM	0.202	0.204	0.197	0.202
LZ77	0.285	0.298	0.273	0.287
ALZ	0.197	0.199	0.203	0.203
LZP	0.334	0.329	0.327	0.313
LZW	0.323	0.343	0.314	0.333
AC	0.731	0.909	0.736	0.894
Huffman	<b>1.423</b>	<b>1.588</b>	<b>1.414</b>	<b>1.549</b>

Table 1 : Performance de plusieurs algorithmes de compression en ce qui concerne les propriétés de monotonie et de distributivité. Les valeurs maximales sont indiquées en gras.

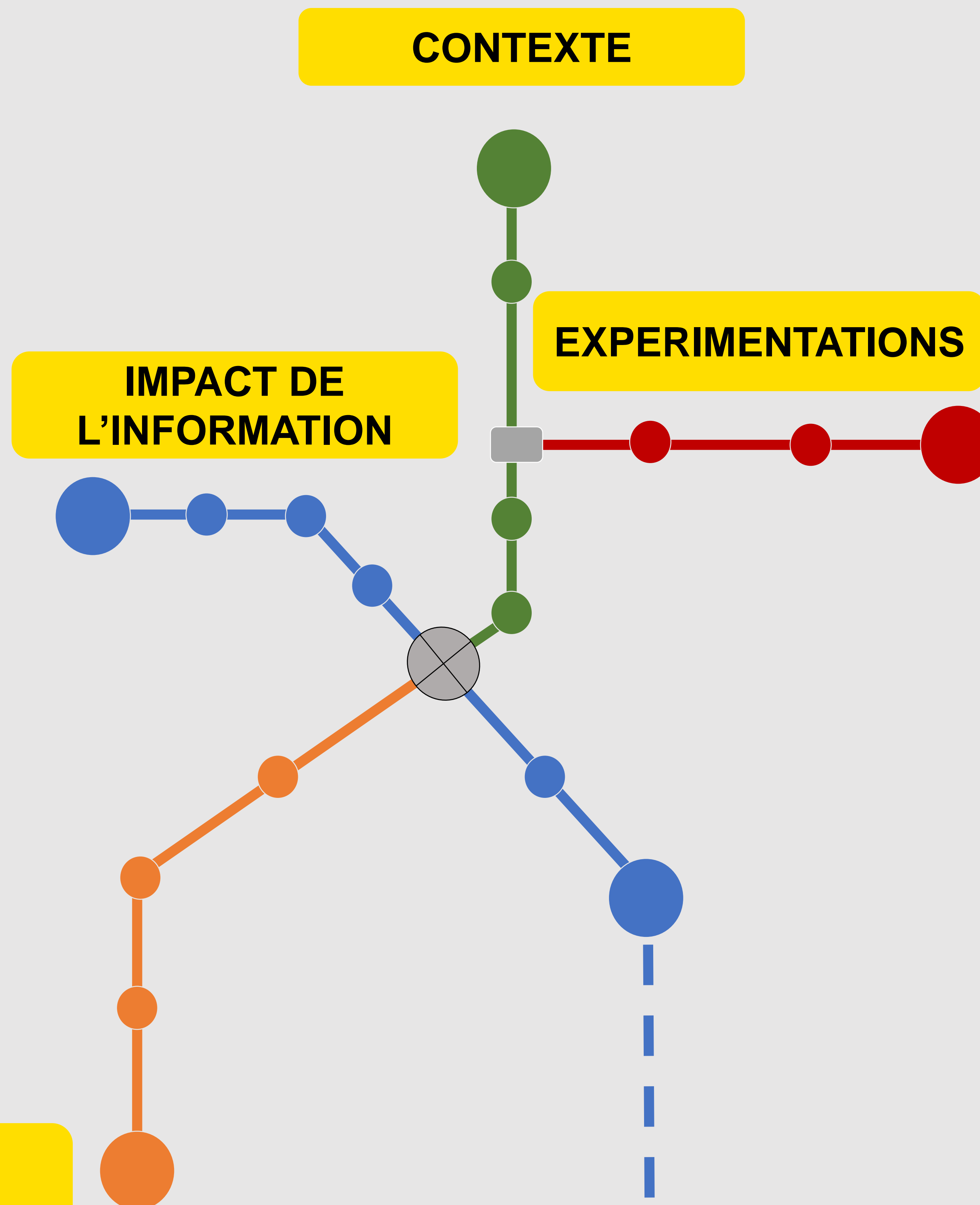
Catégories	Ratio de compression vs Vérité Terrain	ICR vs Vérité Terrain
Scène	0.30	0.40
Publicité	0.56	0.58
Visualisation	0.55	0.31
Objet	0.16	0.16
Design intérieur	0.68	0.67
Art	0.51	0.58
Suprématisme	0.60	<b>0.73</b>

Table 3 : Comparaison entre notre nouvelle métrique et le taux de compression. Les valeurs de la première colonne ont été reportées de la colonne du ratio de compression du tableau 3 dans [4].

Table 2 : Performance de plusieurs algorithmes de compression en ce qui concerne les propriétés d'idempotence et de symétrie. Les valeurs minimales sont indiquées en gras.

Après avoir comparé notre méthode à une autre mesure de la complexité, on constate que les valeurs sont similaires dans cinq des sept catégories. Cependant avec la catégorie *Visualisation* nous obtenons de moins bons résultats. Cela est dû à la forte dissimilarité des images, ce qui pénalise notre approche. A noter, que la catégorie *Suprématisme*, présente les meilleurs résultats, dû au faible aspect sémantique des images.

## CONTEXTE

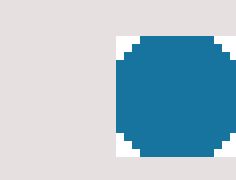


Notre méthode est une mesure abstraite reposant sur des hypothèses fortes concernant les propriétés que devraient avoir le compresseur et l'opérateur de concaténation, et doit être validée sur des applications concrètes. Nous avons donc réalisé deux expérimentations :

- Expérimentation de synthèse
  - Génération aléatoire de **7200** images réparties en 3 types différents.
  - Test avec 8 algorithmes de compression (PNG, PPM, LZ77, ALZ, LZP, LZW, AC, Huffman)
  - But:** Vérifier le respect des propriétés d'idempotence, de symétrie, de monotonie et de distributivité.

Figure 2 :

Les trois types d'information testés : (point de couleur, nombre et texte)



3170

ijlzxagbthpsjdxjkzx

- Expérimentation sur un jeu de données de référence (complexité visuelle) (SAVOIAS) [4]
  - Jeu de données composé de **1400** images réparties en sept catégories.
  - But:** Comparer notre nouvelle méthode au ratio de compression obtenu dans le jeu de données.



Figure 3 : Exemples d'images du jeu de données SAVOIAS (catégories Objet et Scène).

## PERSPECTIVES

Notre but étant de construire une mesure globale de l'information et de son impact, l'intégration de l'aspect sémantique fera partie de nos futurs travaux. L'intégration d'un aspect centré sur la personne est aussi envisagée pour mieux comprendre son comportement vis-à-vis de l'information transmise. Nous souhaitons également améliorer notre méthode en explorant d'autres pistes théoriques (profondeur logique de Bennett [5]).

Nos futurs travaux expérimentaux se focaliseront sur des tests en conditions réelles sur environ 800 étudiants durant une période de 3 mois à l'aide d'une application mobile de traçage que nous développons.

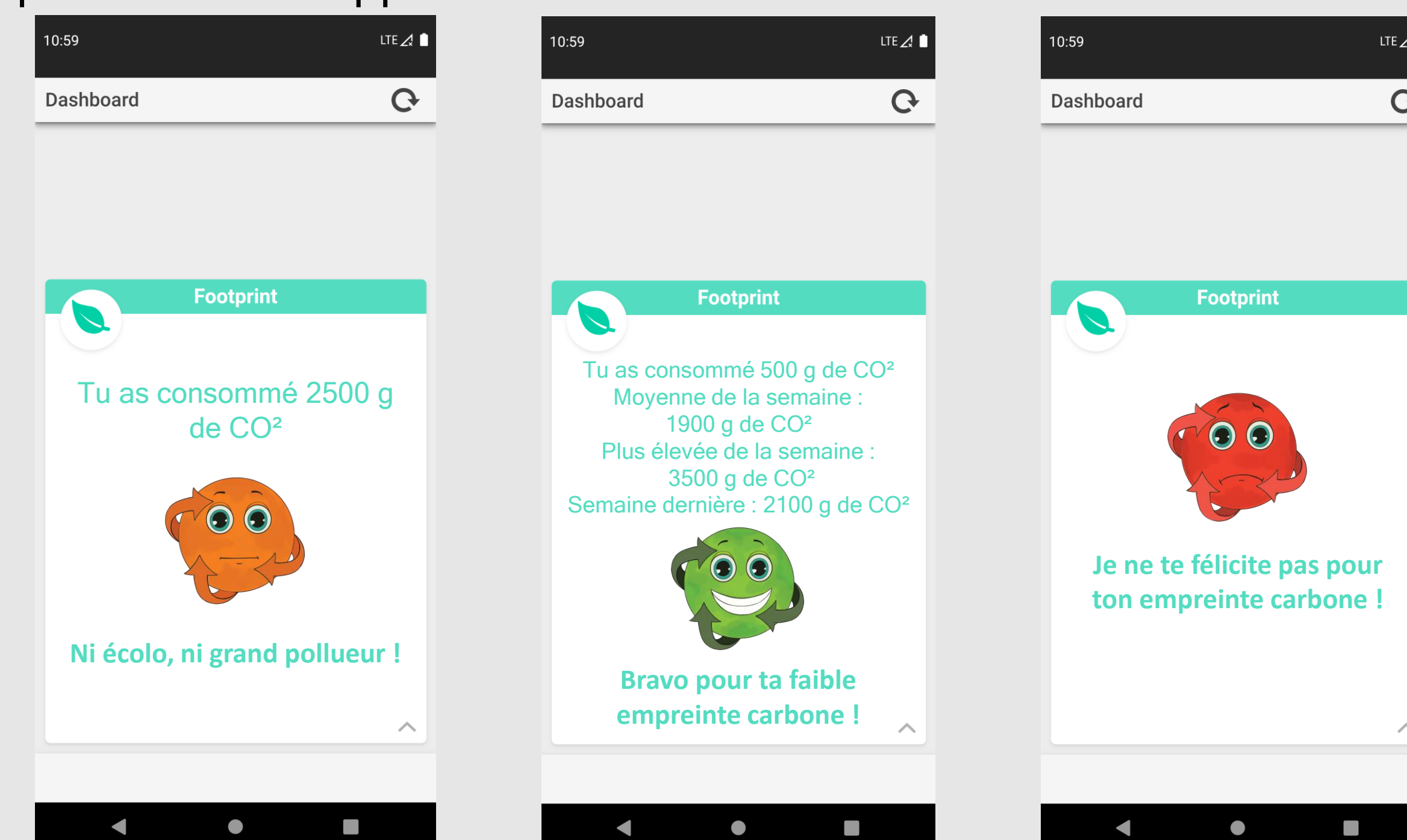


Figure 4 : Exemples d'informations transmises aux personnes à l'aide de l'application mobile de traçage.

[1] R.H. Thaler and C.R. Sunstein. Nudge: Improving decisions about health, wealth, and happiness. 2008.

[2] A. N. Kolmogorov. Three Approaches to Information. Problemy Peredachi Informatsii, 1(1):3–11, 1965.

[3] Rudi Cilibrasi and Paul M.B. Vitanyi. Clustering by compression. IEEE Transactions on Information Theory, 51(4):1523–1545, 2005.

[4] Elham Saraee, Mona Jalal, and Margrit Betke. Savoias: A diverse, multi-category visual complexity dataset, 2018.

[5] C. H. Bennett. Logical depth and physical complexity. In A Half-Century Survey on The Universal Turing Machine, page 227–257, USA, 1988. Oxford University Press, Inc.