

La moitié des messages sur Twitter ne sont pas en anglais

La seconde langue la plus utilisée est le japonais

Une étude quantitative et sémantique de Twitter

Paris, France — 24 février 2010 — 50% des messages sur Twitter ne sont pas en anglais d'après une analyse de 2,8 millions de tweets. Cette étude, menée par Semiocast, montre que les 5 langues les plus utilisées sur Twitter sont : l'anglais, le japonais, le portugais, le malais (officiel et indonésien) et l'espagnol.

Les messages ont été collectés du 8 au 10 février 2010 et ont été traités par les outils analytiques de Semiocast. Ces outils permettent d'identifier la langue d'un micro-message parmi 41 langues dans la majorité des systèmes d'écriture (notamment grec, hébreu, chinois, coréen, tamoul, devanagari, ...).

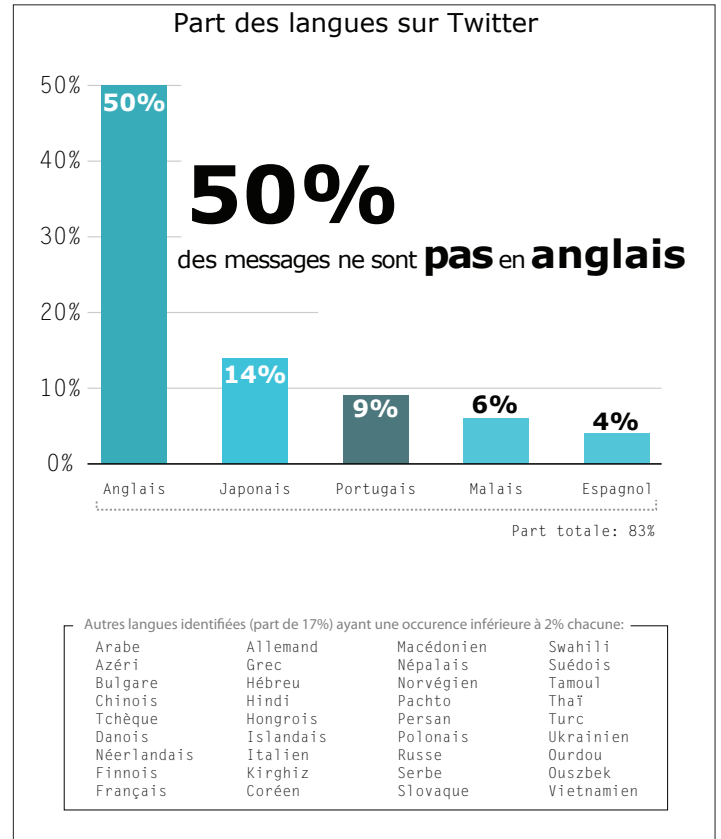
L'anglais est toujours la langue la plus fréquente sur Twitter, avec 50% des messages. Ce pourcentage reflète le fort taux de pénétration de Twitter dans les pays anglophones, ainsi que la tendance des utilisateurs de pays non-anglophones à tweeter en anglais. Cependant, l'internationalisation de Twitter est rapide : en effet, l'anglais représentait encore deux tiers des échanges au premier semestre 2009. La part de l'anglais devrait continuer à décroître à l'avenir, car la croissance de Twitter est tirée par celle des pays non-anglophones.

Le japonais est la deuxième langue sur Twitter, avec 14% des messages. Ce chiffre démontre la popularité de Twitter au Japon et l'importance que ce pays a eu dans le développement international du réseau Twitter.

La troisième langue la plus utilisée est, sans surprise, le portugais, au regard du succès des réseaux sociaux et du Web temps réel au Brésil. 9% des messages sont rédigés dans cette langue, soit près de 4,5 millions de messages par jour.

L'adoption de Twitter a été très rapide en Indonésie et en Malaisie, grâce aux accords conclus avec des opérateurs télécoms. Leur position dans le classement le prouve : le malais officiel et l'indonésien sont en quatrième position avec 6% des messages. Ainsi, tous les jours, 3 millions de messages sont échangés en malais.

L'espagnol arrive en cinquième avec 4% des messages, reflétant l'importance de la population hispanophone, notamment dans le pays d'origine de Twitter, les États-Unis.



Les places 6 à 8 sont occupées par des langues européennes majeures, à savoir l'italien, le néerlandais et l'allemand, chacune représentant entre 1% à 2% des messages journaliers.

Le français est 11^e, avec un peu moins de 1% des messages transitant sur Twitter. Par comparaison avec Facebook, la part du français sur Twitter dénote une relative sous-représentation. En effet, la France est aujourd'hui le 7^e pays le plus présent sur Facebook.

La détection de la langue est un pré-requis à toute analyse sémantique des micro-messages. Même un indicateur standard, tel que le sentiment associé, ne peut être déterminé avec une certaine fiabilité si la langue n'a pas été identifiée correctement au préalable.

Sur Semiocast

Start-up parisienne à l'origine de Titema, premier agrégateur Twitter intelligent, Semiocast a été fondée en février 2009.

Semiocast permet aux entreprises de mesurer et de comprendre les échanges du Web temps réel à propos de leurs marques et d'interagir avec leurs consommateurs.

Notre offre d'études marketing et d'opinion s'appuie sur nos technologies d'indexation du Web temps réel et d'analyse sémantique des micro-messages. Ces technologies nous permettent de donner une vision synthétique des sujets qui structurent les conversations autour d'une marque, et de suivre, en temps réel, l'évolution du buzz ou l'impact d'une campagne de communication.

More information : semiocast.com
 Contact : Paul Guyot (paul@semiocast.com — +33.175000290)