

# Visualisation de l'évolution des sentiments sur Twitter

---



## Mémoire de fin d'étude

Master *Sciences et Technologies*,

Mention *Informatique*,

Parcours DECOL

**Auteur**

Erick CUENCA

**Superviseurs**

Pascal PONCELET

Florence Ying WANG

**Lieu de stage**

LIRMM UM5506 - CNRS, Université de Montpellier

---

soutenu publiquement le 11 juin 2015



## Résumé

L'analyse de textes et de sentiments sur les réseaux sociaux est un enjeu important pour différentes activités telles que la mise en place de stratégies d'affaires ou de politiques publiques. Aujourd'hui les services de microblogging (*e.g.* twitter) attirent l'attention en raison des grandes masses de données qu'ils véhiculent. Analyser ces données; *i.e.* comprendre la grande quantité de messages recueillis à chaque instant et montrer des informations utiles aux utilisateurs est un problème difficile. Il est donc crucial de développer et connaître les méthodes efficaces de visualisation de texte et plus spécifiquement de sentiments pour aider les utilisateurs dans leur tâche d'analyse.

**Mots clés** : Visualisation d'informations, Analyse des sentiments.

---

## Abstract

Text analysis, particularly sentiment analysis of social media, has become an important issue for various activities such as the development of business strategies and public policies. Today, microblogging services (*e.g* twitter) have received increasing attention because of the large amount of information they carry. Social media analysis, for example, understanding of the large amount of text messages collected at every moment, is not a trivial problem. Therefore, it is crucial to study and develop effective methods of text visualization and specifically sentiment visualization to facilitate the analysis tasks.

**Keywords** : Information visualization, Visual analysis.



# Table des matières

<b>Table des matières</b>	<b>v</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Contexte et problématique . . . . .	1
1.2 Objectifs du stage . . . . .	2
1.3 Contributions . . . . .	2
<b>2 Cadre théorique</b>	<b>3</b>
2.1 Visualisation d'informations . . . . .	3
2.2 Visualisation de textes et de sentiments contenus dans les textes . . . . .	4
2.3 Techniques d'interaction visuelle . . . . .	5
2.4 Processus général . . . . .	7
2.5 Modèle de création de visualisations . . . . .	8
<b>3 Conception d'une nouvelle approche visuelle</b>	<b>9</b>
3.1 Caractérisation de la problématique . . . . .	9
3.2 Caractérisation des structures de données et des besoins . . . . .	9
3.3 Conception des techniques de représentation et d'interaction . . . . .	12
<b>4 Mis en œuvre</b>	<b>19</b>
4.1 Conception des algorithmes . . . . .	19
4.2 Étude de cas . . . . .	22
<b>5 Conclusions</b>	<b>25</b>
<b>Bibliographie</b>	<b>27</b>



# Introduction

La visualisation d'informations est la « *présentation visuelle d'un ensemble d'informations issues de données abstraites et traitées par des moyens informatiques* » [1]. La difficulté de la visualisation d'informations est de déterminer quelle représentation choisir pour bien répartir les données abstraites dans l'espace et montrer un message cohérent et pertinent, basé sur l'exploitation des capacités visuelles et cognitives de l'être humain. La visualisation de sentiments est une branche clé pour l'analyse d'informations qui n'a pas été suffisamment développé.

Ce travail est organisé de la manière suivante : la section suivante présente le contexte de la problématique de recherche que nous traitons, l'objectif et les principales contributions de ce projet. Dans le Chapitre 2 nous définissons les principaux concepts de la visualisation d'informations dans le cadre de notre projet. Le Chapitre 3 est consacré à la conception de notre approche visuelle. Le Chapitre 4 présente la mise en œuvre et les études de cas. Finalement, dans le Chapitre 5, nous concluons.

## 1.1 Contexte et problématique

Suite au développement de l'utilisation d'internet et l'explosion des réseaux sociaux dans le monde, il est devenu facile d'exprimer une idée, une opinion ou un sentiment à propos de n'importe quel sujet ou événement. Dans ce contexte, les services de microblogging permettent aux utilisateurs de s'exprimer en postant de courts messages. Comme le présente Wikipédia<sup>1</sup>, « *Twitter est un outil de microblogging géré par l'entreprise Twitter Inc. Il permet à un utilisateur d'envoyer gratuitement de brefs messages, appelés tweets, sur internet, par messagerie instantanée ou par SMS. Ces messages sont limités à 140 caractères* ». Les utilisateurs de Twitter expriment des messages (*tweets*) qui sont visibles par tout le monde et qui sont envoyés directement à leurs abonnés (*followers*). Twitter est devenu un service populaire de microblogging en pleine expansion, il est un moyen d'expression des internautes parce qu'il permet d'échanger en temps réel, sur tous les sujets, des points de vue ou des besoins. Ces messages envoyés (*tweets*) sont bien adaptés à la diffusion et propagation d'informations, parce que ils peuvent être republiés et aussi contenir des hash-tags, c'est-à-dire des étiquettes attribuées par les auteurs des tweets pour caractériser brièvement le sujet du tweet. Les tweets sont munis de métadonnées ainsi que des informations sur leur emplacement, langue, mot clé, sentiments exprimés, etc. Voici un exemple de tweet : « *Félicitations à l'Equipe de France des*

---

<sup>1</sup>Information du site <http://fr.wikipedia.org/wiki/Twitter> [Online ; accessed 23-mai-2015]

*moins de 17 ans, magnifique championne d'Europe face à l'Allemagne. Un présage pour l'#EURO2016 ? »<sup>2</sup>.*

Dans ce projet nous nous intéressons au problème de l'analyse et de la visualisation des sentiments contenus dans les textes, plus précisément dans des tweets. Compte tenu du nombre de messages échangés sur les réseaux sociaux, il est impossible de manipuler les données manuellement, et des outils de traitement automatique doivent donc être mis en place. Comme le mentionnent Wu *et.al* [2] « *sentiment analysis aims to assist users to automatically detect relevant opinions within a large volume of review collection and create a coherent overview of these opinions* ».

## 1.2 Objectifs du stage

L'objectif de ce projet est de proposer une nouvelle approche visuelle permettant d'analyser efficacement et de comprendre les sentiments des gens sur un sujet au fil du temps, afin de restituer le maximum d'informations aux utilisateurs. L'équipe ADVANSE<sup>3</sup> du LIRMM, spécialisée dans l'analyse de données et la visualisation d'informations, a mis en place une première métaphore visuelle, appelée *SentiCompass* [3] qui représente de façon simple et facilement interprétable les sentiments issus de plusieurs milliers de messages sur Twitter. Cependant, cet outil est limité, parce que certains choix graphiques sont trompeurs et il est difficile de comprendre les grandes tendances des jeux de données traités. Le but de mon projet est de repenser une nouvelle approche visuelle afin de palier ces défauts. Pour cela, il est nécessaire d'avoir une connaissance précise des méthodes de visualisation de sentiments à partir de données textuelle. Ceci me permettra de proposer les projections graphiques adéquates. Cette nouvelle approche devra inclure des techniques de visualisation d'informations basées sur représentations visuelles et des modes d'interaction adaptés. A titre d'illustration, le site *Textvis*<sup>4</sup> permet de voir quelques approches visuelles.

## 1.3 Contributions

Nous proposons une nouvelle approche visuelle qui prend en compte les dimensions d'un modèle affectif (*i.e.* émotions) avec une représentation visuelle en flux afin de représenter l'évolution des sentiments au fil du temps. Un des aspects intéressants est la visualisation de l'agrégation et désagrégation des sentiments avec l'aide d'un ensemble des techniques d'interaction et de représentations visuelles. Nous avons mis en place des algorithmes pour générer des zones de transitions (*i.e.* fisheye) au moment de l'agrégation ou désagrégation des sentiments.

---

<sup>2</sup>Tweet de la compte du Président François Hollande [**@fhollande ; 22-mai-2015**]

<sup>3</sup>ADVanced Analytics for data SciencE.

<sup>4</sup><http://textvis.lnu.se/> : Ce site présente des techniques de visualisation, élaborées par le groupe ISOVIS. [**Online ; accessed 27-mai-2015**]

## Cadre théorique

Dans ce chapitre nous présentons l'état de l'art relatif à notre projet. Dans un premier temps, nous définissons certains concepts fondamentaux sur la visualisation d'informations. Ensuite, nous décrivons quelqu'un des principaux paradigmes et la visualisation de sentiments contenus dans les textes. Puis nous décrivons les techniques d'interactions visuelles. Ensuite nous présentons le processus général de la visualisation d'informations et la démarche dans la création de visualisations.

### 2.1 Visualisation d'informations

Tout d'abord nous allons définir les notions d'**information** et de **visualisation** pour arriver à comprendre au final la notion de **visualisation d'informations**.

**L'information** est définie dans [1] comme : « *Élément de connaissance susceptible d'être représenté à l'aide de conventions pour être conservé, traité ou communiqué* ».

[4] propose de considérer la **visualisation** comme : « *L'utilisation de représentations visuelles interactives et informatisées de données pour amplifier la cognition* ». Selon cette approche nous devons prendre en compte les types de données à représenter pour utiliser des techniques de visualisation spécifiques. La plupart des auteurs subdivisent ce domaine en trois classes selon la nature des données [5] : scientifique, géographique et d'informations. La première classe porte sur des données géo-localisées (*e.g.* image-  
rie médicale), la seconde sur des données géographiques (*e.g.* données sur une carte géographique) et la troisième sur des données abstraites (*e.g.* textes, graphes ou matrices).

Dans le cadre de ce projet, nous considérons la **visualisation d'informations** comme la : « *présentation visuelle d'un ensemble d'informations issues de données abstraites et traitées par des moyens informatiques* » [1]. La difficulté de la visualisation d'informations est de déterminer quelle représentation choisir pour bien répartir les données abstraites dans l'espace et montrer un message cohérent et pertinent, basé sur l'exploitation des capacités visuelles et cognitives de l'être humain.

## 2.2 Visualisation de textes et de sentiments contenus dans les textes

De nombreuses approches visuelles utilisant divers métaphores visuelles pour représenter l'information ont été proposées<sup>1</sup>. Parmi les plus utilisées, nous avons :

- **ThemeRiver** [6] : Métaphore visuelle qui représente les thèmes sous la forme d'une rivière, voir figure 2.1. Ici, les thèmes sont représentés par la taille du flot et l'abscisse représente le temps (*e.g.* *OpinionFlow* [7], *EvoRiver* [8], *TextFlow* [9] suivent l'évolution des thèmes, pour montrer les tendances globales et aussi pour coder les événements critiques au fil du temps).
- **Galaxie et ThemeScape** [10] : Le but ici est de transformer les informations du texte en nouvelles représentations visuelles qui révèlent les tendances et les motifs thématiques ainsi que leurs relations. Plus précisément, la métaphore *Galaxie* utilise la spatialisation 2D (voir figure 2.2) pour créer des clusters de points dans l'espace qui suggèrent la similarité de contenu et révèlent les tendances des thèmes. La métaphore *Themescape* utilise 3 dimensions (voir figure 2.3) pour montrer les thématiques dominantes comme des pics, la proximité entre les pics montre la similarité du contenu. Cette approche est utile pour analyser les tendances d'un corpus de documents. En outre, les deux approches offrent des résultats qui améliorent la navigation et l'analyse visuelle.

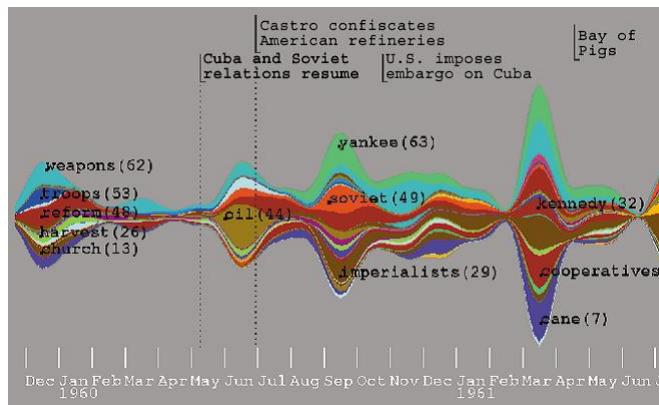
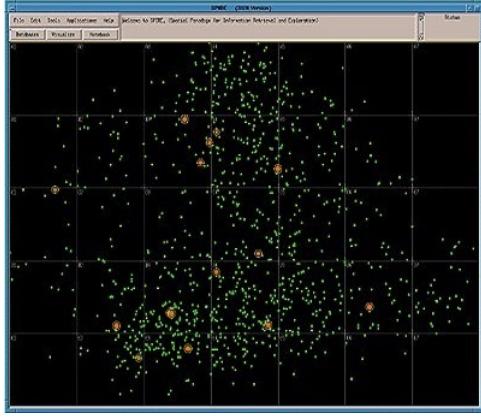
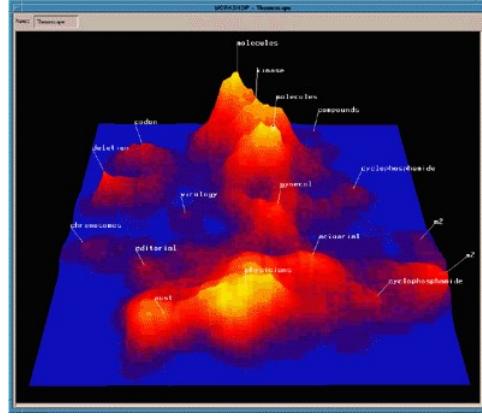


FIGURE 2.1 : *ThemeRiver* [6].

La visualisation de sentiments issus de textes contient généralement trois étapes successives : d'abord, identifier les attributs (c'est-à-dire, les noms, noms composés, *etc.*), deuxièmement identifier le sentiment de ces attributs (généralement positif, neutre, négatif) et finalement projeter dans une carte de sentiments ces attributs [11]. Il est aussi utile d'identifier la force de ces attributs (des sentiments). De nouvelles méthodes sont basées sur une analyse plus précise des tweets pour distinguer des émotions humaines

<sup>1</sup><http://textvis.lnu.se/> : Ce site présente des techniques de visualisation, élaborées par le groupe ISOVIS. [Online ; accessed 27-mai-2015]

FIGURE 2.2 : *Galaxie* [10].FIGURE 2.3 : *Themescapes* [10].

plus complexes comme la tristesse, le bonheur, le stress, la nervosité, *etc.*, comme *SentiCompass* [3] ou *WebSentiment* [12].

### 2.3 Techniques d'interaction visuelle

*« Interaction techniques allow the data analyst to directly interact with the visualizations and dynamically change the visualizations according to the exploration objectives, and they also make it possible to relate and combine multiple independent visualizations »* [13]. Ils représentent la manière d'utiliser une application pour réussir une tâche sur l'espace physique de la visualisation d'informations. Ainsi, il est impératif d'aborder les techniques d'interaction en complément de techniques de projection visuelles. Sans des techniques d'interactions adaptées, les outils visuels ne peuvent parfois pas révéler tout leur potentiel.

Parmi les types de représentations, nous avons les représentations manipulables, c'est-à-dire ceux qui nous permettent de manipuler le processus de génération de vues. (*e.g.* zoom, rotation, défilement). [13] décrit ces techniques comme « *dynamic* » et « *interactive* » selon que les modifications apportées à la visualisation des données sont effectuées automatiquement ou manuellement :

- **Projections dynamiques** : il s'agit de changer dynamiquement les projections afin d'explorer un ensemble de données multidimensionnelles.
- **Filtrage interactif** : il s'agit d'avoir, d'une part, la possibilité de diviser interactivement l'ensemble des données dans des segments et, d'autre part, de se concentrer sur les sous-ensembles intéressants.
- **Zoom interactif** : il s'agit de partir d'une vue globale des données et de permettre l'affichage des détails selon différentes résolutions.
- **Focus + Contexte** (voir Figure 2.4) : comme le mentionnent Card *et.al* [4] « *Focus+Context start from three premises : First, the user needs both overview*

(context) and detail information (focus) simultaneously. Second, information needed in the overview may be different from that needed in detail. Third, these two types of information can be combined within a single (dynamic) display, much as in human vision ».

- **Distorsion interactif** : il s'agit de pouvoir montrer des parties de données avec un niveau élevé de détail tandis que la vue d'ensemble est préservée. Il existe des techniques de déformation hyperboliques ou sphériques souvent employées sur des hiérarchies ou des graphiques. Ces techniques utilisent un fisheye (une sorte de loupe) déformant que l'on promène à son gré sur l'ensemble des données.
- **Fishey** (voir Figure 2.5) : « A fisheye view of a graph shows an area of interest quite large and with detail, and shows the remainder of the graph successively smaller and in less detail. It achieves this smooth integration of local detail and global context by repositioning and resizing elements of the graph » [14].

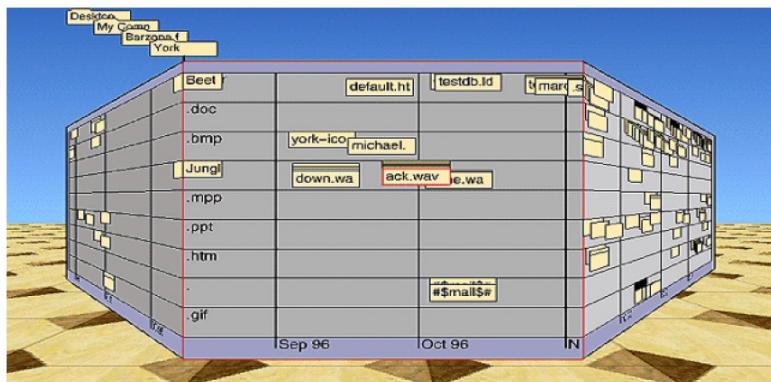


FIGURE 2.4 : Technique Focus + Contexte, qui montre des régions du graphique de une manière plus détaillée (*i.e.* focus) et aussi une vue d'ensemble au détail réduit (*i.e.* contexte) simultanément.

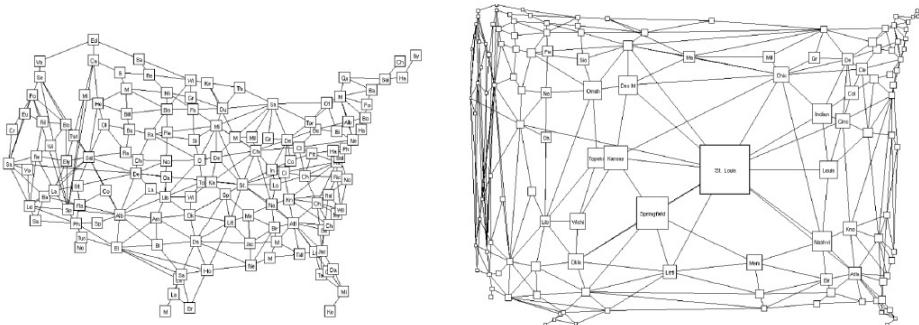


FIGURE 2.5 : A gauche un graphique représentant le métro de Paris. A droite le même graphique utilisant la technique de fisheye [14].

## 2.4 Processus général

Comme tout programme informatique, la visualisation d'informations possède sa propre chaîne de traitements de données. En utilisant le modèle de référence [4] (appelé « *pipeline infovis* », voir la Figure 2.6) nous allons représenter le processus de génération de représentations graphiques. Au départ des **données brutes** sont collectées généralement à l'aide d'un procédé automatisé. Ensuite ces données sont organisées, nettoyées structurées, afin d'obtenir des **tables de données**; *i.e.* un ensemble de relations entre variables et valeurs. Cette structure peut alors être associée à une **structure visuelle** (ensemble d'objets graphiques) par projection visuelle (*visual mapping*). Finalement, la représentation visuelle peut-être manipulée de manière interactive par l'utilisateur afin d'obtenir différentes **vues** à partir d'un même ensemble de données brutes. Sur ce modèle, la phase essentielle est la projection visuelle, qui permet de passer d'un monde de données à un monde de formes visuelles.

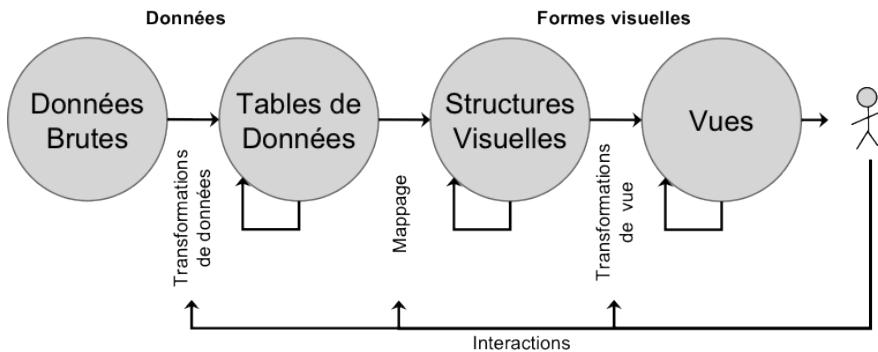


FIGURE 2.6 : Le « *pipeline infovis* » de la visualisation d'informations [4].

A partir de ces vues il devient possible d'analyser plus facilement les différents données. L'analyse visuelle (*visual analytics*) est définie dans [15] comme : « *La science qui étudie les mode de raisonnement basés sur des interfaces visuelles interactives. Un outil d'analyse visuelle est un outil d'application de cette science* ». Sur la figure 2.7 nous voyons l'intégration entre la visualisation et les méthodes d'analyse automatique de données. L'analyse visuelle favorise l'évaluation constructive, la correction et l'amélioration de processus et modèles.

Shneiderman [16] propose un modèle d'analyse visuelle de données « *Information-seeking Mantra* » dans un processus en trois phases : « **Overview first - zoom and filter - details on demand** ». Toute d'abord, dans la **vue d'ensemble** l'utilisateur se fait une idée des données. Après cette étape, en **zoomant et filtrant** il identifie l'information intéressante. Enfin il peut accéder au **détail** des données. Plus récemment Keim *et al.* [17] proposent une approche plus proche de la manière actuelle de travailler : « **Analyse first - show the important - zoom - filter and analyse further - details on demand** ».

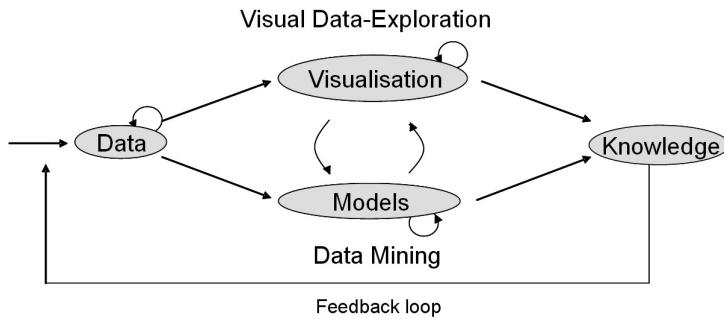


FIGURE 2.7 : Intégration entre la visualisation et les méthodes d'analyse automatique de données.

## 2.5 Modèle de création de visualisations

Nous allons utiliser le modèle imbriqué de Munzner [18] (voir Figure 2.8). Celui-ci propose des étapes de création de visualisations afin de limiter les risques de « hors - sujet ».

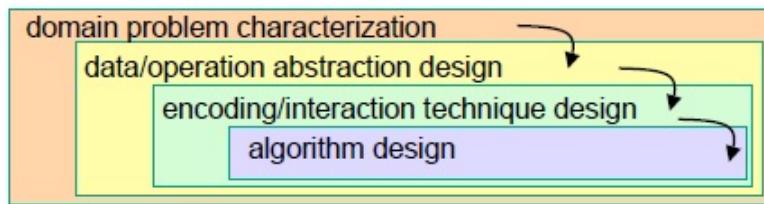


FIGURE 2.8 : Modèle de création de visualisations de Munzner [18].

Les différents étapes de ce modèle sont les suivantes. Tout d'abord nous partons de **la caractérisation des problèmes du domaine cible**. Il s'agit d'apprendre et comprendre la problématique du domaine. L'étape suivante est **l'abstraction des données et des opérations**. Cette étape consiste à transformer les données brutes en données abstraites que les techniques de visualisation peuvent gérer. La troisième étape est **la conception des techniques de représentation et d'interaction** qui définit les modes d'interaction et les modes de représentation. Finalement **la conception des algorithmes** consiste à trouver des algorithmes pour mettre en place les modes définis à l'étape précédente.

Dans ce chapitre, nous avons présenté les concepts généraux de la visualisation d'informations. Le chapitre suivant s'intéresse aux spécificités de la conception de notre approche visuelle.

## Conception d'une nouvelle approche visuelle

Dans ce chapitre, nous reprenons le modèle imbriqué de Munzner [18] qui propose des règles pour la création de visualisations. Concrètement, nous allons commencer ce chapitre par la caractérisation de la problématique que notre application devra résoudre. Ensuite nous décrirons le paradigme visuel que nous avons conçu afin de répondre à cette problématique. Enfin, nous décrivons les techniques conçues pour mettre en place ce paradigme.

### 3.1 Caractérisation de la problématique

Des tweets peuvent être visualisés de nombreuses façons : par le sentiment, par le sujet, par des termes fréquents, *etc.* La visualisation des sentiments au fil du temps a pour but de comprendre comment les gens expriment des émotions à travers des textes et comment ces textes déclenchent des émotions différentes. Certaines approches visuelles (*e.g.* *TwitInfo* [19], *OpinonSeer* [2]) ont déjà été développées dans le domaine de l'abstraction de sentiments comme une variable unidimensionnel (*i.e.* négatif ou positif). Cependant, la plupart de ces approches ne prennent pas en compte l'association entre la temporalité des tweets et leur sentiment. Par conséquent, trouver de meilleures façons de visualiser des sentiments au fil du temps est notre défi.

Comme nous l'avons décrit dans le chapitre d'introduction, notre travail se concentre à trouver des améliorations visuelles pour palier les limites de *SentiCompass* [3] et représenter visuellement les grandes tendances des jeux de données. Dans ce cadre, nous avons utilisé comme départ l'étude déjà faite dans *SentiCompass* [3] dans le domaine de la représentation des émotions et de l'analyse des sentiments des tweets. Maintenant, nous allons définir ces concepts et les approches suivies précédemment.

### 3.2 Caractérisation des structures de données et des besoins

#### Représentation des émotions par modèles

En psychologie, Delay *et.al* [20] définit le terme **humeur** comme : « *La disposition affective fondamentale riche de toutes les instances émotionnelles et instinctives qui donne à chacun de nos états d'âme une tonalité agréable ou désagréable, oscillant entre les deux pôles extrêmes du plaisir et de la douleur* ». **L'émotion** peut être définie comme

une certaine forme de plaisir ou de déplaisir. Le **sentiment** est une composante de l'émotion qui traduit un état physiologique. Le but d'un modèle émotionnel est améliorer la visualisation et l'abstraction des sentiments. Ces modèles utilisent souvent des dimensions émotionnelles pour positionner émotions sur un plan 2D. Ils représentent le plaisir sur l'abscisse (de très désagréable à une extrémité à très agréable sur l'autre). Des modèles émotionnels ont été proposés dans des outils pour définir et comparer les états émotionnels (*e.g.* *WebSentiment* [12], *Pearl* [21]).

Certains modèles montrent la multi-dimensionnalité des émotions (*e.g.* Watson *et.al* [22], Tayer [23], Russell [24]). L'un des modèles les plus utilisés, est celui que nous avons pris comme référence dans notre travail, *i.e.* le modèle *circumplex* de Russell [24] (voir Figure 3.1). Ce modèle propose de distribuer les sentiments dans un espace bidimensionnel, l'abscisse étant *valence* (*pleasant - unpleasant*) et l'ordonnée étant *arousal* (*activation - deactivation*). Différentes émotions peuvent être organisées autour du cercle basé sur des combinaisons de ces deux axes. Ce modèle s'adapte à notre travail, principalement parce que les deux principaux axes du sentiment (*valence* et *arousal*) sont les plus couramment utilisés dans les stimuli de mots affectifs [25].

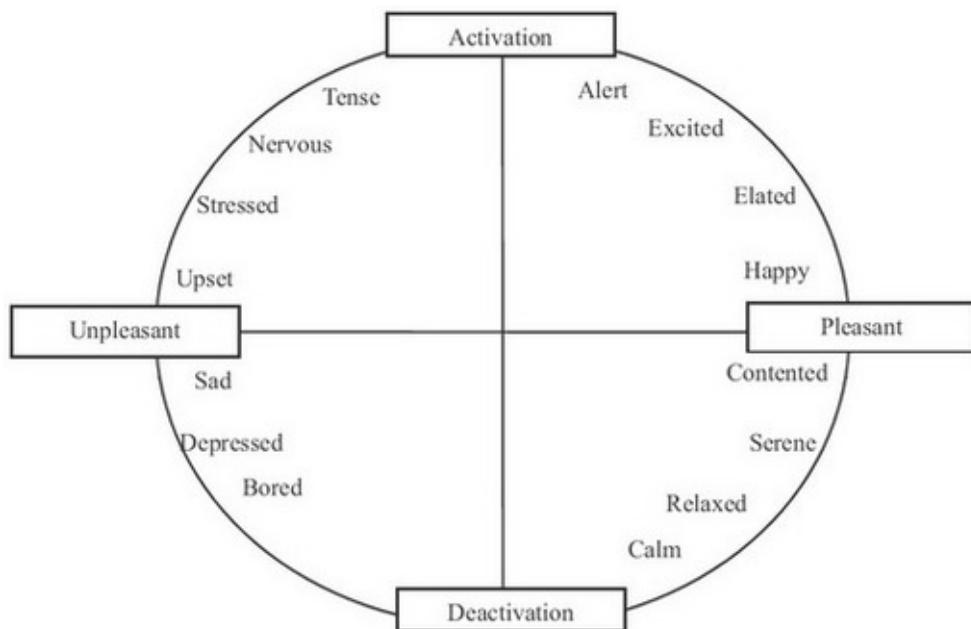


FIGURE 3.1 : Modèle émotionnel *circumplex* de Russell [24].

### Agrégation et désagrégation des émotions

Comme le présente Wikipédia<sup>1</sup>, « *Agrégation désigne l'action d'agréger, de regrouper des éléments* ». Dans notre contexte, nous allons utiliser ce terme pour définir le regroupement des émotions contenus dans un segment du modèle émotionnel *circumplex* de Russell [24]. Par exemple (voir Figure 3.2), lorsqu'un sentiment (*i.e.* émotion) se

<sup>1</sup>Information du site <http://fr.wikipedia.org/wiki/Aggregation> [Online ; accessed 01-juin-2015]

### 3.2. CARACTÉRISATION DES STRUCTURES DE DONNÉES ET DES BESOINS<sup>1</sup>

trouve dans l'espace *happy - elated - excited - alert*, nous devons pouvoir les regrouper en le projetant dans l'espace du *pleasant - activation*. Au contraire de la désagregation, comme le présente Larouse<sup>2</sup>, est « *Séparation des parties qui composent un tout* ». C'est-à-dire, nous devons pouvoir séparer les émotions contenus dans un segment du modèle émotionnel *circumplex* de Russell [24]. Par exemple (voir Figure 3.3), lorsqu'un sentiment se trouve dans l'espace *unpleasant - activation*, nous devons pouvoir le préciser en le projetant dans l'espace *upset - stressed - nervous - tense*.

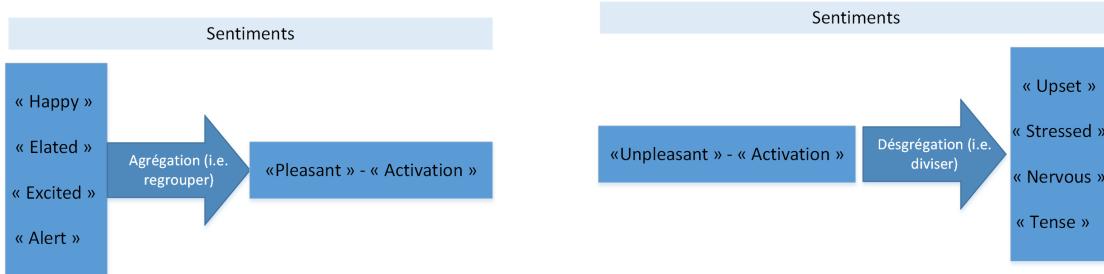


FIGURE 3.2 : Nous montrons l'agrégation des sentiments (*happy - elated - excited - alert*) en (*pleasant - activation*).

FIGURE 3.3 : Nous montrons la désagrégation des sentiments (*unpleasant - activation*) en (*upset - stressed - nervous - tense*)

## Analyse des sentiments

Pour l'analyse d'opinion, une méthode hybride a été choisi précédemment, qui combine un dictionnaire affectif avec une classification traditionnelle supervisée (Naïve Bayes [26]) pour estimer les sentiments des tweets. L'utilisation d'un dictionnaire affectif est une solution pour estimer un sentiment, plus efficace pour de courts extraits de texte comme des commentaire en ligne, conversations de réseaux sociaux (*e.g.* tweets). Ainsi, c'est une méthode qui convient à notre travail.

Parmi les dictionnaires affectif nous avons : POMS (Profile of Mood States) [27], POMS-ex (Extended Profile of Mood States) et ANEW (Affective Norm for English Words) [28]. Maintenant nous allons décrire comment le dictionnaire ANEW a été utilisé pour exprimer les deux dimensions des sentiments (*valence* et *arousal*) dans le travail précédent de *SentiCompass* [3].

ANEW contient des évaluations affectives pour 1034 mots anglais obtenues à partir d'études empiriques. Pour chaque mot, les évaluations consistent à trouver la moyenne et l'écart type pour la *valence*, *arousal* et la *domination*. Pour calculer la *valence* et *arousal* de tweets, des mots de signalisation affectifs (qui sont des mots enregistrés dans ANEW) sont d'abord recherchés et localisés. Puis statistiquement sont pesées la *valence* et *arousal* dans de multiples mots affectifs. Pendant ce temps, pour compenser l'approche du dictionnaire affectif, la polarité a été calculée (*i.e.* négatif ou positif) du tweet en utilisant une classification naïve bayésienne. Finalement, pour chaque tweet, les deux dimensions de *valence* et *arousal* sont estimées, puis coordonnées sur l'espace affectif.

<sup>2</sup>Information du site <http://www.larousse.fr/dictionnaires/francais/desagregation/24166> [Online ; accessed 01-juin-2015]

La figure 3.4 représente une vue d'ensemble du système. Maintenant nous allons décrire notre approche dans la partie de la visualisation d'évolution des sentiments (en rouge sur le schéma).

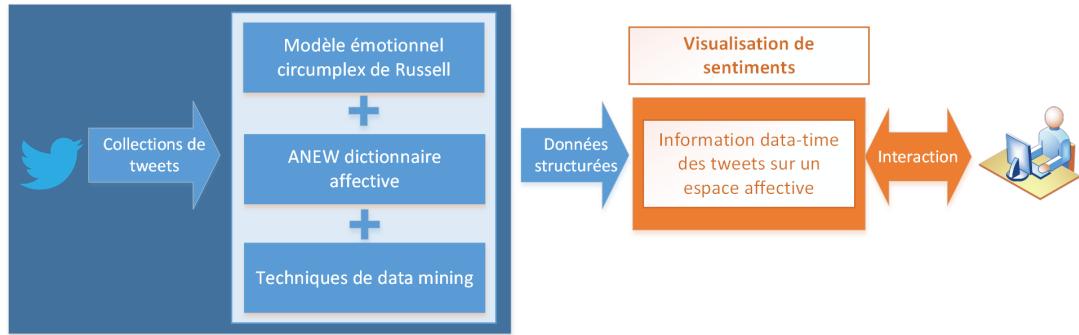


FIGURE 3.4 : Vue d'ensemble du système. Notre travail correspond à la partie visualisation d'évolution des sentiments.

### Besoins de la visualisation

Ici, nous présentons les besoins que notre outil de visualisation devra considérer. Nous devons gérer la temporalité des tweets et leurs sentiments pour arriver à résoudre les limitations des précédentes approches. Les tâches à gérer sont les suivantes :

- Visualiser la distribution des tweets et la force d'un sentiment à un moment donné.
- Visualiser l'ensemble des sentiments pour un segment à un moment donné.
- Visualiser la signification sémantique des tweets basées dans le modèle émotionnel circumplex de Russell [24] à un moment donné.
- Visualiser les sentiments à différent niveau de désagrégation et agrégation à un moment donné.
- Visualiser les sentiments dominants d'un événement à différentes échelles de temps.
- Visualiser les sentiments utilisant divers techniques d'interaction.

Dans ce travail, nous nous efforçons de répondre aux problèmes abordés (information temporelle, représentation des sentiments) en utilisant des techniques visuelles et d'interactions adaptées à nos besoins.

### 3.3 Conception des techniques de représentation et d'interaction

Pendant l'étude bibliographique, nous avons remarqué certains paradigmes visuels utilisés pour la visualisation de sentiments contenus dans les textes. Après l'analyse de

la problématique dans la section précédente, nous avons choisi comme métaphore visuelle *ThemeRiver* [6]. Basé sur les besoins de la visualisation, nous devrons résoudre le problème de la représentation temporelle des sentiments sur les tweets, ainsi, la métaphore de la « rivière » s'adapte à notre contexte, car ce paradigme permet de représenter horizontalement l'évolution des thèmes au fil du temps (abscisse) et la force relative des thèmes par le largeur du flux.

Le paradigme visuel *ThemeRiver* [6] utilise Stacked Graph [6] comme technique de visualisation, dans notre approche nous allons utiliser une dérivation de cela, appelée *Stream Graph* [29]. Nous allons maintenant voir plus en détails l'utilisation de cette technique comme outil de modélisation de notre approche visuelle.

### Visualisation de l'évolution des sentiments

Grâce à la popularité de Twitter et à la facilité d'accès aux contenus textuels, plusieurs travaux se sont intéressés à l'analyses des tweets (*e.g.* *TwitInfo* [19], *Civitas* [30]). La contrainte de la taille à 140 caractères dans les tweets encourage les utilisateurs à construire des mises à jours dans le temps gérer le flux d'un événement. Par exemple, pendant la finale de la coupe du monde Brésil 2014, le coup de sifflet final marqua un nouveau record du nombre de tweets par minute sur un événement sportifs. Ce qui nous intéresse ici est de représenter le déroulement de l'évolution des sentiments des tweets lors d'un événement. Pour cela nous devons trouver une métaphore visuelle qui permet d'exploiter tous ses caractéristiques sans trop de surcharge cognitive. Basé sur le modèle émotionnel *circumplex* de Russell [24], notre outil doit prendre compte la variation du temps des tweets et l'espace affectif à laquelle ils appartiennent, ce qui nous permettrait par exemple de savoir quelles sont les émotions les plus prédominantes lors d'un événement.

### Visualisation avec Stream Graph

Nous allons définir **Stacked Graph** [6] comme un graphique qui représente une variable quantitative contre une autre variable quantitative (généralement le temps comme variable indépendante, sur l'abscisse). Les différentes zones sont empilées l'une sur l'autre, de manière à ce que la hauteur de chaque zone représente la valeur pour chaque variable, et la hauteur totale leur somme. **Stream Graph** [29] (voir Figure 3.5), est une type de Stacked Graph autour d'un axe central, donnant une forme fluide à la visualisation.

Maintenant voyons comment combiner la technique de visualisation *Stream Graph* [29] avec la modélisation de notre approche visuelle. Chaque flux représente un segment dans l'espace du modèle émotionnel *circumplex* de Russell [24], nous allons représenter les flux sur l'abscisse. La longueur des flux correspond à la durée d'un événement et la largeur représente le nombre de tweets qui appartiennent à ce segment spécifique du modèle émotionnel. Nous affrontons le problème de la désagrégation et agrégation des sentiments comme une interpolation des flux, c'est-à-dire, un flux peut être divisé pour montrer les sentiments contenus dans ce segment du modèle émotionnel.

*Stream Graph* [29] nous permettre d'afficher de grands ensembles de données [29]. Pour bien exploiter tout son potentiel, il est nécessaire de le munir de techniques d'interaction, permettant aux utilisateurs d'explorer en détail chaque élément du flux. Pour

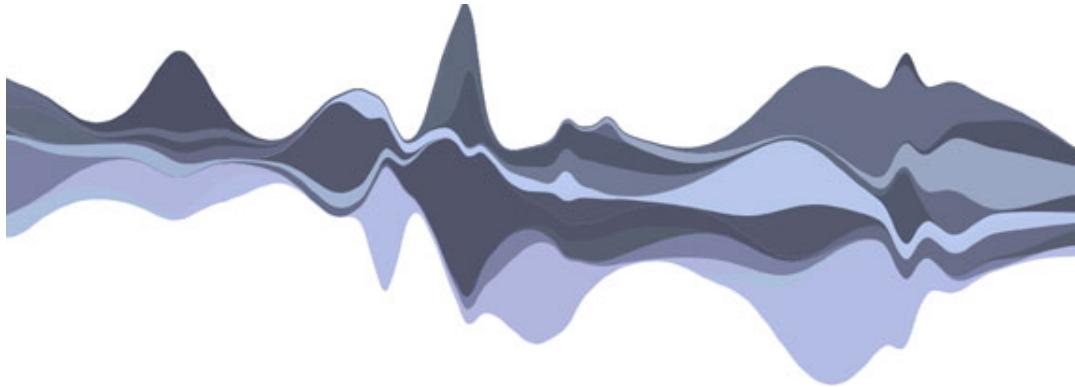


FIGURE 3.5 : Exemple de Stream Graph [29].

approfondir sur un niveau macro et micro, nous allons mettre en place des techniques d’interaction, telles que le filtrage, la sélection, le zoom.

### Représentations visuelles

Maintenant nous allons proposer les projections visuelles pour représenter la métaphore *ThemeRiver* [6] dans notre approche. Nous allons donner des couleurs sur les flux de notre approche pour communiquer le segment dans l'espace du modèle émotionnel de Russell [24]. Nous utilisons deux paires de couleurs complémentaires (voir Figure 3.6) pour représenter les deux dimensions du sentiment : *valence* (vert au rouge, indiquant le niveau de *pleasant*) et *arousal* (bleu au jaune, indiquant le niveau *activation*). Les couleurs froides comme le bleu et le vert indiquent un niveau faible, tandis que les couleurs chaudes telles que le rouge et le jaune indiquent un niveau haut.

En utilisant une interpolation des couleurs, nous allons représenter l’agrégation et désagrégation des sentiments dans les flux (voir Figure 3.7). Dans cet exemple, nous pouvons voir comment se réalise une transition entre des flux pour montrer plus en détails quelles sont les sentiments contenus dans le segment du modèle de Russell [24] (*i.e. désagrégation*).

### Techniques d’interaction

Nous allons mettre en place trois types de techniques de visualisation : focus + contexte, zoom et fisheye. En utilisant le focus + contexte (voir Figure 3.8) nous gardons une vue d’ensemble de tous les flux d’un événement au-dessous de l’écran (*i.e. contexte*), tout en permettant de visualiser en détail une partie des données au-dessus de l’écran (*i.e. focus*). Nous avons mis en place, une fonctionnalité de filtrage interactif (*fenêtre glissante*) dans le contexte pour nous déplacer et générer sur le focus un zoom et une zone de transition. Cette zone de transition va nous aider pour représenter visuellement la transition des flux au moment de l’agrégation et désagrégation. Dans le prochain chapitre, nous allons décrire des algorithmes mis en place pour ces techniques.

Le système d’interaction mis en place ici est innovant, car il montre des techniques simples qui peuvent être exploitées ensemble pour donner plus d’interaction aux utilisateurs.

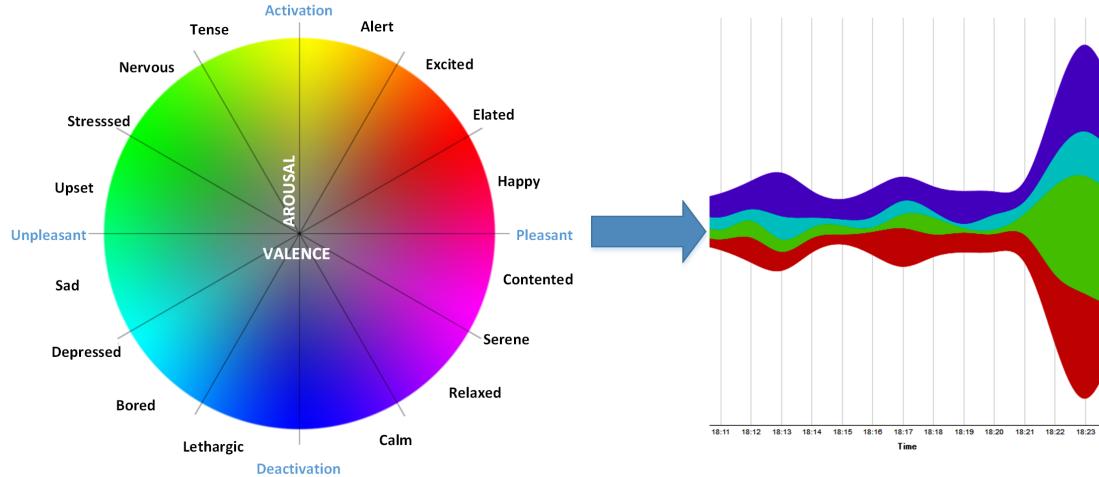


FIGURE 3.6 : Représentation des couleurs. A gauche, une représentation visuelle du modèle émotionnel *circumplex* de Russell [24]. A droite, notre approche visuelle en représentant les couleurs selon le segment émotionnel.

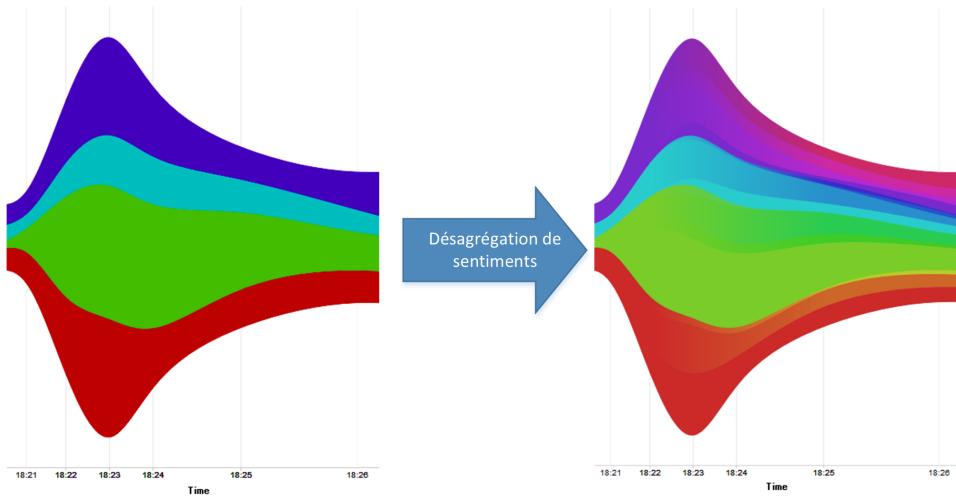


FIGURE 3.7 : Notre outil peut gérer divers niveaux de segmentation de sentiments, pour passer en fluidité parmi ces segments nous utilisons une interpolation des couleurs. Dans cet exemple, il s'agit d'une désagrégation des sentiments dans les flux. Nous observons comme les couleurs se divisent au moment de la transition.

teurs de notre outil. En utilisant le filtrage interactif dans la vue contexte nous pouvons générer d'interactions intéressantes sur la vue focus. Par exemple, si nous mettons l'intervalle de la partie zoom à 0 nous obtenons une vue de type fisheye sur le focus, néanmoins si nous mettons les intervalles des parties normales et fisheye à 0 nous obtenons une vue Focus + Contexte classique (*i.e.* partie zoom en haut et contexte en bas).

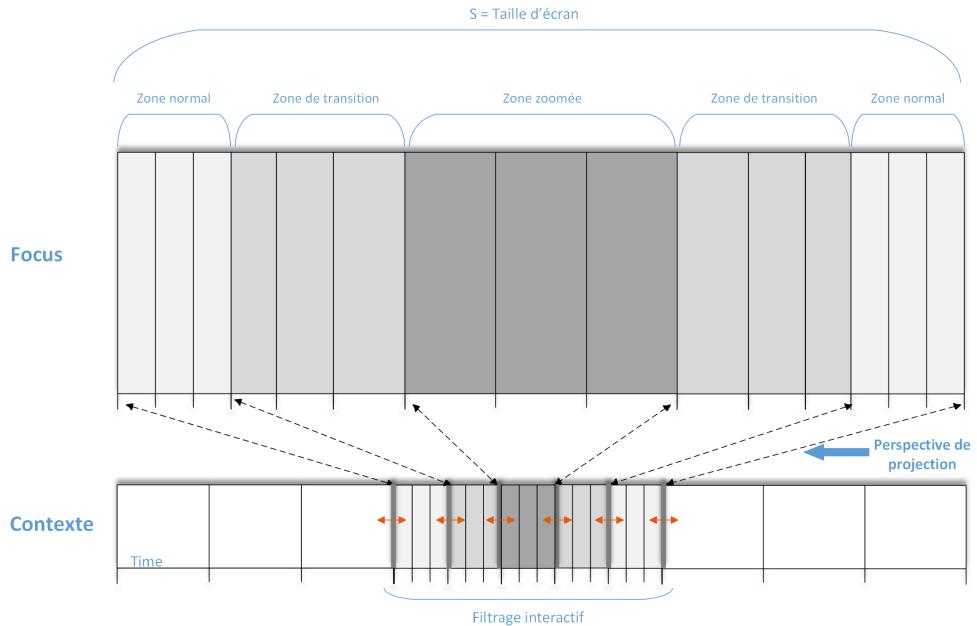


FIGURE 3.8 : Technique Focus + Contexte utilisé dans notre approche. Sur la vue contexte, nous avons développé un filtrage interactif, qui permet de diviser interactivement l'ensemble des données dans des segments. Nous utilisons une fenêtre glissante pour gérer cette partie. La vue focus montre la projection de ces données par rapport à la fenêtre glissante.

## Maquette

Des utilisateurs disposent d'une interface graphique en mode web pour gérer notre outil. De manière plus précise, nous avons décrit les caractéristiques de notre application sur la Figure 3.9 :

1. **Contrôle de temporalité** : Nous fournissons de cette fonctionnalité pour permettre la sélection des segmentations de temps à granularité souhaités.
2. **Vue contexte** : En utilisant la technique Focus + Contexte, nous fournissons de la vue contexte pour permettre une exploration dans tout le jeu de données. Ce point de vue nous permet de trouver des motifs facilement.
3. **Vue focus** : Nous fournissons de la vue focus pour visualiser plus en détail une partie des données, c'est ici que nous avons intégré des zones avec différent niveaux d'agrégation de sentiment. Une zone de transition (*i.e.* fisheye) (couleur de fond gris foncé) qui permet de passer de la zone normal (couleur de fond gris clair) à la zone zoomée (couleur de fond fonde gris).
4. **Filtrage interactif** : Nous fournissons d'une fenêtre glissante pour diviser interactivement l'ensemble des données dans des segments. Cette fenêtre nous permet de naviguer sur la vue contexte et contrôler les données sur la vue focus.

5. **Désagrégation des sentiments** : Nous permet voir une transition des couleurs pour diviser les segmentes des sentiments du modèle émotionnel *circumplex* de Russell [24].
6. **Agrégation des sentiments** : Nous permet voir une transition des couleurs pour regrouper les segmentes des sentiments du modèle émotionnel *circumplex* de Russell [24].
7. **Segments des sentiments** : Nous pouvons voir divers niveaux de segmentation de sentiments sur la zone zoomée (couleur de fond gris).
8. **Tool tip** : Nous fournissons d'un tool tip pour naviguer sur la vue focus et de montrer des sentiments contenus dans les flux.

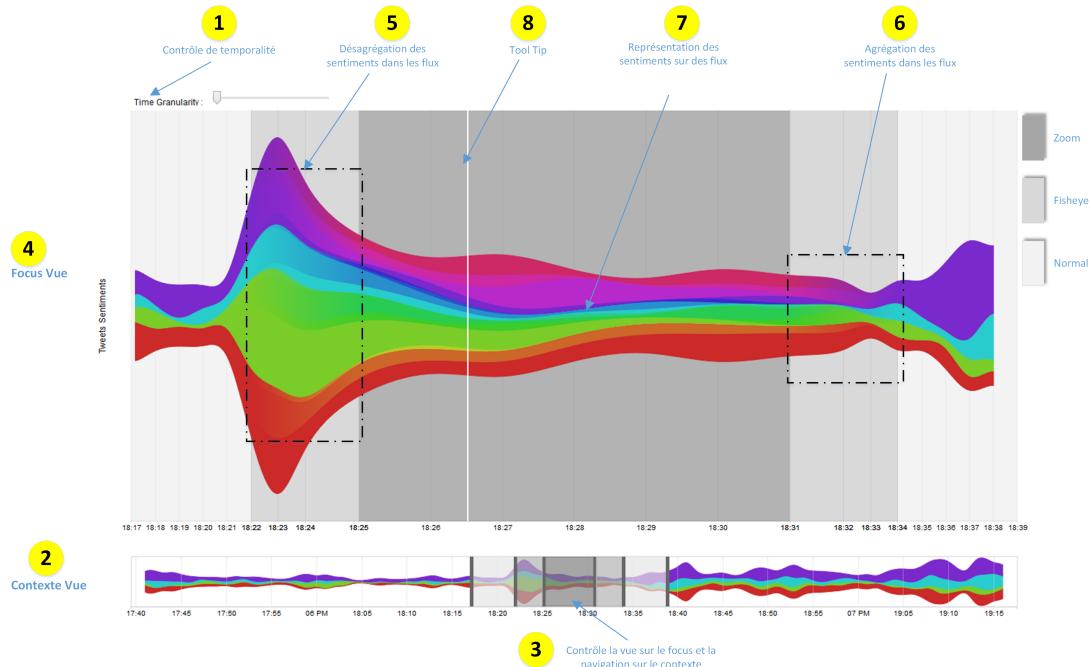


FIGURE 3.9 : Maquette de notre outil.



## Mis en œuvre

Dans ce chapitre, nous décrivons les algorithmes mis en œuvre pour créer les méthodes d'interaction. Plus précisément, nous présentons un algorithme conçu pour gérer la zone de transition sur la vue focus de notre application.

### 4.1 Conception des algorithmes

Sur la Figure 4.1, nous regardons la vue focus sur laquelle nous avons développé des zones pour montrer des données avec un niveau élevé de détail. Ces zones sont gérées depuis la vue contexte (*i.e.* filtrage interactif). Nous allons définir les concepts utilisés :

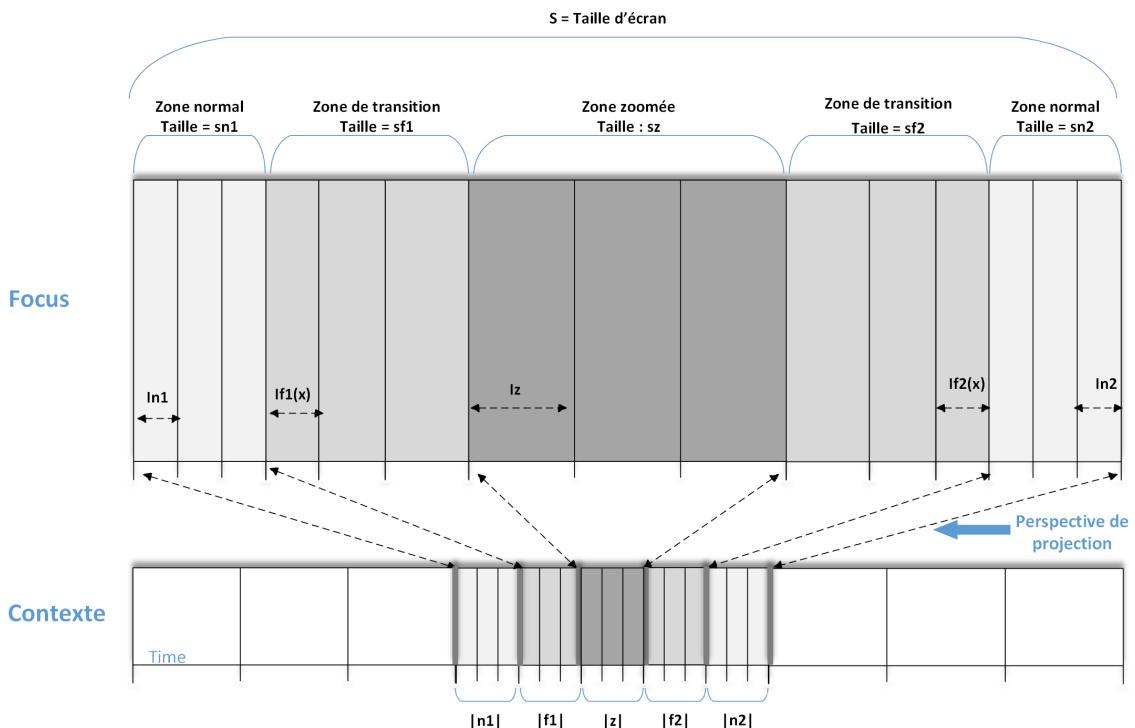


FIGURE 4.1 : Technique Focus + Contexte, en montrant en détails leur zones.

- Variables dans la vue focus

- $S =$  c'est la taille d'écran, c'est-à-dire, le nombre de pixels que peut afficher la vue focus sur un écran.

- **Variables dans la vue contexte**

- $|n_1| =$  nombre d'intervalles de temps dans la zone  $normal_1$ .
- $|n_2| =$  nombre d'intervalles de temps dans la zone  $normal_2$ .
- $|f_1| =$  nombre d'intervalles de temps dans la zone  $fisheye_1$ .
- $|f_2| =$  nombre d'intervalles de temps dans la zone  $fisheye_2$ .
- $|z| =$  nombre d'intervalles de temps dans la zone  $zoom$ .

Il n'y aucune contrainte dans la taille d'écran ou dans le nombre d'intervalles. Ainsi, des utilisateurs peuvent choisir leurs intervalles dans la vue contexte selon leur besoin.

### Projeter la temporalité

Nous calculons le nombre d'intervalles de temps à projeter dans la vue focus par rapport aux intervalles choisis dans la vue contexte. Cet intervalle est la somme de toutes les zones que conforment notre filtrage interactif (*i.e.*  $|n_1|$ ,  $|n_2|$ ,  $|f_1|$ ,  $|f_2|$ ,  $|z|$ ). La variable  $N$  représente ce calcul dans la formule suivante :

$$N = |n_1| + |n_2| + |f_1| + |f_2| + |z|$$

### Projeter la proportionnalité

Pour projeter les zones dans la vue focus par rapport à la vue contexte, nous allons nous servir de constantes de projection. Constante  $\alpha$  pour gérer la zone zoomée,  $\beta$  pour gérer les zones de fisheye et  $\gamma$  pour gérer les zones normales. La contrainte :  $\alpha > \beta > \gamma$  nous permettra d'avoir une projection toujours proportionnel à nos besoins (*i.e.* la zone zoomée plus grande que les zones de fisheye, et ceux-ci plus grandes que les zones normales). Voyons, le calcul des proportions dans la vue focus :

$$\begin{aligned} pn_1 &= \alpha \cdot \frac{|n_1|}{N} \\ pn_2 &= \alpha \cdot \frac{|n_2|}{N} \\ pf_1 &= \beta \cdot \frac{|f_1|}{N} \\ pf_2 &= \beta \cdot \frac{|f_2|}{N} \\ pz &= \gamma \cdot \frac{|z|}{N} \end{aligned}$$

$P$  représente la somme de tous ces proportions par rapport à la taille d'écran :

$$P = pn_1 + pn_2 + pf_1 + pf_2 + pz$$

### Taille des zones

Maintenant, nous allons calculer la taille des zones dans la vue focus (*zone zoomée, zones de fisheye et zones normales*), par rapport à la taille de l'écran (*variable S*) et sa proportionnalité :

$$\begin{aligned} sn_1 &= pn_1 \cdot \frac{S}{P} \\ sn_2 &= pn_2 \cdot \frac{S}{P} \\ sf_1 &= pf_1 \cdot \frac{S}{P} \\ sf_2 &= pf_2 \cdot \frac{S}{P} \\ sz &= pz \cdot \frac{S}{P} \end{aligned}$$

Comme résultat, nous obtenons la taille en pixels de chaque zone que nous montrerons dans la vue focus.

### Taille des intervalles de temps dans les zones zoomées et normales

Pour trouver la taille des intervalles de temps dans ces zones, nous prenons la taille de chaque zone et nous la divisons par le nombre d'intervalles à montrer :

$$\begin{aligned} In_1 &= \frac{sn_1}{|n_1|} \\ In_2 &= \frac{sn_2}{|n_2|} \\ Iz &= \frac{sz}{|z|} \end{aligned}$$

Le résultat est la taille en pixels de chaque intervalle de temps dans la zone zoomée et les zones normales.

### Taille des intervalles dans la zone de fisheye

Dans notre approche, nous implémentons une zone de fisheye *i.e.* une zone de transition entre la zone normal et la zone zoomée (*i.e.* désagrégation des sentiments) et entre la transition de la zone zoomée et la zone normal (*i.e.* agrégation des sentiments). Le but de la zone fisheye est donner aux utilisateurs un espace de transition au moment de l'interpolation des flux.

Pourtant, les tailles des intervalles de temps dans cette zone ne sont pas égales, voici l'équation d'une fonction croissante pour donner cet effet :

$$g(x) = sf_i \cdot \frac{C^{(x-1)}}{C^{|f_i|-1}} \quad (4.1)$$

Où :

- La variable  $i$  représente la zone de fisheye  $[1, 2]$
- La variable  $x$  représente le nombre d'intervalles de temps dans la zone de fisheye  $[1, |f_i|]$
- La constante  $C$  représente le niveaux de croissance de la fonction.

Maintenant, nous allons calculer les intervalles de temps dans la zone de fisheye :

$$\forall x \in [1, |f_i|]$$

$$If_i(x) = \begin{cases} g(x) & \text{si } x = 1 \\ g(x) - g(x-1) & \text{sino;} \end{cases}$$

## 4.2 Étude de cas

Pour notre cas d'étude nous allons utiliser des tweets collectés lors d'un événement sportif et lors d'un événement politique. Nous montrerons les principales caractéristiques d'interaction et de visualisation de notre outil. Nous verrons comme les besoins de visualisation définies précédemment sont remplis (voir Section 3.2).

### Cas 1

Le premier cas d'étude porte sur un match de rugby entre les équipes du Pays de Galles et de Nouvelle Zélande, des tweets ont été collectés le 22 novembre 2014, de 17h40 à 19h17 (UTC). Dans la Figure 4.2, nous avons défini une granularité temporelle de 1 minute, car la quantité des tweets correspondent à un événement de quelques heures. La vue de contexte nous permet voir qu'il y a un pic des tweets entre 18h20 et 18h27, cette période correspond au début de la seconde moitié du match<sup>1</sup> et nous voyons que les sentiments plus dominants se trouvent dans le segment *activation - unpleasant*. En utilisant la vue focus nous *zoomons* sur ce segment pour le désagréger. Nous pouvons ainsi voir que le sentiment dominant est *tense*. Cela correspond au registre du match, parce que les deux équipes étaient à l'égalité de points à ce moment.

Dans la Figure 4.3 en utilisant notre vue focus sur tout le jeu de données nous allons désagréger tous les segments des sentiments, et maintenant nous voyons certains motifs des sentiments dans le match. Par exemple, nous remarquons que depuis le début du match (17h30) le segment du sentiment *bored* reste faible et constant, mais après la seconde moitié du match (18h35) ce sentiment commence à grandir. Nous observons aussi que le sentiment *tense* augmente seulement au début de la seconde moitié et après retourne à son état initial.

### Cas 2

Notre deuxième cas d'étude porte sur la période électorale australienne 2013. Nous avons des tweets du 04 au 09 septembre (GMT +10), qui prennent en compte une période de temps avant et après du jour même des élections, le 07 septembre. Dans ce cas, nous avons

---

<sup>1</sup><http://www1.skysports.com/rugbyunion/live/match/66561/report> [Online ; accessed 23-mai-2015]

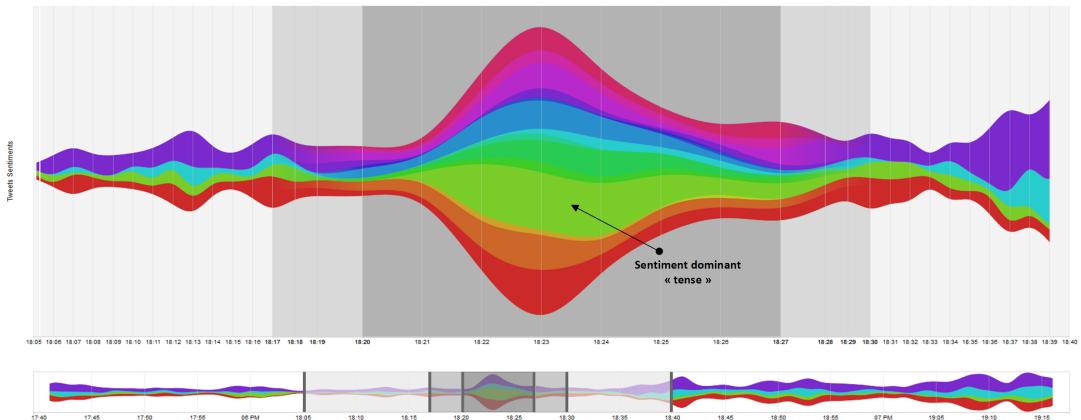


FIGURE 4.2 : Cas d'étude (Jeu de Rugby. Pays de Galles vs de Nouvelle Zélande).

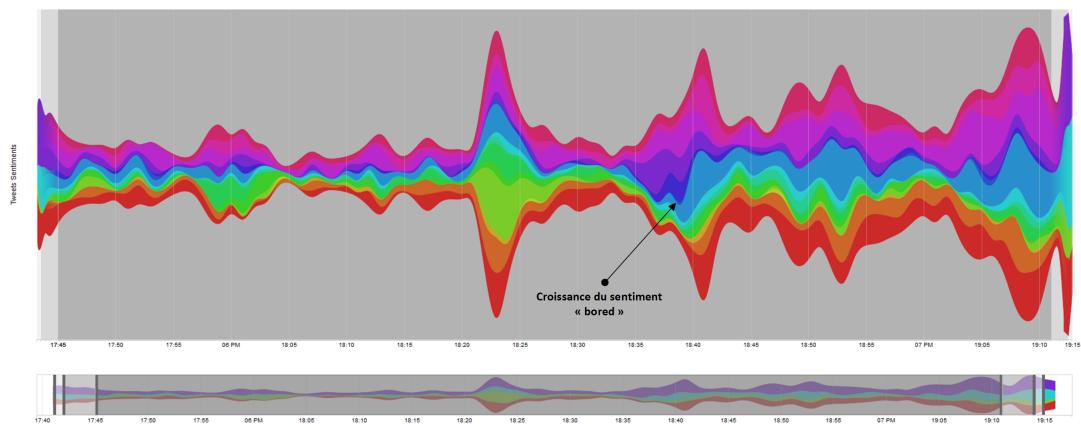


FIGURE 4.3 : Cas d'étude (Jeu de Rugby. Pays de Galles vs de Nouvelle Zélande).

défini la granularité temporelle à 1 heure, parce que le nombre de tweets appartiennent à plusieurs jours. Dans la Figure 4.4 la vue contexte montre la distribution des tweets pendant toute le période électorale, nous pouvons voir comme dans le jour 07 septembre le volume de tweets est plus élevé car c'est le jour des élections. La vue focus nous montre un zoom des jours avant les élections (04 à 06 septembre), nous pouvons voir une segmentation homogène des sentiments, cependant nous observons que les sentiments *stressed* et *nervous* sont plus élevés avant du jour des élections.

Maintenant, nous parlerons des sentiments dans le jour des élections. Dans la Figure 4.5 nous pouvons voir comme les sentiments du segment *pleasant* et *deactivation* restent constants, par exemple (*calm*, *relax*). Tandis que le segment *pleasant* et *activation* commence à augmenter, nous trouvons ici que le sentiment *elated* a la plus grande présence. Après le jour des élections, les flux de sentiments reviennent à une distribution homogène.

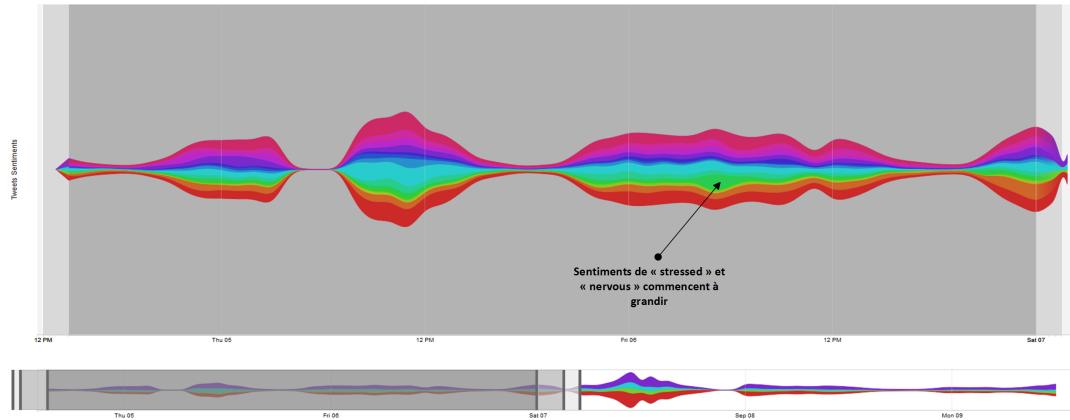


FIGURE 4.4 : Cas d'étude (Période électorale australienne 2013).

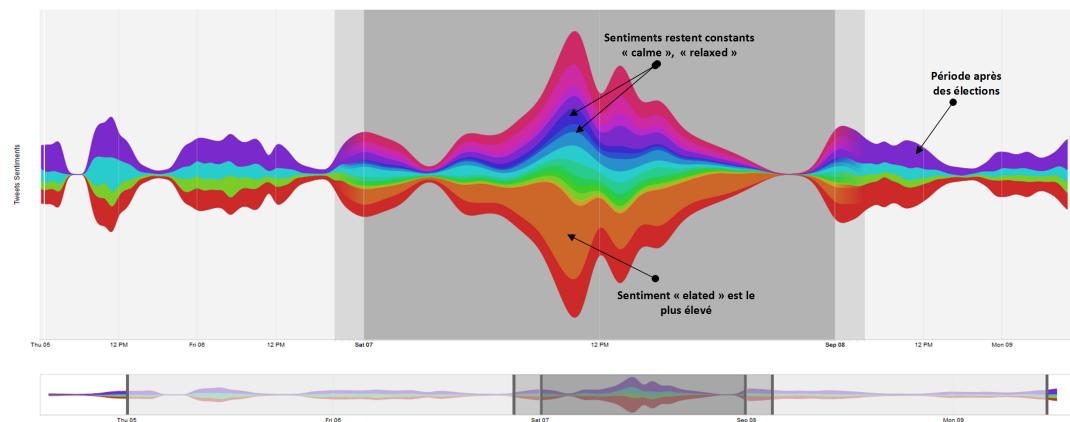


FIGURE 4.5 : Cas d'étude (Période électorale australienne 2013).

## Conclusions

Actuellement la visualisation de sentiments des *tweets* est un domaine en croissance. Il est donc impérative d'explorer de nouveaux outils pour traiter ce problème. Les travaux présentes dans ce mémoire ont pour but de donner une solution visuelle aux utilisateurs pour leur permettre d'analyser l'évolution des sentiments sur Twitter. Nous avons présenté la conception d'un nouvel outil qui utilise divers techniques d'interaction et de visualisation pour discerner les sentiments exprimés lors d'un événement. Sur la base des travaux existants (*SentiCompass* [3]), nous avons réussi à ajouter des fonctionnalités innovantes permettant une intégration de vues compatibles et facilement appréhendables par l'utilisateur. Nous avons fait une étude des principales métaphores visuelles, pour proposer un travail qui donne aux utilisateurs les moyens d'interagir avec notre approche pour trouver des motifs au fil du temps, qui pourraient être utilisés pour des études dans d'autres domaines.



## Bibliographie

- [1] Arnaud Sallaberry. *Visualisation d'information : de la théorie sémiotique à des exemples pratiques basés sur la représentation de graphes et d'hypergraphes*. PhD thesis, Université Sciences et Technologies-Bordeaux I, 2011.
- [2] Yingcai Wu, Furu Wei, Shixia Liu, Norman Au, Weiwei Cui, Hong Zhou, and Huamin Qu. OpinionSeer : interactive visualization of hotel customer feedback. *IEEE Transactions on Visualization and Computer Graphics*, 16(6) :1109–1118, 2010.
- [3] Florence Ying Wang, Arnaud Sallaberry, Karsten Klein, Masahiro Takatsuka, and M Roche. SentiCompass : Interactive visualization for exploring and comparing the sentiments of time-varying twitter data. In *Proceedings of the 8th IEEE Pacific Visualization Symposium (PacificVis 2015)*, pages 129–133, 2015.
- [4] Stuart K Card, Jock D Mackinlay, and Ben Shneiderman. *Readings in information visualization : using vision to think*. Morgan Kaufmann, 1999.
- [5] R. Spence. *Information Visualization : Design for Interaction*. Pearson/Prentice Hall, 2007.
- [6] S. Havre, B. Hetzler, and L. Nowell. ThemeRiver : Visualizing theme changes over time. In *Proceedings of the IEEE Symposium on Information Visualization*, pages 115–123, 2000.
- [7] Y. Wu, S. Liu, K. Yan, M. Liu, and F. Wu. OpinionFlow : Visual analysis of opinion diffusion on social media. *IEEE Transactions on Visualization and Computer Graphics*, 20(12) :1763–1772, 2014.
- [8] G. Sun, Y. Wu, S. Liu, T. Peng, J.J.H. Zhu, and R. Liang. EvoRiver : Visual analysis of topic coopetition on social media. *IEEE Transactions on Visualization and Computer Graphics*, 20(12) :1753–1762, 2014.
- [9] Weiwei Cui, Shixia Liu, Li Tan, Conglei Shi, Yangqiu Song, Zekai Gao, Huamin Qu, and Xin Tong. TextFlow : Towards better understanding of evolving topics in text. *IEEE Transactions on Visualization and Computer Graphics*, 17(12) :2412–2421, 2011.
- [10] James A. Wise, James J. Thomas, Kelly Pennock, David Lantrip, Marc Pottier, Anne Schur, and Vern Crow. Visualizing the non-visual : spatial analysis and

- interaction with information from text documents. In *Proceedings of the IEEE Symposium on Information Visualization (InfoVis'95)*, pages 51–58, 1995.
- [11] Ming C Hao, Christian Rohrdantz, Halldor Janetzko, Daniel A Keim, Umeshwar Dayal, Lars Erik Haug, Meichun Hsu, and Florian Stoffel. Visual sentiment analysis of customer feedback streams using geo-temporal term associations. *Information Visualization*, pages 273–290, 2013.
  - [12] Jianwei Zhang, Yukiko Kawai, Tadahiko Kumamoto, and Katsumi Tanaka. *A novel visualization method for distinction of web news sentiment*. Springer, 2009.
  - [13] Daniel A Keim. Information visualization and visual data mining. *IEEE Transactions on Visualization and Computer Graphics*, 8(1) :1–8, 2002.
  - [14] Manojit Sarkar and Marc H Brown. Graphical fisheye views. *Communications of the ACM*, 37(12) :73–83, 1994.
  - [15] James J Thomas. *Illuminating the path : the research and development agenda for visual analytics*. IEEE Computer Society, 2005.
  - [16] Ben Shneiderman. The eyes have it : A task by data type taxonomy for information visualizations. In *Proceedings of the IEEE Symposium on Visual Languages*, pages 336–343, 1996.
  - [17] Daniel Keim, Gennady Andrienko, Jean-Daniel Fekete, Carsten Görg, Jörn Kohlhammer, and Guy Melançon. *Visual analytics : Definition, process, and challenges*. Springer, 2008.
  - [18] Tamara Munzner. A nested process model for visualization design and validation. *IEEE Transactions on Visualization and Computer Graphics*, 15(6) :921–928, 2009.
  - [19] Adam Marcus, Michael S Bernstein, Osama Badar, David R Karger, Samuel Madden, and Robert C Miller. TwitInfo : aggregating and visualizing microblogs for event exploration. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 227–236, 2011.
  - [20] Jean Delay and Gustave Roussy. Les dérèglements de l'humeur. 1949.
  - [21] D Randy Garrison, Terry Anderson, and Walter Archer. Critical inquiry in a text-based environment : Computer conferencing in higher education. *The internet and higher education*, 2(2) :87–105, 1999.
  - [22] David Watson and Auke Tellegen. Toward a consensual structure of mood. *Psychological bulletin*, 98(2) :219, 1985.
  - [23] Robert E Thayer. *The biopsychology of mood and arousal*. Oxford University Press, 1989.
  - [24] James A Russell. A circumplex model of affect. *Journal of personality and social psychology*, 39(6) :1161, 1980.

- [25] Nancy A Remington, Leandre R Fabrigar, and Penny S Visser. Reexamining the circumplex model of affect. *Journal of personality and social psychology*, 79(2) :286, 2000.
- [26] Albert Bifet and Eibe Frank. Sentiment knowledge discovery in twitter streaming data. In *Discovery Science*, pages 1–15. Springer, 2010.
- [27] Douglas M McNair, Maurice Lorr, and Leo F Droppleman. *Profile of mood states*. Univ., 1971.
- [28] Margaret M Bradley and Peter J Lang. Affective norms for english words (anew) : Instruction manual and affective ratings. Technical report, Technical Report C-1, The Center for Research in Psychophysiology, University of Florida, 1999.
- [29] Lee Byron and Martin Wattenberg. Stacked graphs–geometry & aesthetics. *IEEE Transactions on Visualization and Computer Graphics*, 14(6) :1245–1252, 2008.
- [30] Nicholas Diakopoulos, Mor Naaman, and Funda Kivran-Swaine. Diamonds in the rough : Social media visual analytics for journalistic inquiry. In *IEEE Conference on Visual Analytics Science and Technology (VAST)*, pages 115–122, 2010.
- [31] Enamul Hoque and Giuseppe Carenini. ConVis : A visual text analytic system for exploring blog conversations. *Computer Graphics Forum*, 33(3) :221–230, 2014.
- [32] W. Cui, S. Liu, Z. Wu, and H. Wei. How hierarchical topics evolve in large text corpora. *IEEE Transactions on Visualization and Computer Graphics*, 20(12) :2281–2290, 2014.
- [33] Dongning Luo, Jing Yang, M. Krstajic, W. Ribarsky, and D. Keim. EventRiver : Visually exploring text collections with temporal references. *IEEE Transactions on Visualization and Computer Graphics*, 18(1) :93–105, 2012.
- [34] John Stasko, Carsten Görg, and Zhicheng Liu. Jigsaw : supporting investigative analysis through interactive visualization. *Information visualization*, 7(2) :118–132, 2008.
- [35] Alan M MacEachren, Anuj Jaiswal, Anthony C Robinson, Scott Pezanowski, Alexander Savelyev, Prasenjit Mitra, Xiao Zhang, and Justine Blanford. Senseplace2 : Geotwitter analytics support for situational awareness. In *IEEE Conference on Visual Analytics Science and Technology (VAST)*, pages 181–190, 2011.
- [36] Allison June-Barlow Chaney and David M Blei. Visualizing topic models. In *Proceedings of the Sixth International AAAI Conference on Weblogs and Social Media*, 2012.
- [37] Bongshin Lee, Nathalie Henry Riche, Amy K Karlson, and Sheelagh Carpendale. SparkClouds : Visualizing trends in tag clouds. *IEEE Transactions on Visualization and Computer Graphics*, 16(6) :1182–1189, 2010.

- [38] Daniel Archambault, Derek Greene, Pádraig Cunningham, and Neil Hurley. The-meCrowds : Multiresolution summaries of twitter usage. In *Proceedings of the 3rd international workshop on Search and mining user-generated contents*, pages 77–84, 2011.
- [39] Colin Ware. *Information visualization : perception for design*. Elsevier, 2012.
- [40] Romain Vuillemot. *Un cadre de conception pour la Visualisation d'Information Interactive*. PhD thesis, INSA de Lyon, 2010.
- [41] Matthew O Ward, Georges Grinstein, and Daniel Keim. *Interactive data visualization : foundations, techniques, and applications*. CRC Press, 2010.
- [42] Stuart K Card and Jock Mackinlay. The structure of the information visualization design space. In *IEEE Symposium on Information Visualization*, pages 92–99, 1997.
- [43] Riccardo Mazza. *Introduction to information visualization*. Springer Science & Business Media, 2009.
- [44] Lisa Feldman Barrett and James A Russell. Independence and bipolarity in the structure of current affect. *Journal of personality and social psychology*, 74(4) :967, 1998.