



CONSULTING | TECHNOLOGY

Helping enterprises realize their
Full Potential

AGILITY | COLLABORATIVE INTELLIGENCE | INNOVATION | PERFORMANCE

L'ingénierie linguistique au service de la relation client

Laëtitia Guillot

Vendredi 9 octobre 2015



1. MON PARCOURS
2. PRÉSENTATION DE KEYRUS
3. LA FOUILLE DE DONNÉES TEXTUELLES ET LES TYPES D'ÉTUDES POSSIBLES
4. PREMIER CAS D'ÉTUDE : ANTICIPER LA RÉSILIATION DE CONTRAT SUITE À UN SINISTRE
5. DEUXIÈME CAS D'ÉTUDE : AMÉLIORER LA SATISFACTION CLIENT
6. CONCLUSION



1. MON PARCOURS

2. PRÉSENTATION DE KEYRUS

3. LA FOUILLE DE DONNÉES TEXTUELLES ET LES TYPES D'ÉTUDES POSSIBLES

4. PREMIER CAS D'ÉTUDE : ANTICIPER LA RÉSILIATION DE CONTRAT SUITE À UN SINISTRE

5. DEUXIÈME CAS D'ÉTUDE : AMÉLIORER LA SATISFACTION CLIENT

6. CONCLUSION





Stage de fin d'études (6 mois)

Guide sur bonnes pratiques de développement d'un composant d'extraction

Développement de ressources linguistiques



Ingénieur linguiste (2 ans)

Mise en place d'un workflow d'annotation dans le domaine juridique



Consultante en text mining (depuis septembre 2012)

Pilotage et réalisation de projets de fouille de données textuelles



1. MON PARCOURS
- 2. PRÉSENTATION DE KEYRUS**
3. LA FOUILLE DE DONNÉES TEXTUELLES ET LES TYPES D'ÉTUDES POSSIBLES
4. PREMIER CAS D'ÉTUDE : ANTICIPER LA RÉSILIATION DE CONTRAT SUITE À UN SINISTRE
5. DEUXIÈME CAS D'ÉTUDE : AMÉLIORER LA SATISFACTION CLIENT
6. CONCLUSION



2. PRÉSENTATION DE LA SOCIÉTÉ KEYRUS

UN GROUPE SOLIDE ET AGILE

173 M€

Chiffre d'affaires 2014

Les infrastructures et les processus (qualité / RH) d'un grand groupe de services professionnels

Une gouvernance simple et formalisée pour conserver l'agilité durablement

Un centre de décision proche des clients

Société cotée au NYSE-Euronext Paris

SPÉCIALISTE DE LA PERFORMANCE DES ORGANISATIONS

400 Clients
Grands Comptes*
& MGE

*dont 80 Global Fortune 500

3800 Clients PME

Une capacité d'action sur la stratégie, les organisations et les systèmes de Gestion de la Performance

Des Business Units distinctes pour servir chaque typologie de clients (Grands Comptes, MGE, PME)

Des compétences fonctionnelles, sectorielles et technologiques

CA par secteur



- Industries : 32%
- Banques - Assurances : 17%
- Télécoms : 10%
- Services – Distribution : 23%
- Service public : 10%
- Utilities : 8%

DES VALEURS AU SERVICE DE NOS CLIENTS

+2000 Collaborateurs

L'esprit d'entreprise diffusé par un fondateur toujours aux commandes

Une forte proximité clients

La qualité de service comme base de construction de notre marque

La culture de l'innovation au cœur de nos modes de fonctionnement

La diversité, vecteur clé de notre politique RH

UNE DIMENSION INTERNATIONALE

15 pays sur 4 continents

Une expertise des déploiements de projets internationaux

De fortes capacités nearshore & offshore



1. MON PARCOURS
2. PRÉSENTATION DE KEYRUS
- 3. LA FOUILLE DE DONNÉES TEXTUELLES ET LES TYPES D'ÉTUDES POSSIBLES**
 1. DÉFINITION DU TEXT MINING
 2. LES DIFFÉRENTS TYPES D'ANALYSE
4. PREMIER CAS D'ÉTUDE : ANTICIPER LA RÉSILIATION DE CONTRAT SUITE À UN SINISTRE
5. DEUXIÈME CAS D'ÉTUDE : AMÉLIORER LA SATISFACTION CLIENT
6. CONCLUSION



- Le **Text Mining** est l'ensemble des techniques permettant de mettre au jour de manière automatisée l'information utile « cachée » dans des masses importantes de données textuelles (web, système documentaire d'une entreprise, bases de données scientifiques, etc...)
- Ces techniques sont issues :
 - ▶ du **traitement automatique du langage naturel** qui permet de formaliser et de structurer les données textuelles et de les rendre informatiquement exploitables
 - ▶ du **data mining** qui permet d'analyser des données structurées
- **Applications** : veille, gestion documentaire, ressources humaines, gestion de la relation client, etc...



■ Quelles données ?

- ▶ Enquêtes de satisfaction (questions ouvertes)
- ▶ Données des SI (champs commentaires)
- ▶ Courriers et mails
- ▶ Retranscriptions des conversations en centre d'appels
- ▶ Données web (presse, forums, etc.)

■ Pourquoi faire ?

- ▶ Mieux connaître les clients à travers leur discours
- ▶ Mieux comprendre les effets de la communication commerciale
- ▶ Mieux analyser le comportement des conseillers clientèle
- ▶ S'intéresser à des événements ponctuels
- ▶ Développer de nouvelles offres



1. MON PARCOURS
2. PRÉSENTATION DE KEYRUS
- 3. LA FOUILLE DE DONNÉES TEXTUELLES ET LES TYPES D'ÉTUDES POSSIBLES**
 1. DÉFINITION DU TEXT MINING
 - 2. LES DIFFÉRENTS TYPES D'ANALYSE**
4. PREMIER CAS D'ÉTUDE : ANTICIPER LA RÉSILIATION DE CONTRAT SUITE À UN SINISTRE
5. DEUXIÈME CAS D'ÉTUDE : AMÉLIORER LA SATISFACTION CLIENT
6. CONCLUSION



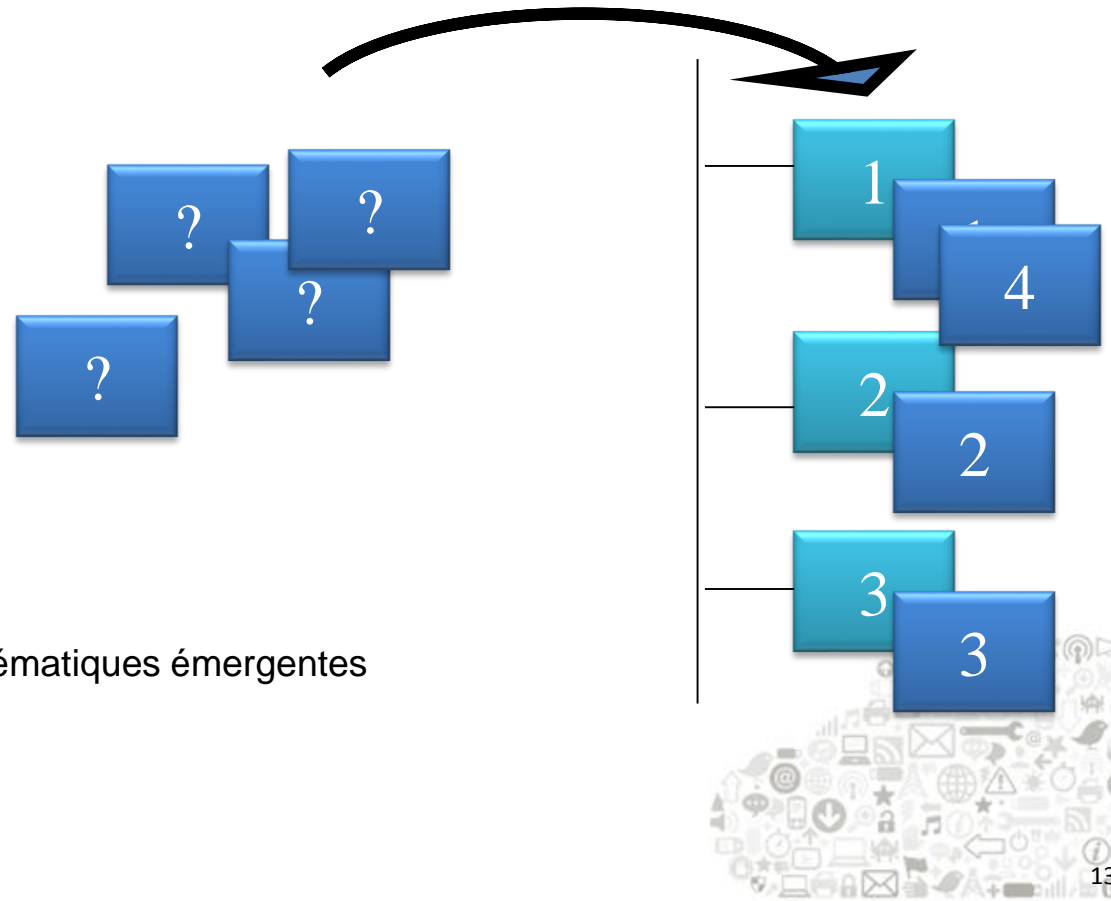
- Objectif : **découverte sans a priori** des thématiques abordées dans un corpus par rapprochement des documents partageant un vocabulaire proche
- Type d'utilisation : exploration de **corpus nouveaux**



- **Avantages :**
 - ▶ Permet de découvrir de nouvelles thématiques
 - ▶ Possibilité d'ajuster la finesse de la classification
- **Inconvénients :**
 - ▶ Beaucoup de tâtonnements avant de parvenir à des résultats satisfaisants
 - ▶ Longueur de l'interprétation

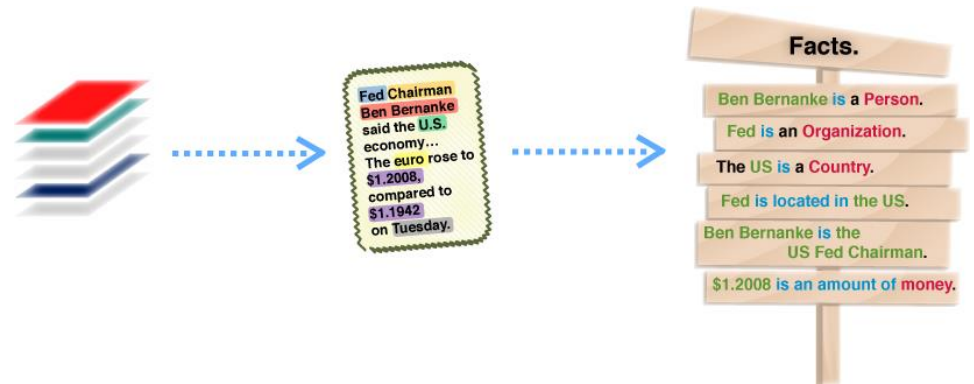


- **Objectif** : catégorisation thématique d'un corpus à partir d'un modèle dont les thèmes sont **déterminés à l'avance**
- Type d'utilisation : sur des types de **corpus déjà étudiés**, suivi de l'évolution d'un thème dans le temps
- Avantages :
 - ▶ Automatisation des traitements
 - ▶ Gain de temps
 - ▶ Possibilité de modéliser un thème
- Inconvénients :
 - ▶ Possibilité de passer à côté de thématiques émergentes



3.2. LES DIFFÉRENTS TYPES D'ANALYSE : L'EXTRACTION D'INFORMATION

- Objectif : modélisation d'une thématique prédéfinie
- Type d'utilisation : **extraire d'un corpus toutes les données abordant une thématique spécifique** (noms de personnes pour anonymiser le corpus, cibler les clients qui parlent de déménager, ...)
- Avantages :
 - ▶ Précision de l'extraction
- Inconvénients :
 - ▶ Exhaustivité difficile à atteindre



1. MON PARCOURS
2. PRÉSENTATION DE KEYRUS
3. LA FOUILLE DE DONNÉES TEXTUELLES ET LES TYPES D'ÉTUDES POSSIBLES
- 4. PREMIER CAS D'ÉTUDE : ANTICIPER LA RÉSILIATION DE CONTRAT SUITE À UN SINISTRE**
5. DEUXIÈME CAS D'ÉTUDE : AMÉLIORER LA SATISFACTION CLIENT
6. CONCLUSION



■ Contexte du projet

- ▶ Analyse de réponses à une question ouverte issue d'un questionnaire envoyé suite à une clôture de sinistre automobile : « *Avez-vous des remarques ou suggestions ?* »
 - Domaine : assurance
 - Segment : clients particuliers et professionnels
 - Volumétrie : 6 000 verbatims environ (octobre 2014 à avril 2015)

■ Objectifs du projet

- ▶ Identifier les thématiques présentes dans les verbatims
- ▶ Cibler les clients insatisfaits et disposer de règles de priorisation des clients à rappeler
- ▶ Savoir différencier les motifs liés aux sinistres d'autres motifs
- ▶ Développer une approche ROIste



■ Contexte client

- ▶ Novice des études de fouille de données textuelles
- ▶ Consultant autonome et détaché physiquement sur la réalisation du projet
- ▶ Contraintes techniques
- ▶ POC (étude de faisabilité)
 - budget limité : logiciel open source
 - contrainte de temps : 14 jours

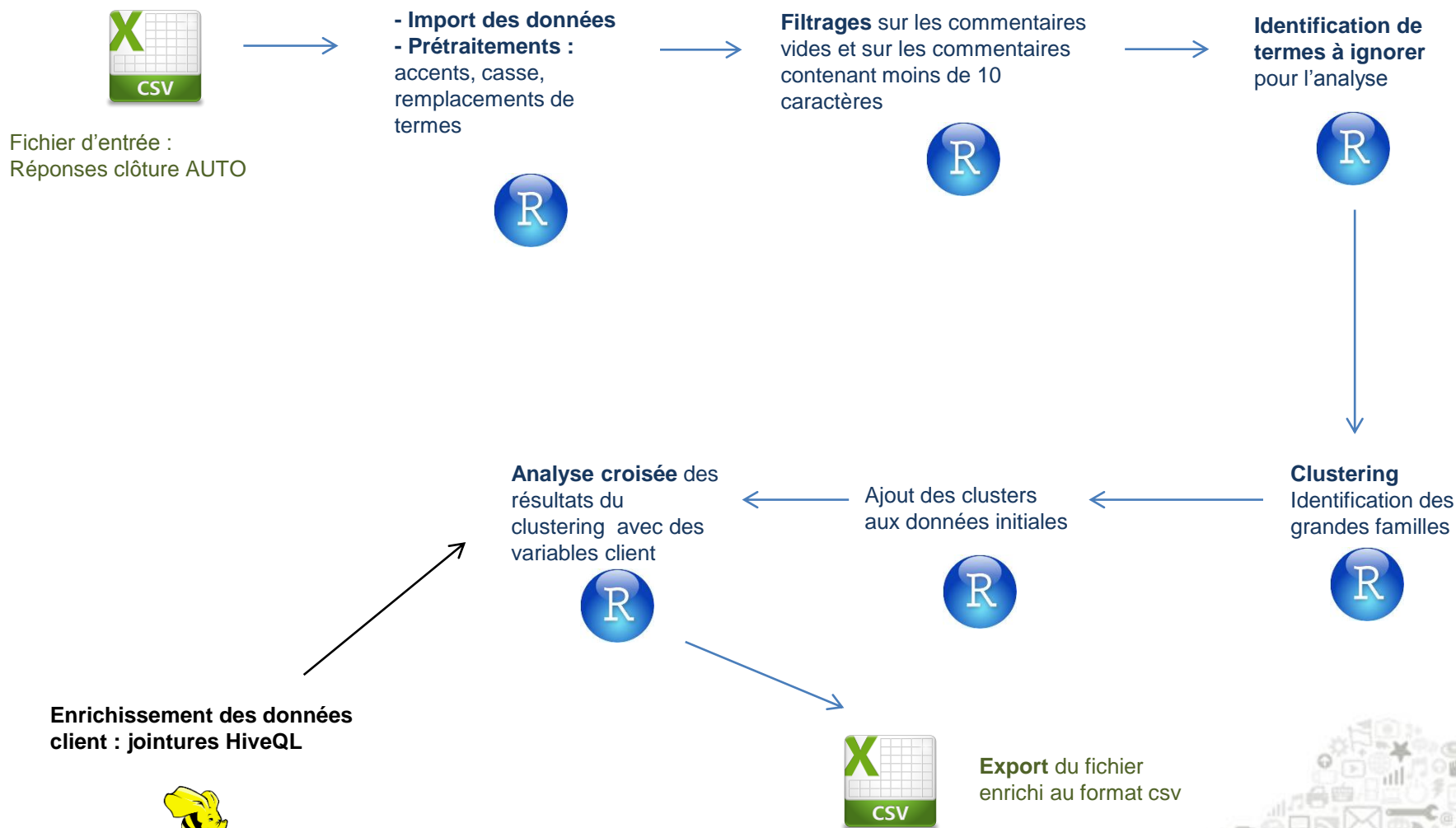
■ Méthode : classification sans a priori

■ Outil utilisé : R package tm



4. ANTICIPER LA RÉSIATION DE CONTRAT SUITE À UN SINISTRE

Chaîne de traitement



4. ANTICIPER LA RÉSILIATION DE CONTRAT SUITE À UN SINISTRE

Classification sans a priori

- **L'analyse sémantique latente** est une méthode d'analyse exploratoire largement utilisée en Statistique. D'une manière générale, elle s'applique sur une table de contingence, ici elle est appliquée sur une matrice termes-documents. Celle-ci contient les occurrences pondérées des termes au sein des documents du corpus. Dans le cas de la matrice termes-documents, le nombre de lignes détermine le nombre de termes et le nombre de colonnes détermine le nombre de documents.
- Le **clustering** de documents est une méthode non supervisée largement utilisée en Statistique qui a ici pour objectif de regrouper ensemble les documents véhiculant une même "idée". Le clustering (ou classification en français) peut être appliqué sur les "vecteurs documents" directement issus de la matrice documents-termes, ou appliqué sur les vecteurs coordonnées issus d'une analyse factorielle (analyse sémantique latente).
- Nous construisons ici la classification sur la base des distances entre coordonnées issues de l'analyse factorielle (LSA). Dans ce cas, la distance la plus utilisée est la distance euclidienne.



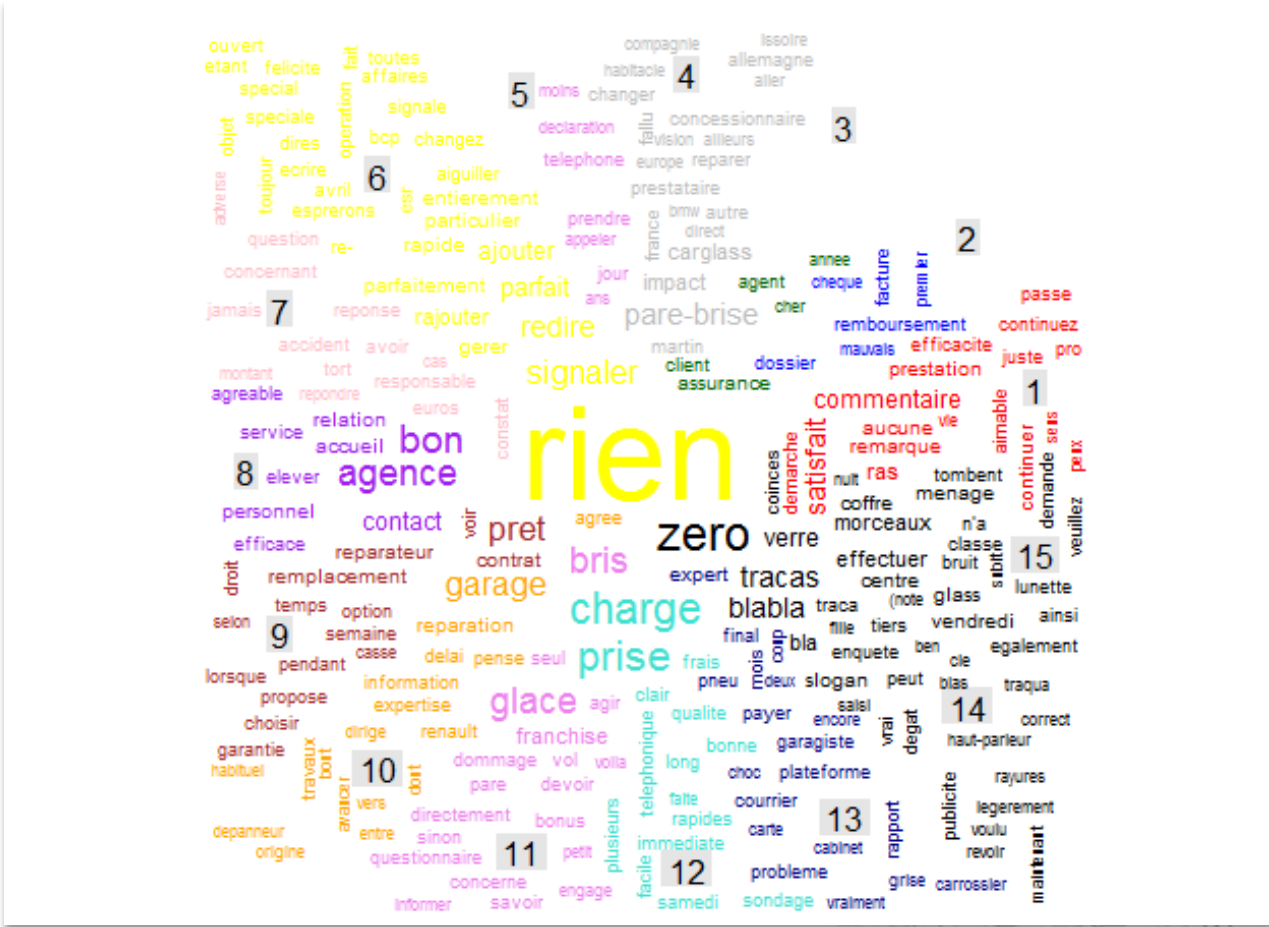
4. ANTICIPER LA RÉSILIATION DE CONTRAT SUITE À UN SINISTRE

Résultats

Répartition du nombre de documents par cluster

clusters														
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
328	152	950	24	263	58	262	142	46	42	45	50	23	33	1

Nuage de mots discriminant par cluster



4. ANTICIPER LA RÉSILIATION DE CONTRAT SUITE À UN SINISTRE

Résultats

Cluster	Thématique	Nombre de documents
1	Rien à signaler	328
2	Délai de remboursement	152
3	Zéro tracas zéro blabla	950
4	Sinistre lié au pare-brise	24
5	Incompétence de l'expert	263
6	Rien à dire	58
7	Franchise élevée	262
8	Contact avec les agences	142
9	Option véhicule de remplacement	46
10	Garage concerné pour les réparations	42
11	Franchise sur les sinistres bris de glace	45
12	Prise en charge du sinistre	50
13	Gestion du sinistre	23
14	Zéro tracas zéro blabla	33
15	Bris de glace	1
	Total	2419

Exemples cluster 1 :

« Dieu merci, peu de sinistre mais toujours satisfaite »

« Très très bien pourvu que sa dure.... »

Exemples cluster 3 :

« Que du tracas et que du bla-bla XXX »

« Avec XXX pas de tracas »

Exemples cluster 5 :

« Je ne suis pas content de la décision de l'expert vu que mes pneus n'avait que 1000 kilomètre . »

« Service extra de XXX Montreuil - incontestablement.

Expert utilisé par XXX peu poli voir pas poli du tout - a changer.

Vu la rythme de ma travail, je ne peut pas me déplacer comme je veux pour l'expertiser... »

Exemples cluster 7 :

« franchise trop importante !!!! 290 euros »

« la franchise de 290 euros est trop élevée »

- **Conclusion** : résultats satisfaisants, conformes aux attentes du client

- **Difficultés rencontrées** :
 - ▶ Contexte particulier : étude de faisabilité (POC)

 - ▶ Nombreuses contraintes sur le projet
 - Budget
 - Temps
 - Outil utilisé (linguistique restreinte)
 - Contexte technique
 - Attentes clients fortes



1. MON PARCOURS
2. PRÉSENTATION DE KEYRUS
3. LA FOUILLE DE DONNÉES TEXTUELLES ET LES TYPES D'ÉTUDES POSSIBLES
4. PREMIER CAS D'ÉTUDE : ANTICIPER LA RÉSILIATION DE CONTRAT SUITE À UN SINISTRE
- 5. DEUXIÈME CAS D'ÉTUDE : AMÉLIORER LA SATISFACTION CLIENT**
6. CONCLUSION



■ Contexte du projet

- ▶ Analyse des réclamations client issus des centres de relation client
 - Segment : clients professionnels
 - Volumétrie : 10 000 verbatims à traiter par trimestre
 - Outil : TEMIS
 - Durée longue : 6 mois environ dont 4 mois pour la construction du modèle

■ Objectifs du projet

- ▶ Classer chaque verbatim dans une catégorie prédéfinie
- ▶ Produire des analyses
- ▶ Suivre et mesurer les évolutions au fil du temps
- ▶ Faire un retour aux unités concernées
- ▶ Maintien de la satisfaction client



■ Contexte client

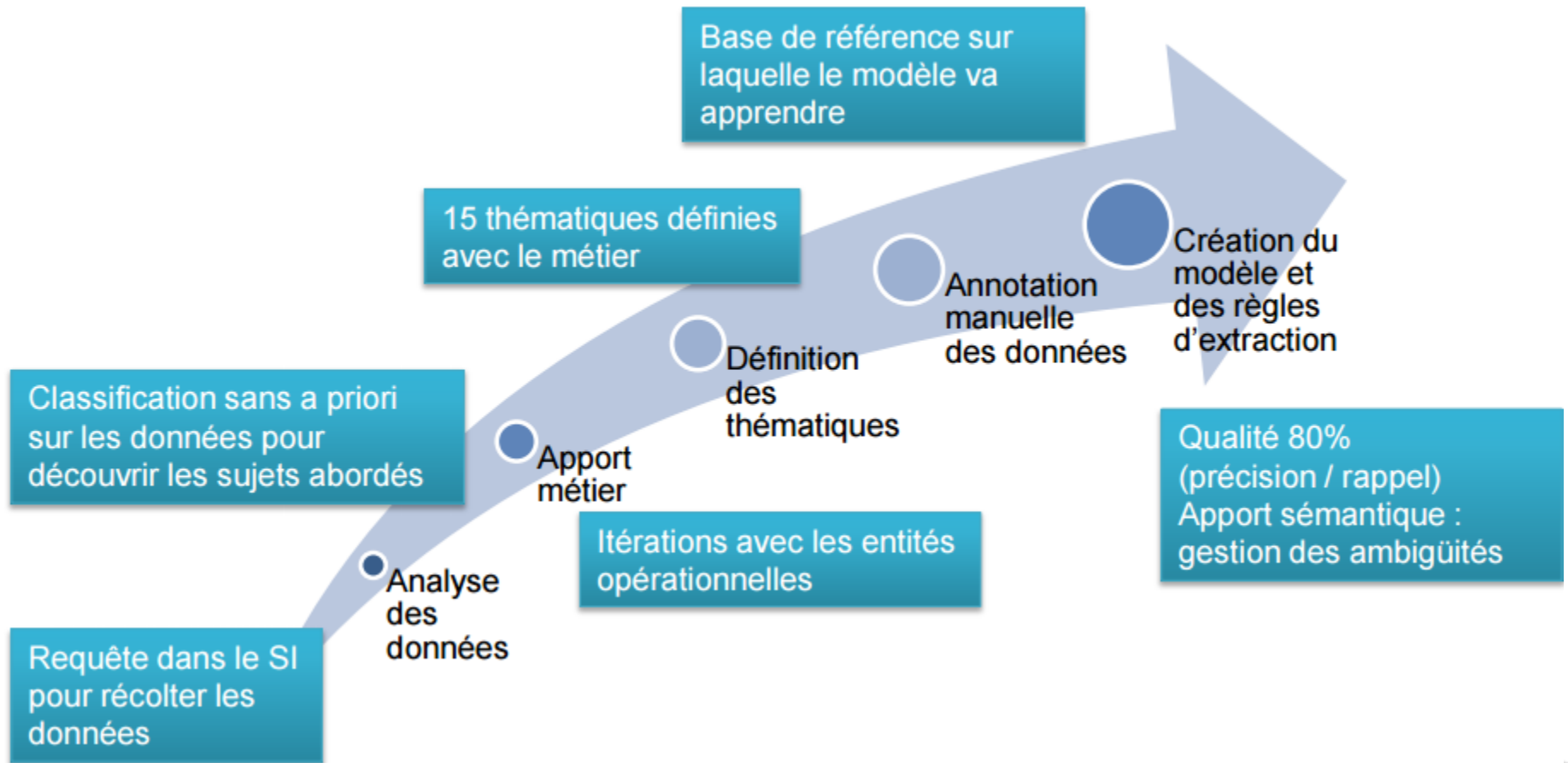
- ▶ Familier des technologies de text mining, réalise des études depuis 7 ans
- ▶ Possède leur logiciel de text mining
- ▶ Appui du consultant au sein de l'équipe text mining
- ▶ Projet au sein d'une mission longue

■ Méthode : construction d'un modèle de catégorisation



5. AMÉLIORER LA SATISFACTION CLIENT

Création d'un modèle de catégorisation



5. AMÉLIORER LA SATISFACTION CLIENT

Définition des thématiques

- ✓ **Modèle principal par processus**

- Les thématiques sont déclinées en sous-thématiques au sein d'un plan détaillé.



5. AMÉLIORER LA SATISFACTION CLIENT

Le modèle : les choix techniques

- Chaque commentaire est classé dans une seule thématique processus et dans une ou des catégories transverses.
- Catégorisation sur les thématiques processus
 - Utilisation des modèles de catégorisation couplés à des **composants d'extraction métier**
→ la statistique est insuffisante
 - La sémantique permet de pallier les faiblesses du modèle statistique par la gestion des ambiguïtés
 - Qualité délibérée de **80%** pour identifier les thématiques émergentes par mi les verbatims dits « non classés »
- Thématiques complémentaires ou transverses
 - **Exclusivement par composants d'extraction métier**
 - Permettent une vue différente du modèle
 - Traitement des thématiques émergentes sans casser le modèle



5. AMÉLIORER LA SATISFACTION CLIENT

Le modèle : l'apport déterminant de la sémantique

- **Faiblesses de l'approche purement statistique** : regroupement de document en fonction du vocabulaire employé mais sans prise en compte du contexte métier (les concepts métiers, la négation...)

« je suis satisfait » vs. « je ne suis pas satisfait »

« réception de facture » vs. « compréhension de facture »

- **Nécessité de prendre en compte le contexte**, pas seulement des mots clés !
 - Règles d'extraction s'appuient sur la sémantique



■ Conclusion :

- ▶ Utile pour suivi sur le long terme
- ▶ Amélioration de la productivité
- ▶ Suivi régulier, fiable et aligné sur la vision métier

■ Difficultés rencontrées :

- ▶ Les données dépendent du conseiller
 - Bruit : erreurs orthographiques, abréviations non normalisées → substitution orthographique et normalisation
 - Subjectivité du conseiller : retranscrit sa compréhension du problème du client
- ▶ Longue mise en place des modèles
 - Coopération avec les entités opérationnelles
 - Nécessité de bien connaître les données à analyser
 - Phases de tâtonnements importantes avant de parvenir à un résultat correct

1. MON PARCOURS
2. PRÉSENTATION DE KEYRUS
3. LA FOUILLE DE DONNÉES TEXTUELLES ET LES TYPES D'ÉTUDES POSSIBLES
4. PREMIER CAS D'ÉTUDE : ANTICIPER LA RÉSILIATION DE CONTRAT SUITE À UN SINISTRE
5. DEUXIÈME CAS D'ÉTUDE : AMÉLIORER LA SATISFACTION CLIENT
- 6. CONCLUSION**



6. CONCLUSION

Prise de recul vis-à-vis du master et de la vie professionnelle

Master

- Double compétence : importance de la programmation
- Stage = tremplin
- Culture du réseau
 - ▶ Collègues
 - ▶ Réseaux sociaux professionnels (LinkedIn, Viadeo)

Vie professionnelle

- Autonomie
- Prise d'initiative
- Polyvalence / adaptation
- Communication (ne pas avoir peur d'aller voir les gens)
- Nouveaux apprentissages réguliers (outils, langages)
- Veille
- Savoir se vendre : domaine méconnu



QUESTIONS

