

Analyse Quantitative et Qualitative de données textuelles

Normand Péladeau, Ph.D.
Président
Recherches Provalis

Les Produits de Recherches Provalis

SIMSTAT (1989)
Analyses Statistiques

Simstat v2.5

Simstat pour Windows

Fichier Edition Données Statistiques Graphes Script Outils Fenêtres Aide

Données - cars2.DBF 8/406

NAME	MPG	CYLIND	DISPLA	HRSPWR	ACCEL
chevrolet chevelle malibu	18.00	8	307	130	3504.0
buick skylark 320	15.00	8	350	165	3693.0
plymouth satellite	18.00	8	318	150	3436.0
amc rebel sst	16.00	8	304	150	3433.0

ACCEL) + (.00000504 < > 7 de 7

(.00737 * ACCEL) + (.00000504 * ACCEL^2)
(r = .8712)

Bloc-notes - SANS TITRE

RÉGRESSION HRSPWR par ACCEL

Régression

R = .8712 R Carré = .7589 Prob de R = .0000

Analyse de Variance

Source	D.l.	Somme des carrés	Carré Moyen	F	Prob
Régression	2	455140.3981	227570.1991	624.951	.0000
Résidu	397	144563.8794	364.1408		

Équation: HRSPWR = 34.5304 + (.007371 * ACCEL) + (.000005035 * ACCEL^2)

Variable	B	E-T B	Int. de confiance à 95%	t	Prob
Origine	34.5304				
Degré 1	.007371	.008953	-.01023 à .02497	.823	.4108
Degré 2	.000005035	.000001390	.000002303 à .000007768	3.623	.0003

Page 8 de 8

Analyses / Index

cars2.DBF - 03/14/2006 9:01:40 AM X : 2 Y : 1 Enregistrer

Les Analyses Statistiques

- **Analyses Descriptives**

(fréquences et tableaux croisés, réponses multiples, etc.)

- **Analyses Comparatives**

(test de Student, Anova, GLM-Anova, etc)

- **Analyses de Corrélations et Régressions**

(linéaire, non linéaire, multiple par bloc hiérarchique, pas-à-pas, etc.)

- **Analyses Psychométriques**

(analyse de fidélité de Cronbach, analyse d'items, courbes ROC)

- **Tests Non Paramétriques**

- **Analyses Multivariées (avec module optionnel)**

(analyse factorielle, composantes principales, correspondance, cluster, etc.)

Particularités de Simstat

- **Analyses spécifiques**
 - Simulation et échantillonnage “Bootstrap”
 - Analyse classique d’items
 - Analyse de sensibilité (courbes ROC)
 - Calcul des accords inter-juges
 - Devis expérimentaux à cas unique
- **Variables mémos et documents “rich text”**

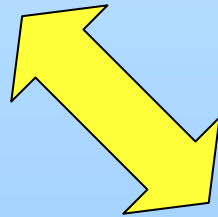
Analyse de texte

The screenshot displays a text analysis software interface. On the left, a newspaper clipping from 'Le Monde' is shown, featuring the headline 'Lionel Jospin, sa part de vérité' and a cartoon. Below the clipping is a transcript titled 'Fig. 3 Example of a focus group transcript' with dialogue between an interviewer and four participants (A, B, C, D). In the center, a document titled 'ARTS-INDUSTRY DESIGN' is visible. On the right, a data table lists analysis results with columns for 'UNDER_AGE', 'START_P', 'END_P', 'PROG_NO', 'COMMENT', and 'PROG_NAME'. A red circle highlights the 'COMMENT' column, and red lines connect it to the transcript and the central document.

UNDER_AGE	START_P	END_P	PROG_NO	COMMENT	PROG_NAME
14,0			131	(TEXT)	Weather
24,0			131	(TEXT)	Weather
24,0	1130	1215	13	(TEXT)	Bargain Hunt
24,0			131	(TEXT)	Weather
23,0	1930	2000	28	(TEXT)	Coronation Street
23,0	2100	2200	80	(TEXT)	Paul McKenna Will Make
23,0			131	(TEXT)	Weather
12,0			131	(TEXT)	Weather
25,0	1930	2000	28	(TEXT)	Coronation Street
16,0	2200	2235	7	(TEXT)	BBC News; Regional Ne
10,0		2100	79	(TEXT)	Outtake TV
10,0		1870	24	(TEXT)	Central News at Six and'
10,0		1900	54	(TEXT)	Joey
10,5		2405	133	(TEXT)	Without a Trace
10,5			131	(TEXT)	Weather
10,5			131	(TEXT)	Weather
10,5		2030	53	(TEXT)	Is Your Medicine Fake? ,
10,5			131	(TEXT)	Weather
10,5		2150	12	(TEXT)	Balderdash and Piffle
10,5		2335	78	(TEXT)	One Foot in the Grave
10,5		2100	111	(TEXT)	The Root of All Evil
10,5			131	(TEXT)	Weather
10,5		1930	4	(TEXT)	Around the World in 80 T

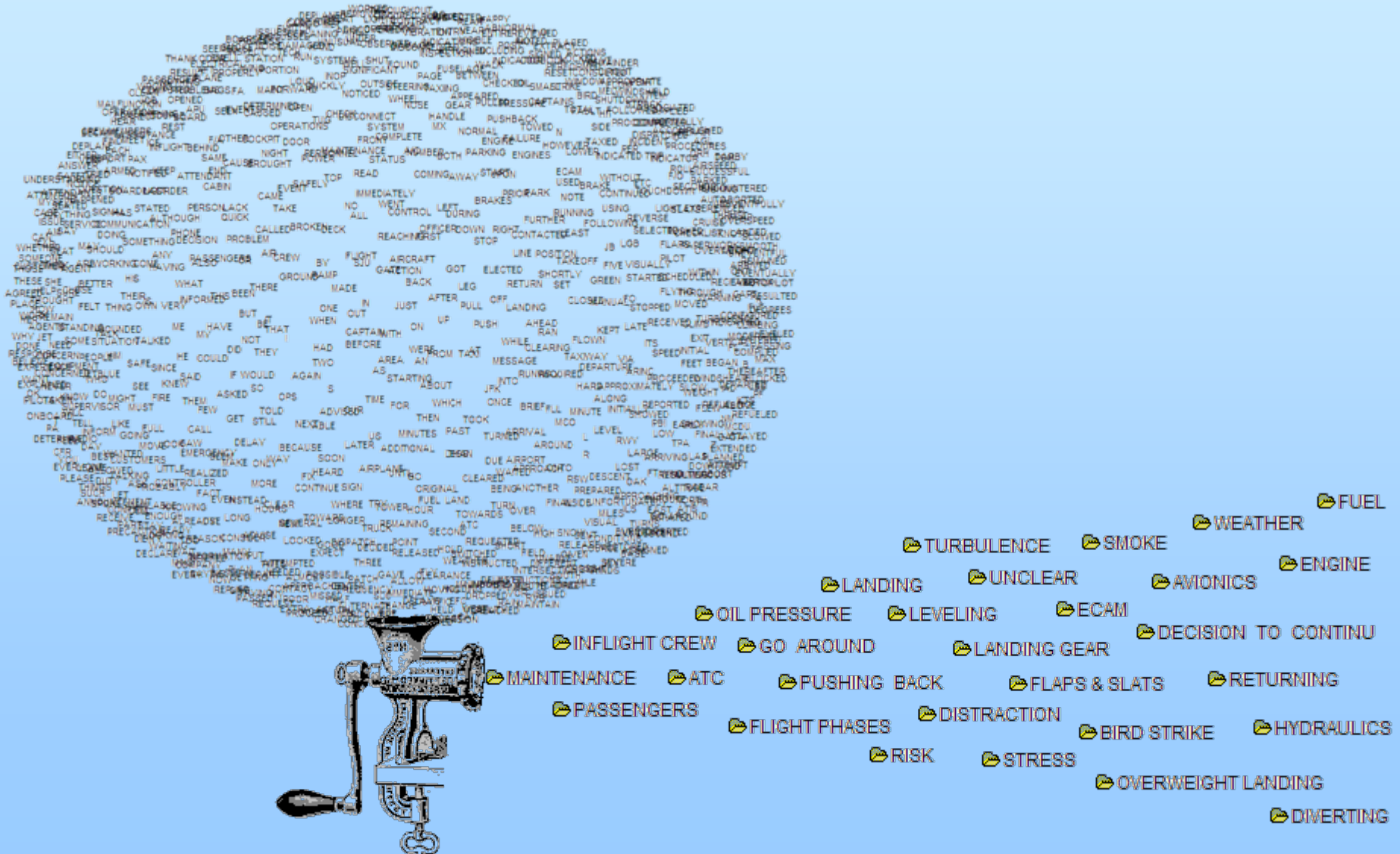
Les Produits de Recherches Provalis

SIMSTAT (1989)
Analyses Statistiques



WORDSTAT (1998)
Analyse de Textes Assistée par Ordinateur
& Exploration de Textes (text mining)

Le défi de l'analyse de texte



Le défi de l'analyse de texte



0 1 0 0 0 0 0 2 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 6 0 0 0 0 0 0 0 0 0 0 1 1 1 0 0 0 0 0 0 0 3 0 0 0 0 0 ... 0

Prétraitement Linguistique

- Élimination des mots outils (“stop list”)
- Extraction des radicaux (“stemming”) ou lemmatisation

2 0 0 0 1 0 1 0 0 0 0 1 0 0 0 0 0 0 6 0 0 0 0 0 0 0 0 0 0 0 1 2 0 0 0 0 0 0 3 0 0 0 0 0 ... 0

Catégorisation Sémantique

Extraction et catégorisation du vocabulaire technique, des mots signifiants et chaînes de mots

- Extracteur de chaînes de mots et du vocabulaire technique
- Catégorisation des mots et chaînes de mots par un dictionnaire usager (taxonomie)
- Thésaurus et bases de données sémantiques
- Liste des mots en contexte (KWIC) pour validation
- Utilisation de règles de désambiguïisation

2 0 0 3 0 6 0 0 0 1 1 0 3 0 0 1 2 0 1 0 0 0 3 0 0 2 0 ... 3

Le défi de l'analyse de texte

4- ENVIRONMENTAL FACTORS

4-1 QUICK SUCCESSION

- ALL_OF_THE_SUDDEN (1)
- ALMOST_IMMEDIATELY (1)
- FOLLOWED_IMMEDIATELY (1)
- IMMEDIATELY (1)
- IMMEDIATELY_AFTER (1)
- IMMEDIATELY_FOLLOWED (1)
- IMMEDIATELY_FOLLOWING (1)
- MOMENTS (1)
- QUICKLY_BECAME (1)
- QUICKLY_FOLLOWED (1)
- QUICKLY_RESOLVED_TO (1)
- SECONDS_LATER (1)
- SHORTLY (1)
- SIMULTANEOUSLY (1)
- THEN_VERY_QUICKLY (1)

4-2 MULTIPLICITY

- 2_AIRCRAFT (1)
- BOTH_CONFLICTS (1)
- BOTH_OF_THE_OTHER_AIRCRAFT (1)
- FIRST_RA (1)
- MANY (1)
- MULTIPLE (1)
- NUMEROUS (1)
- SECOND_RA (1)
- TARGETS (1)
- SEVERAL (1)

5- HUMAN FACTORS

5-1 ASSERTIVENESS

5-2 AWARENESS

5-3 ARGUMENTATIVE

5-4 DISTRACTION

- DISTRACTED (1)
- DISTRACTION (1)
- INATTENTION (1)
- INATTENTIVE (1)
- UNALERT (1)
- UNAWARE (1)
- UNOBSERVANT (1)
- UNVIGILANT (1)
- UNWATCHFUL (1)

5-5 MISUNDERSTANDING

5-6 MISTAKE

5-7 OMISSION

- DISREGARDED (1)
- FORGET (1)
- FORGOTTEN (1)
- IGNORE (1)
- IGNORED (1)
- NEGLECTED (1)
- NEGLIGENCE (1)
- OMISSION (1)
- OMIT (1)
- OMITTED (1)
- OVERLOOKED (1)

Le défi de l'analyse de texte

• Liste des mots en contexte (KWIC)

RECNO		KEYWORD	
2766	UND TO THE PASSENGERS IN HOPES OF ALLEVIATING THEIR	STRESS	. THE PART ARRIVED AT THE AIRRAFT AT APPROXIMATELY 1
2690	is still not seated. I had to make a second PA announcement to	stress	the need for all pax to take their seats at this time without furth
209	prior to seatbelt sign going on. I reiterated the announcemnt to	stress	that everyone one would not be able to get up later due to the
3068	is signed off. And we left in a matter of minutes. I would like to	stress	the OUTSTANDING job of everyone on the ground in FLL, espe
7922	ecovering from brain surgery and wasn't suppose to be under	stress	and wasnt going to discuss it any further. We were able to pl
8702	and eyes rolled back. F4 - Passenger was had been under	stress	, tired and had very little food. She had taken dramamine and l
7972	d very tired, 3 days with no good sleep. Death in family under	stress	. No allergies, no medication on board, gave him oxygen, we p
8702	F2 - Passenger was under	stress	, tired and had little food. She took dramamine and had 1 merlo
387	ea were in agreement that row ten was causing unnecessary	stress	. I have previously worked as a flight attendant and between
3512	er there either. I finally got through and got Maureen again and	stressed	we had everything but the flight plan filed. She said she would
5448	: Mr.Koch took the fit back to Den even after we explained and	stressed	the fact that although we would accomodate her for the night
1231	her behavior but I also took into consideration that she must be	stressed	out from the holiday. They advised me (the GSC) that there wa
10273	:endants have the proper training in this regard, as it should be	stressed	more than it apparently is. I would like this report to be consi

• Règles de désambiguïsation

TRANSFER THE IS OR STRESS* THAT Emphases

TRANSFER STRESS* IS NOT État ÉMOTIONNEL

INSATISFAIT OR (SATISFAIT IS AFTER #NÉGATION)

Le défi de l'analyse de texte



0 1 0 0 0 0 0 2 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 6 0 0 0 0 0 0 0 0 0 1 1 1 0 0 0 0 0 0 3 0 0 0 0 ... 0

Prétraitement Linguistique

Élimination des mots outils ("stop list")
Extraction des radicaux ("stemming")
Lemmatisation

2 0 0 0 1 0 1 0 0 0 0 1 0 0 0 0 0 0 6 0 0 0 0 0 0 0 0 0 0 1 2 0 0 0 0 0 0 3 0 0 0 0 ... 0

- Classification hiérarchique ascendante des mots ou des documents
- Groupement des documents (k-means)
- Réduction statistique de la matrice mots x documents (SVD, analyse factorielle, composante principale, etc.)
- Classification automatique de documents par algorithme d'apprentissage (Naive Bayes, SVM K-nearest Neighbor, etc.)

Réductions Statistiques

des mots, des documents
ou de la matrice mots x documents

2 0 0 3 0 6 0 0 0 1 1 0 3 0 0 1 2 0 1 0 0 0 3 0 0 2 0 ... 3

Le défi de l'analyse de texte



0 1 0 0 0 0 0 2 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 6 0 0 0 0 0 0 0 0 0 0 1 1 1 0 0 0 0 0 0 0 3 0 0 0 0 0 ... 0

Prétraitement Linguistique

Élimination des mots outils ("stop list")
Extraction des radicaux ("stemming")
Lemmatisation

2 0 0 0 1 0 1 0 0 0 0 0 1 0 0 0 0 0 0 0 6 0 0 0 0 0 0 0 0 0 0 0 1 2 0 0 0 0 0 0 3 0 0 0 0 0 ... 0

Catégorisation Sémantique

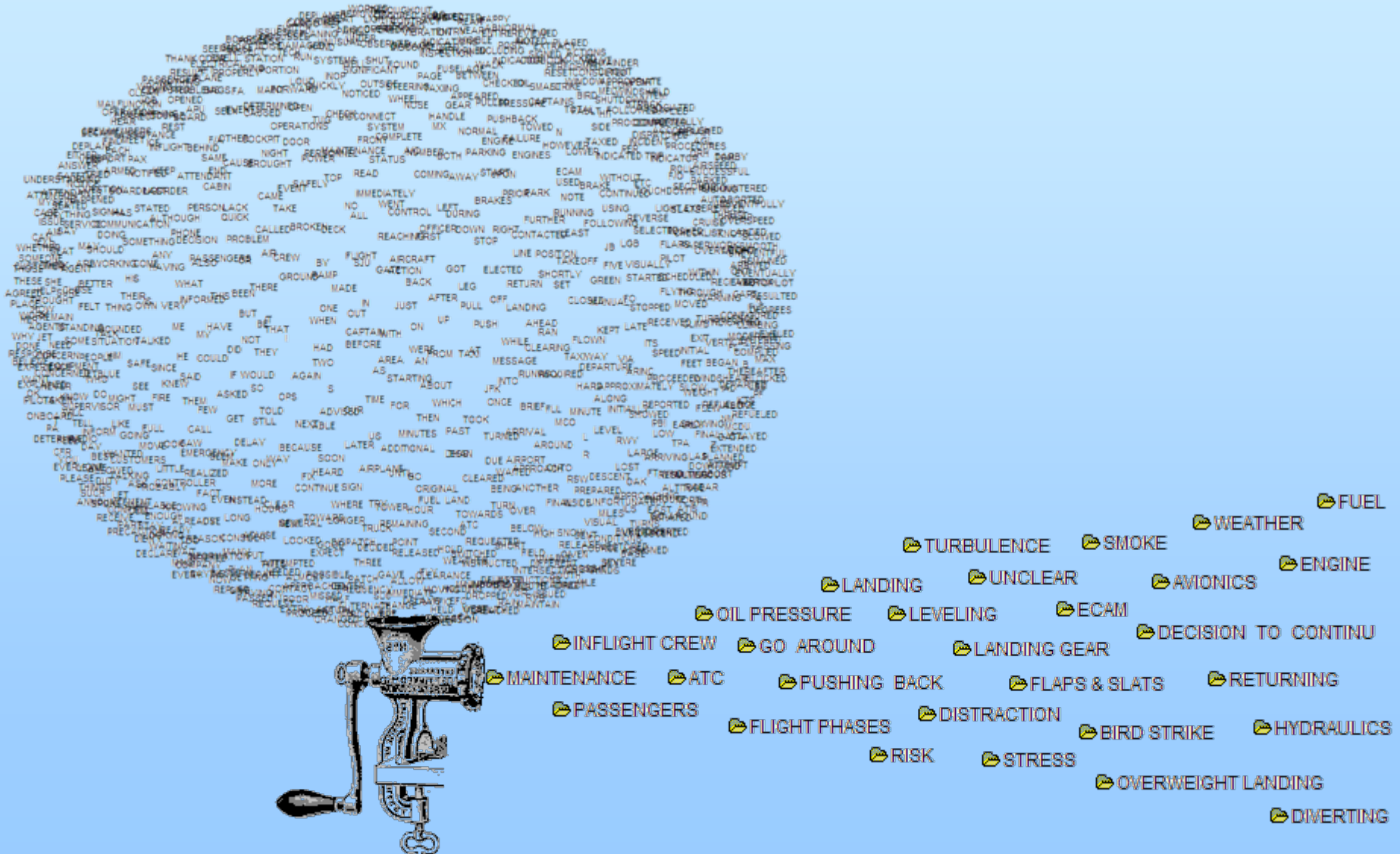
Extraction et catégorisation du vocabulaire technique, des mots signifiants et chaîne de mots

Réductions Statistiques

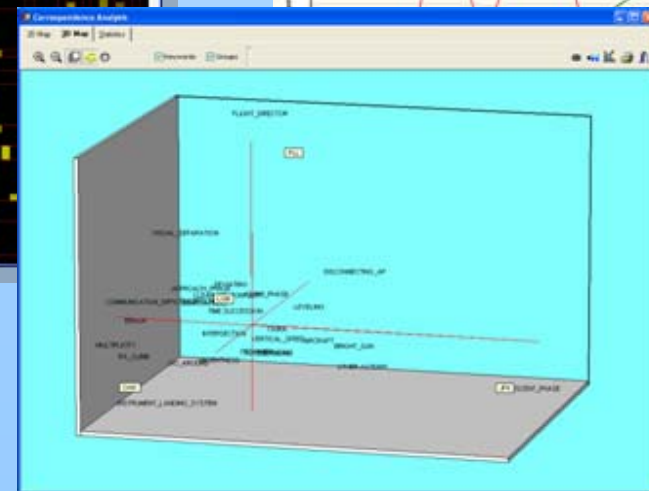
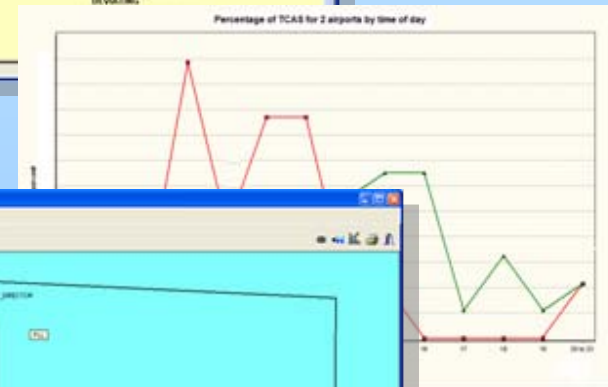
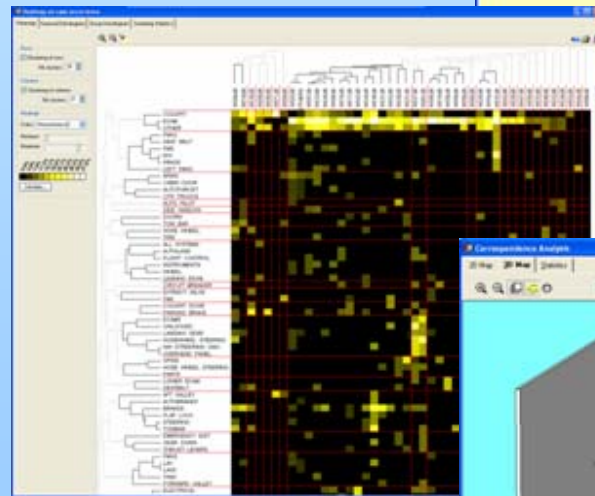
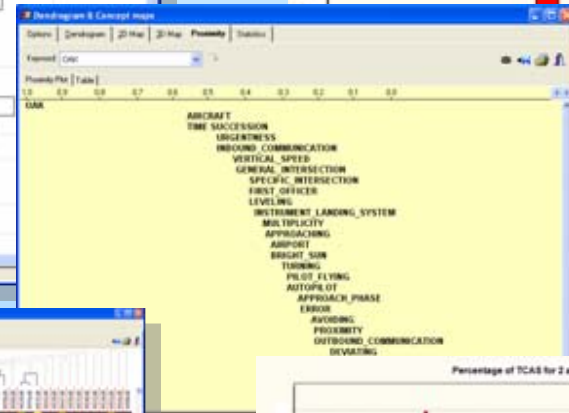
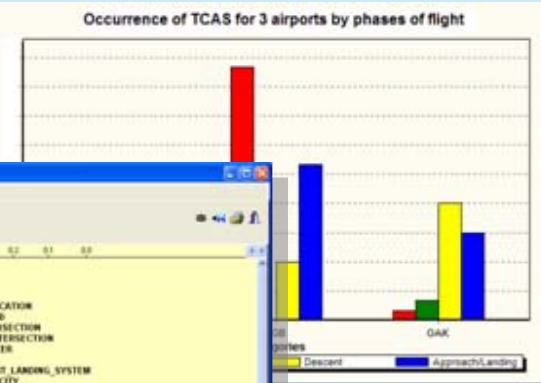
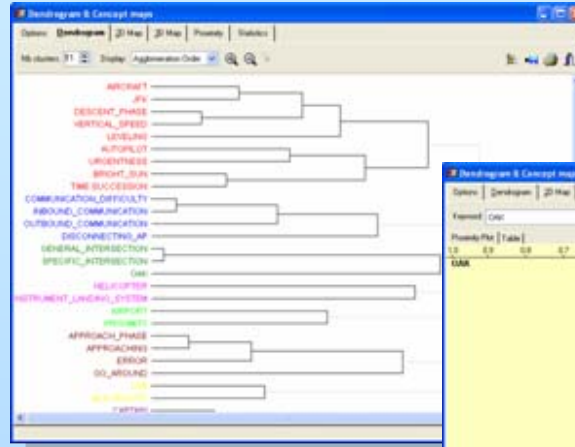
des mots, des documents
ou de la matrice mots x documents

2 0 0 3 0 6 0 0 0 1 1 0 3 0 0 1 2 0 1 0 0 0 3 0 0 2 0 ... 3

Le défi de l'analyse de texte



Le défi de l'analyse de texte



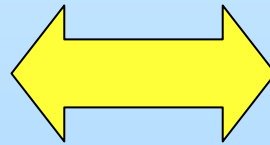
- WEATHER
- FUEL
- ENGINE
- AVIONICS
- DECISION TO CONTINUE
- RETURNING
- HYDRAULICS
- WEIGHTLANDING
- DIVERTING

Possibilité d'Analyses

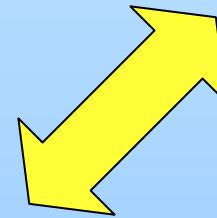
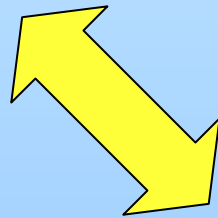
- 1. Analyse de fréquences (mots, chaînes, et catégories)**
 - Analyse univariée de fréquences
 - Comparaison avec des données normatives (fréquence de mots)
- 2. Analyse de co-occurrence & similarité de documents**
 - Classification hiérarchique, positionnement multidimensionnel (MDS), graphes de proximité
- 3. Mots clés x variables numériques ou catégorielles**
 - Tableaux croisés (avec tests statistiques), diagrammes en barres et linéaires, cartes thermiques, analyse de correspondance, etc.
- 4. Classification automatique de documents**
 - Algorithmes d'apprentissage machine dont Naive Bayes & Nearest Neighbour

Les Produits de Recherches Provalis

SIMSTAT (1989)
Analyses Statistiques



QDA MINER (2004)
Analyse Qualitative



WORDSTAT (1998)
Analyse de Texte Assistée par Ordinateur
& Exploration de Textes (text mining)

Analyse Qualitative

Bush for President Announcement

What a pleasure it is to visit with you, to shake your hands. Laura and I are so grateful for your welcome, your enthusiasm, your confidence.

There will come a time for formal speeches and 10 point plans. But I know the question on your mind: Why are you thinking about running for president? So I'll tell you what's on my heart.

I'll have a formal announcement sometime in the Fall. I have come here today to tell you this: I am running for President of the United States. There's no turning back, and I intend to be the next President of the United States.

I'm running because our country must be prosperous. But prosperity must have a purpose. The purpose of prosperity is to make sure the American dream touches every willing heart. The purpose of prosperity is to leave no one out... to leave no one behind. I'm running because my party must match a conservative mind with a compassionate heart. And I'm running to win.

Patriotism
Representativeness

Prosperity is not a given. Some in this administration think they invented it. But they did not invent prosperity, any more than they invented the Internet. Governments don't create wealth. Wealth is created by Americans -- by creativity and enterprise and risk-taking. But government can create an environment where businesses and entrepreneurs and families can dream and flourish.

Local Economy

We'll be prosperous if we reduce taxes. I'll have a plan that reduces marginal rates to create jobs, but a plan that also helps struggling families on the outskirts of poverty. I believe that after we meet priorities, all that remains must be passed back to Americans, so it will not be spent by Washington.

We'll be prosperous if we reduce the regulations that strangle enterprise. And I will do what I did in Texas: fight for meaningful, real tort reform.

We'll be prosperous if we embrace free trade. I'll work to end tariffs and break down barriers everywhere, entirely, so the whole world trades in freedom. The fearful build walls. The confident demolish them. I am confident in American workers and farmers and producers. And I am confident that America's best is the best in the world.

Globalization
Freedom

We must be prosperous to keep our commitments to the health and security and dignity of the elderly. And we should trust Americans by giving them the option of investing part of their Social Security contributions in private accounts.

And we must be prosperous to keep the peace. This is still a world of terror and missiles and madmen. And we are challenged by aging weapons and failing intelligence.

I will rebuild our military power - because a dangerous world still requires a sharpened sword.

I will move quickly to defend our people and our allies against missiles and blackmail.

And I will have a foreign policy with a touch of iron - driven by American values and American interests.

Power
Freedom

America must seize this moment. America must lead. Because America's greatest export to the world is, and always will be, freedom.

America will be prosperous and strong if we do the right things. But prosperity alone is simple materialism. Prosperity must have a greater purpose. The success of America has never been proven by cities of gold, but by citizens of character. Men and women who work hard, dream big, love their family, serve their neighbor. Values that turn a piece of earth into a neighborhood, a community, a chosen nation.

Patriotism
Tradition

- Topics
 - Environment
 - Ethic
 - Globalization
 - Local Economy
 - Power
- Values
 - Family
 - Freedom
 - Patriotism
 - Novelty
 - Protectionism
 - Race/Ethnic relation
 - Representativeness
 - Tradition

Analyse Qualitative

Report

File Edit Search

ABSALOM 8

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19

CASE: Bradley - Foreign Policy
CODE: Local Economy CODER: Admin DATE: 07/22/2005

We are at a special moment in American history, not just because we are on the eve of a new millennium, but because our country and the world are changing at a dazzling rate. There are 2 billion more people in the world market today than only 10 years ago. More of our jobs are dependent on exports than ever. Interest rates in Crystal City are set by millions of individual investors worldwide who everyday render their verdict on the economic health of the United States.

CASE: Bradley - Foreign Policy
CODE: Environment CODER: Admin DATE: 07/22/2005

Isn't it just common sense that we make sure every child in America is covered by healthcare? Isn't it just common sense that we protect our natural world from destruction, and do what it takes to achieve racial unity? Isn't it common sense that all our schools should perform well, and that more Americans should do better economically?

CASE: Bradley - Foreign Policy
CODE: Local Economy CODER: Admin DATE: 07/22/2005

Today, I'm telling the American people that if they elect me, I'll define more clearly America's role in the world. I'll guard the economic fundamentals of our prosperity and invest in our common future. I'll use the growth of the new economy and do some of the big things that need to be done in this country:

CASE: Bradley - Foreign Policy
CODE: Local Economy CODER: Admin DATE: 07/22/2005

Finally, together we can bolster the economic security of working families and thereby set the table for future economic growth. We will do fewer things, but they will be essential things, and we will do them more thoroughly.

CASE: Bradley - Announcement
CODE: Ethic CODER: Admin DATE: 07/24/2005

I believe, as the most powerful nation in the world today, we have an obligation to give the world a map to democracy, a sense of physical security against blatant aggression, and a set of economic institutions that allow more people a chance to turn their dreams for a better life into a reality. I believe in a type of leadership that doesn't stand in the spotlight as much as call attention to millions of Americans who shine everyday. I believe in the kind of leadership that faces up to our problems. Let me give you a specific example. More than 2.8 million children in America live in what is called deep poverty. They live in homes that don't even reach the poverty level. As they grow up, many of them become armed, dependent on drugs, preoccupied with a distorted sense of "respect" and uninterested in the future. We won't reach our potential as a nation until we improve this unacceptable situation .

Patriotism
Representativeness

Local Economy

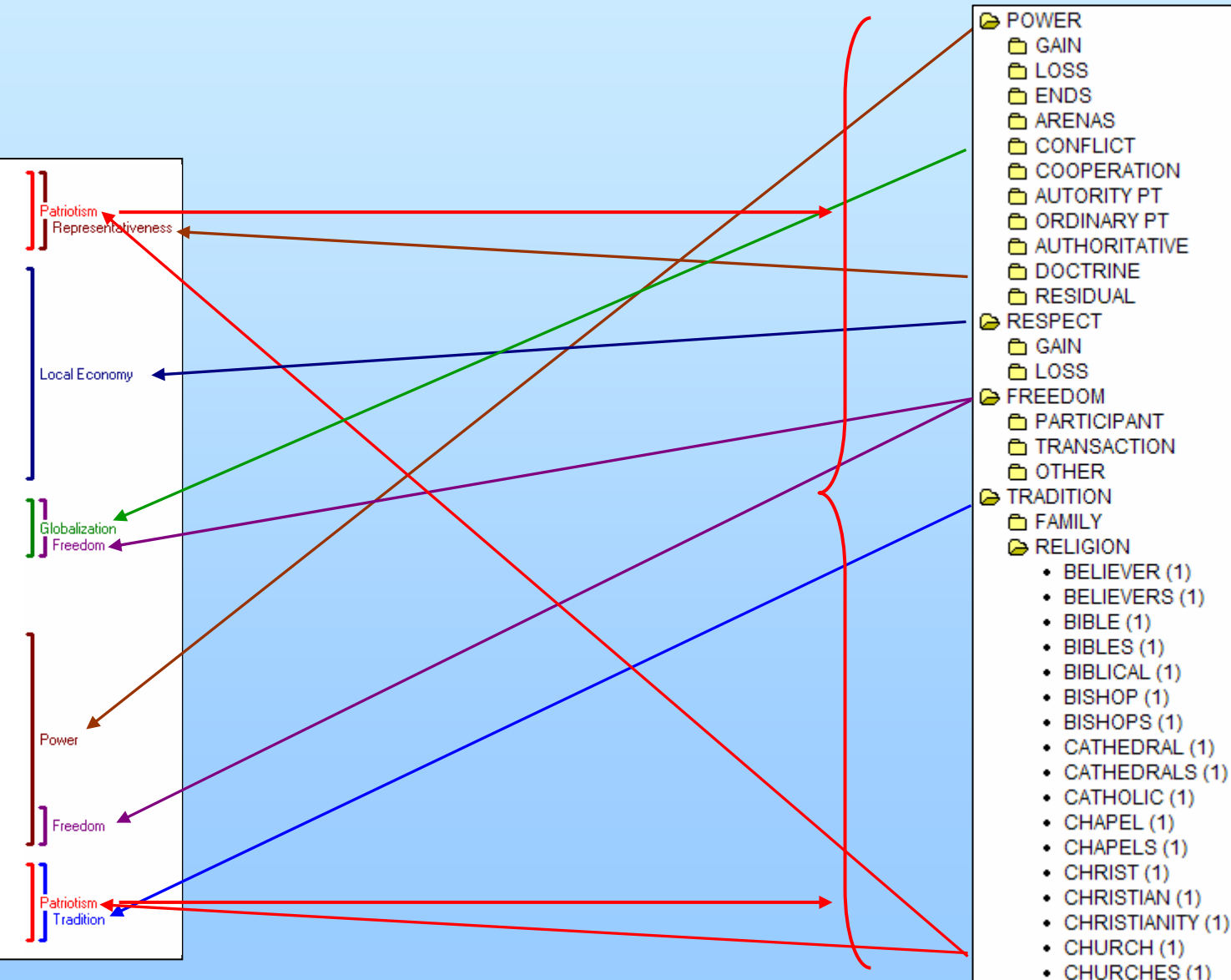
Globalization
Freedom

Power

Freedom

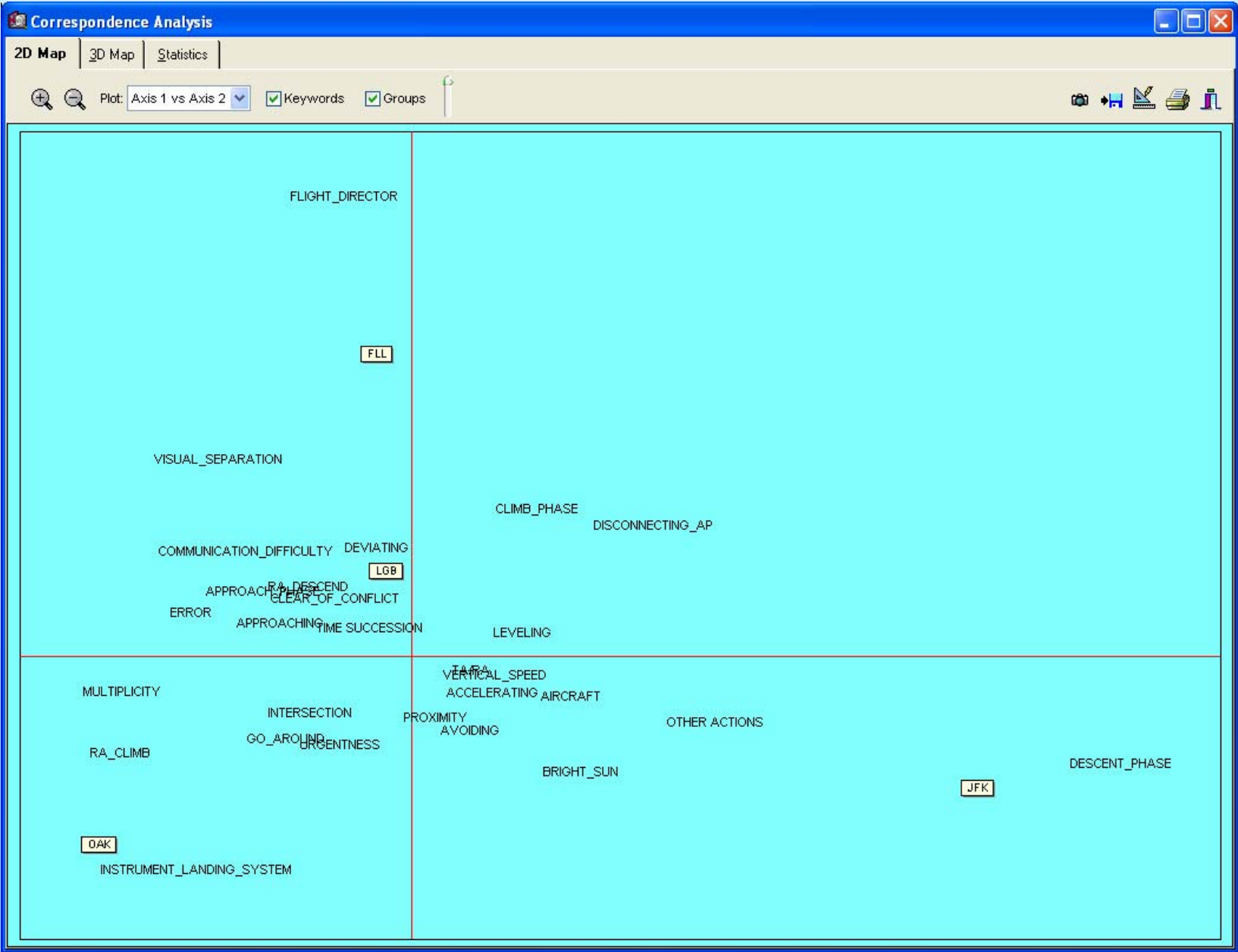
Patriotism
Tradition

Méthode Mixte



Étude de Cas #1

- **CLIENT:** Federal Aviation Administration (FAA) & JetBlue Airline
- **PRODUITS:** WordStat & SimStat
- **APPLICATION:** Découverte de connaissances
 - Identification des erreurs humaines dans les rapports d'incidents et les rapport d'irrégularités de vols.
 - Comparaison des risques de collisions à différents aéroports (analyse des rapports d'incidents TCAS)
 - Développement d'une taxonomie pour le vocabulaire de la sécurité aérienne.



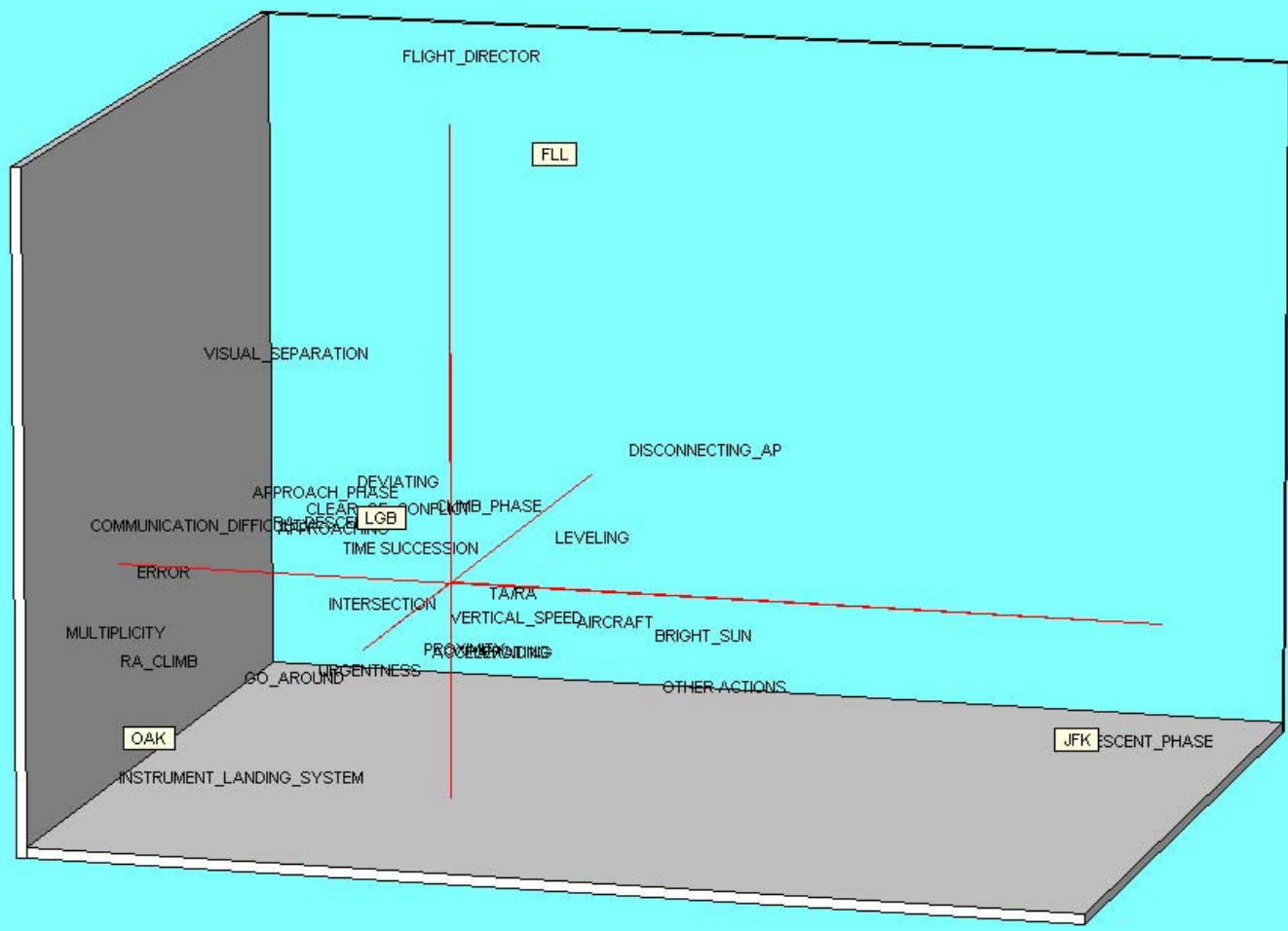
Correspondence Analysis



2D Map | **3D Map** | Statistics



Keywords Groups

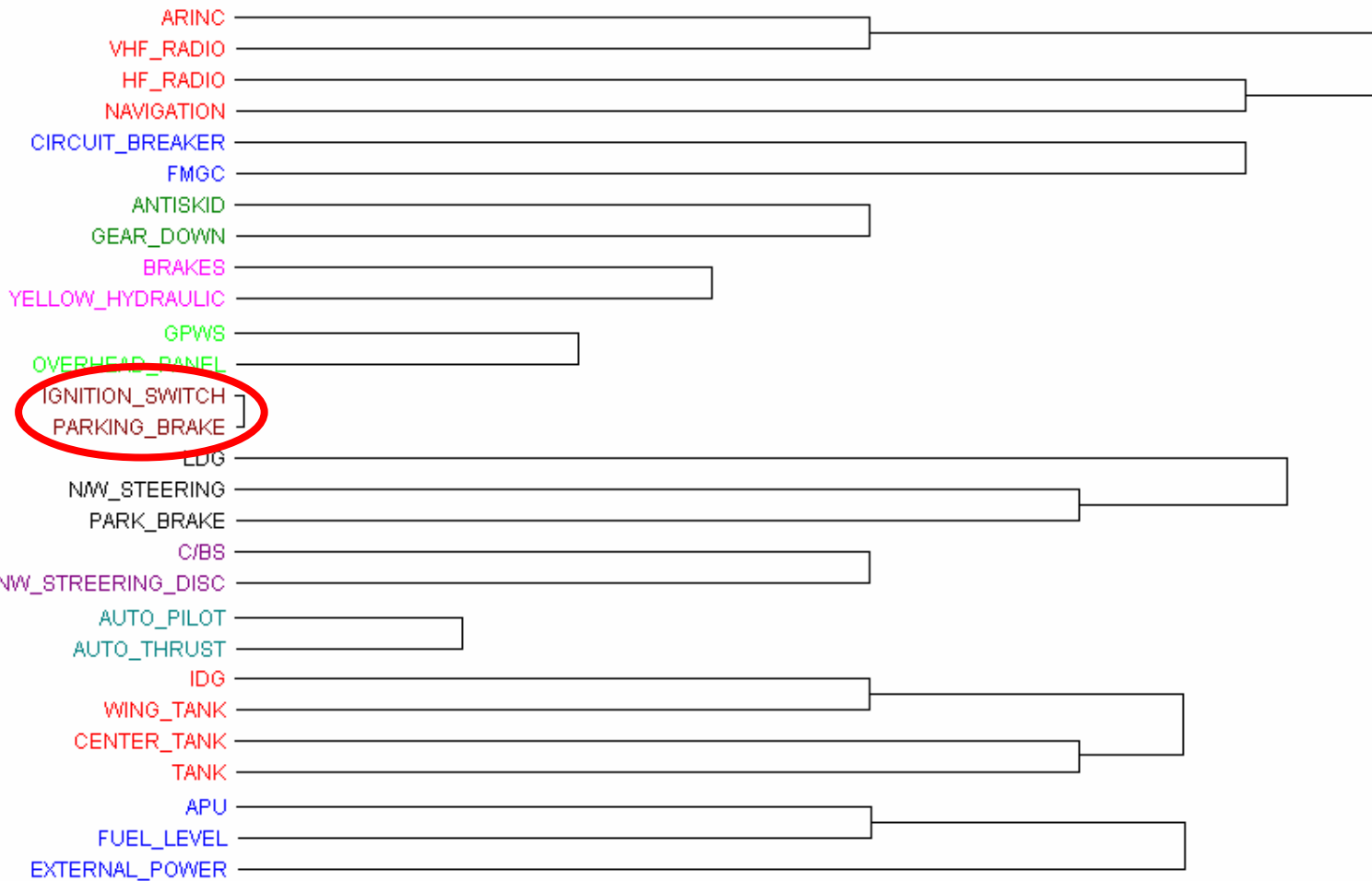


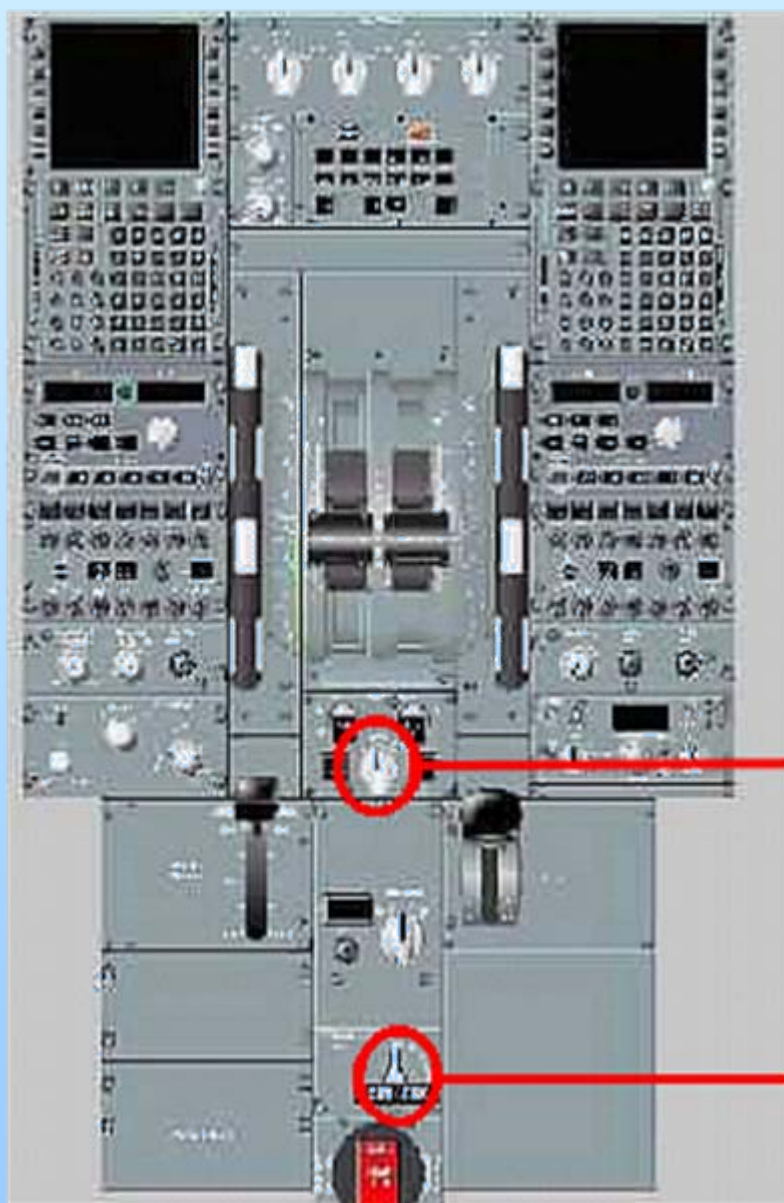
Dendrogram & Concept maps

Options **Dendrogram** | 2D Map | 3D Map | Proximity | Statistics

Nb clusters: 78

Display: Similarity Index



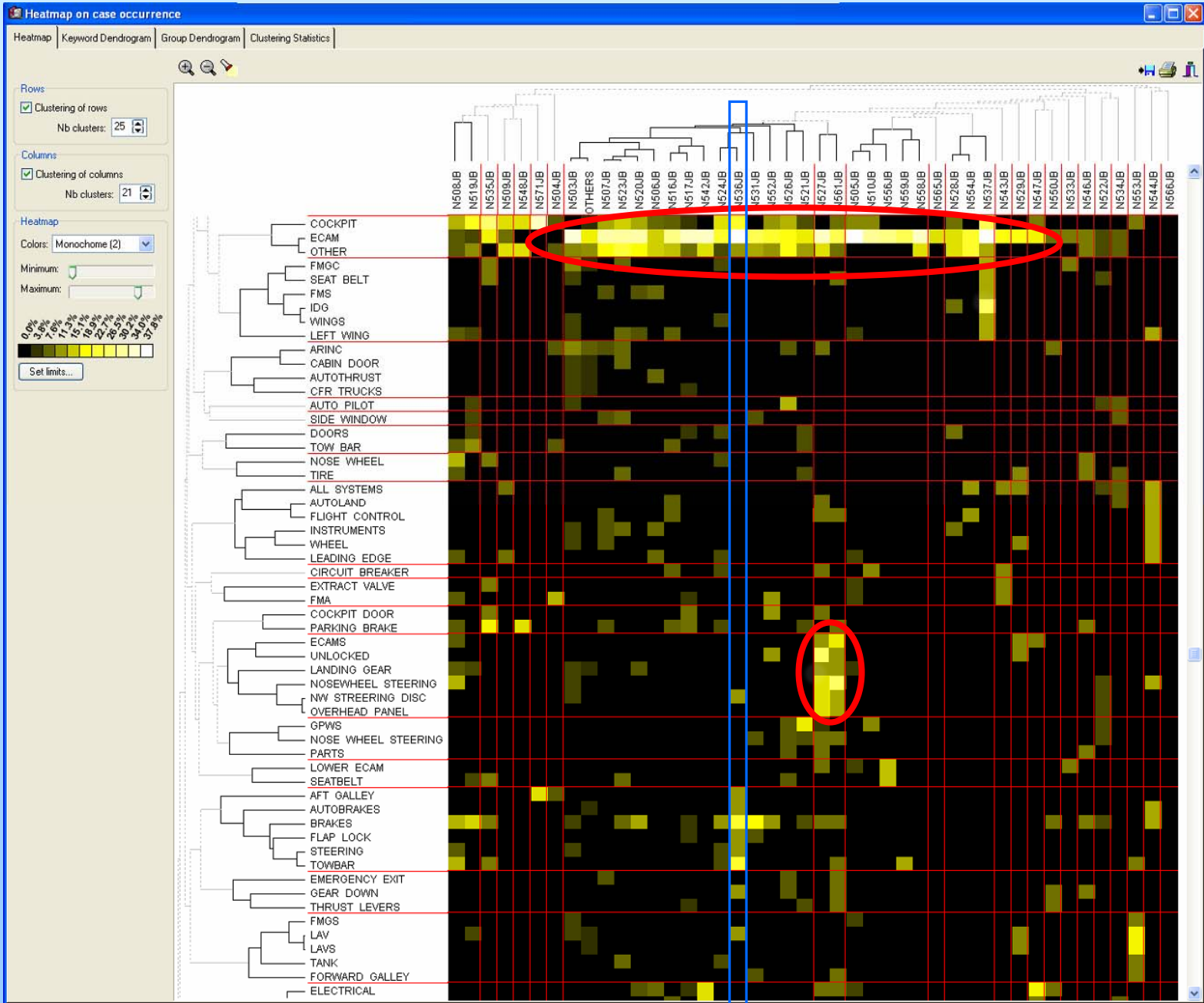


Ignition Switch

Park Brake

Los Angeles, 21 septembre 2005



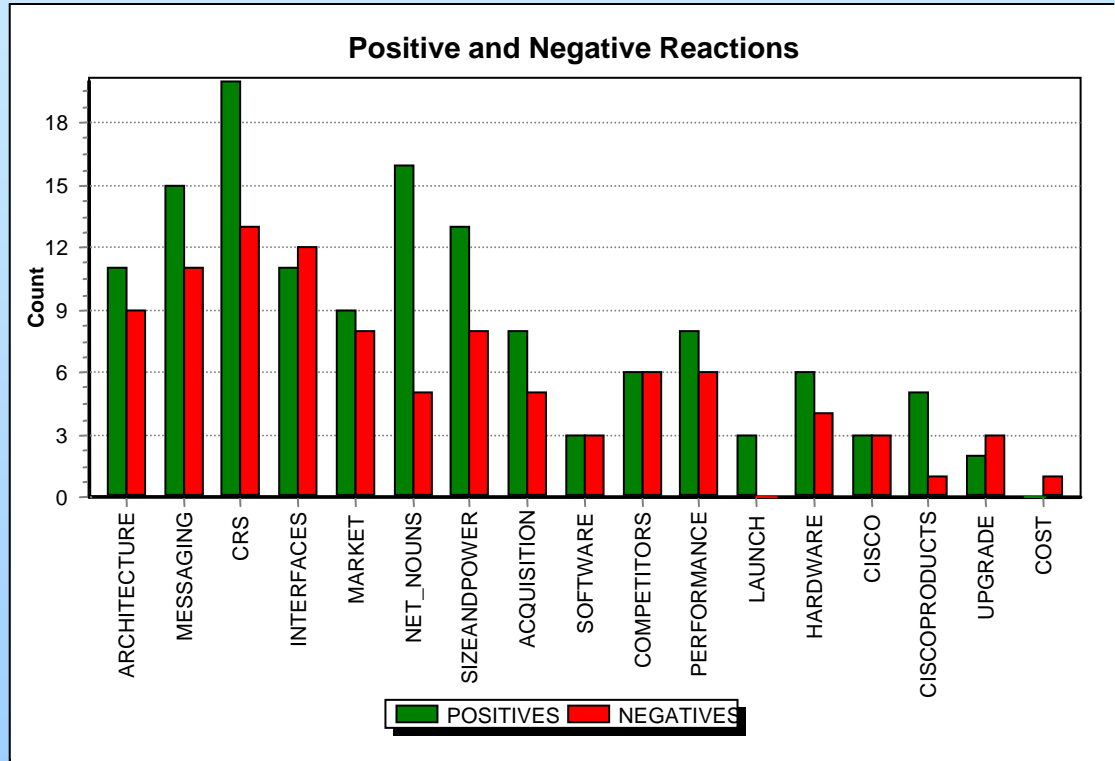


Étude de Cas #2

- **CLIENT:** CISCO Systems Inc
(Product Marketing Department)
- **PRODUITS:** WordStat, Simstat & QDA Miner
- **APPLICATION:** Étude de marché
 - Analyse de l'impact d'une campagne publicitaire (CRS-1) par une analyse de contenu des échanges sur des forums de discussion.
 - Analyse de la satisfaction de la clientèle face à différents produits et services (base de données de commentaires).



Réactions au lancement du CRS-1



Étude de Cas #3

- **CLIENT:** The Planning Commission
Hillsborough County (Florida)
- **PRODUITS:** WordStat & SimStat
- **APPLICATION:** Analyse thématique des consultations publiques sur l'aménagement urbain
 - Analyse de contenu de:
 - Environ 3000 commentaires de citoyens.
 - Transcription des assemblées communautaires et des audiences publiques.
 - Identification des préoccupations et des enjeux majeurs relatifs aux différentes communautés, différentes régions.

Étude de Cas #4

- **CLIENT:** US Office of Personnel Management
- **PRODUITS:** WordStat & SimStat
- **APPLICATIONS:** Analyse d'items de questionnaires, analyse d'incidents critiques, enquêtes
 - Identification des biais sexistes et raciaux, de langage inapproprié, d'imprécision dans les consignes ou questions.
 - Évaluation des habiletés et compétences de juges à partir de rapport d'incidents critiques.
 - Analyse des réponses à des questions ouvertes.

Étude de Cas #5

- **CLIENT:** Sciences Politiques
University of Michigan; Princeton University
- **PRODUITS:** WordStat & QDA Miner
- **APPLICATION:** Analyse de contenu thématique des documents judiciaires & des discours politiques
 - Identification des différences d'argumentation des groupes...
...en faveur des programmes de discrimination positive
...opposées aux programmes de discrimination positive
 - Changements stratégiques des politiciens russes en réponse à la mobilisation et l'utilisation des forces armées américaines en Asie Centrale suite au 11 septembre 2001.

Autres Applications

- Analyse des réponses à des questions ouvertes
- Résumé de transcription d'entrevues, groupes de discussion
- Identification des variations d'usage du vocabulaire
- Profilage de la littérature d'un domaine de recherche
- Identification des tendances dans des archives historiques
- Détermination des auteurs de documents
- Système expert pour les jeunes
- Mesure des processus et des désordres psychologiques
- Analyse des interactions nourritures - médicaments
- Détection de fraudes