

# Impact of Spontaneous Speech Features on Business Concept Detection: a Study of Call-Centre Data

Charlotte Danesi

EDF R&D

1 avenue du Général de Gaulle,  
92141 Clamart - FRANCE  
0033147654539

danesi.charlotte@gmail.com

Chloé Clavel

EDF R&D

1 avenue du Général de Gaulle,  
92141 Clamart - FRANCE  
0033147654315

chloe.clavel@edf.fr

## ABSTRACT

This paper focuses on the detection of business concepts in call-centre conversation transcriptions. In the literature, information extraction behavior has been rarely deeply analyzed on such spontaneous speech data. We highlight here the various problems that are encountered when we attempt to extract information from such data. The recall and precision, which are obtained by comparing the concept detection method on automatic *vs.* manual transcription, are respectively at 74.8% and 77.7%. We find that, even though the concept detection is similar on the whole between manual and automatic transcriptions, spontaneous speech features tend to cause different behaviors of opinion-related concept detection on both transcriptions. On the one hand, spontaneous speech features, which frequently occur in these data, provokes silence (lack of detection) when detecting concepts on both transcriptions. On the other hand, ASR errors (*e.g.* due to homophony or disfluencies) tend to provoke noise (excessive detection) when detecting concept on automatic transcription.

## Categories and Subject Descriptors

I.7.0 [*Computing Methodologies General*]: Document and Text Processing – General J.1 [*Computer Applications*]: Administrative Data - Marketing

## General Terms

Experimentation, Languages.

## Keywords

Disfluencies, spontaneous speech, conversational speech, ASR errors, linguistic pattern, call-centre data, information extraction.

## 1. INTRODUCTION

The mining of call-centre data is strategic. It contributes to improve customer insight, ensure customer satisfaction and therefore develop loyalty. So far, given the great number of available data, only few calls are listened to. Audio analysis and automatic processing of recordings (known as speech analytics methods) provide an answer to this problem.

Our goal is to extract and organize the information contained in the various phone interactions between the French power supply company EDF and the customers: what are the reasons for the

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

SSCS'10, October 29, 2010, Firenze, Italy.

Copyright 2010 ACM 978-1-4503-0162-6/10/10...\$10.00.

call? How do customer needs and preoccupations evolve? Which opinion do customers have on EDF? The text mining tools, which have been so far used by EDF are suited to written text corresponding to satisfaction inquiries [9]. Recently, EDF R&D has tackled the speech mining task through various studies such as in [10] or through the collaborative projects INFOM@GIC-Callsurf [2] [8] and Vox Factory (Cap Digital competitive cluster). A pilot in a call-centre is being put in place with the project partners. It aims at offering to the user an interface for search and navigation in call-centre data [3]. The tool chain behind this interface consists of text mining techniques in order to classify the automatic transcription of conversations. In this paper, we address one of the research issues required to improve the efficiency and the reliability of the existing solution. It consists of the adaptation of EDF text mining tools to conversational speech. We use the Skill Cartridge™ technology provided by TEMIS ([www.temis.com](http://www.temis.com)) Company, which offers the capability to model expressions linked to various *concepts* (topics, reason for satisfaction/dissatisfaction). The concept detection is thus made through rules that have not been built for transcription of conversational data containing transcription errors (on words or on punctuations) and spontaneous speech features (hesitations, stutter on beginning of words, unfinished sentences, etc.). Research has unraveled many aspects concerning the impact of ASR (Automatic Speech Recognition) errors both on spoken document retrieval thanks the TREC evaluation campaigns [6] and on named entity detection thanks to the last ESTER2 evaluation campaign [7]. Business concept detection can be compared with methods used for named entities detection (rule-based approach [15], or machine learning approach [14]). However, spontaneous speech features are more frequent in call-centre data than in the audio data that are frequently studied such as broadcast news. Besides, performance of speech recognition systems falls down on call-centre data and linguistic pattern detection is thus more difficult.

In this paper, we highlight the spontaneous speech features occurring in call centre data and their impact both on automatic transcriptions and on concept detection. Next section is dedicated to the description of the EDF call-centre corpus that is used for the present study and on the methodology that is used to detect concepts. Third section presents an analysis of concept detections on both manual and automatic transcriptions of the calls.

## 2. CORPUS AND METHOD

### 2.1 Callsurf corpus

The Callsurf corpus has been developed through the Infom@gic project. We work here on the test set of the corpus which consists

of transcriptions of 89 call centre conversations in French between EDF agent and customers. About 10 hours have been thus manually and automatically transcribed in order to evaluate the speech recognition system. The automatic transcription consists in an automatic speech-to-text conversion performed by Vecsys Company, based on LIMSI-CNRS and Vecsys Research technology [8]. The word error rate obtained on this corpus is at 29%. The manual transcription has also been provided by Vecsys and contains a total of 10770 speaker turns. The number of words by speaker turn is equivalent for the customer (8.4) and the agent (8.9). We carry out a speaker turn alignment between the two transcriptions using the manual transcription as a reference. There are 697 speaker turns with a manual transcription but without any corresponding automatic transcription and 551 with an automatic transcription but without any corresponding manual transcription. This difference between the transcriptions is due to various reasons (time codes are more accurate in the manual transcription than in the automatic one, back channels corresponding to a speaker turn are not always caught by the speech recognition system)

The Callsurf corpus contains a great number of spontaneous speech features. Indeed, people adjust their talk when speaking, are they a customer or an agent. Those adjustments are called disfluencies. Benveniste's definition [1] is the most often cited to describe French disfluencies: a disfluency occurs when the syntagmatic progress is disrupted. We present in Table 1 the various examples of disfluencies contained in Callsurf corpus:

**Table 1 - Example of disfluencies in Callsurf corpus**

Type of disfluency	Callsurf examples
Repetition [11]	« alors je je entre tout de suite la mensualisation »
Repair [11]	« j'ai vérifié elle est pas <b>de la dans les</b> documents en attente »
Fragments [4]	« ils interviennent sur la partie <b>privative</b> du client »
Restart [11]	« elle va récupérer, <b>voilà, on résiliera</b> celui-ci qui est toujours actif chez nous »
Filled Pauses [12]	« donc <b>euh</b> apparemment, il fallait bien leur dire»
Discourse Markers [12]	«bon <b>ben</b> alors , une fois qu' on a fait tout ça»

The disfluencies are not the sole spontaneous speech features of Callsurf corpus, which contains also 1274 back channels and 631 overlaps between speakers, which inhibit the manual transcription. The number of disfluencies provided by the manual transcription is at 4285. This number does not include repair, fragments, repair and repetition which are not tagged in the manual transcription. Table 2 shows the number of disfluencies according to the speaker role (customer, agent and agent/agent). Hesitations are more frequent in customer speaker turns. Indeed, customers are less conversant with specific technical terms and seem to hesitate more frequently. As expected, conversations between agents contain fewer hesitations, because they are less stressful conversations for the agent. The most frequent disfluency is the French filled pause marker "euh" (2366).

**Table 2 - Manual transcription disfluencies according to the speaker role.**

	Agent	Customer	Between Agents	Total
<b>Number of disfluencies</b>	1892	1954	439	4285
<b>Average by speaker turn</b>	0.36	0.45	0.21	0.36

## 2.2 Concept detection method

We use the Skill cartridge™ technology from TEMIS to detect concepts. It consists of a set of linguistic components such as lexicons or extraction rules, which describe the information that we want to extract (business concepts or topic). The principal goal is to model the information contained in a large corpus using adequate linguistic patterns. The first step consists in a morpho-syntactic analysis (Xelda tagger) and the second step in a semantic analysis. During this last analysis, the corpus is labeled by pre-defined concepts and sub-concepts. These concepts are modeled by a set of linguistic rules, which list all possible formulations of a concept. We use for this study a cartridge defined by EDF R&D in order to analyze customer satisfaction. Table 3 presents the main concepts defined in the cartridge, which consist both on opinion-related and topic-related concepts.

**Table 3 - Description of main concepts of EDF cartridge**

Concepts	Descriptions	Examples
<b>Duration</b>	Three sub-concepts (short, average, long duration)	Three months, every year, very long
<b>Positive-general</b>	Customer general satisfaction	No worries, no problem, is perfect
<b>Expectations</b>	Customer expectations	it would be nice (better)
<b>EDF competition</b>	Competitors	GDF, supplier, operator
<b>Contract</b>	Expressions related to the contract	Contract, subscription, read, off-peak
<b>Negative Contact</b>	Expressions related dissatisfaction on the contact	No contact, nobody has contacted you
<b>Positive Contact</b>	Expressions related satisfaction on the contact	Very good contact
<b>Bill</b>	Expressions related to bill	To follow up, actual consuming, spending, refund
<b>Price</b>	Expressions related price	It is not excessive, It is expensive, It is cheaper
<b>Technical</b>	Expressions related technical field	Commissioning, technical, meter installation
<b>EDF</b>	Expressions related to EDF	Edf pro, EDF, agent

Concepts are organized into various levels, which correspond to the order according to which rules are applied. The levels allow us to call concepts inside a concept and to manage priority between concepts. An example of level use can be found in the definition of the concept "Expectations":

```
<concept name="Les_attentes" level="3">
<e><!-- Je souhaite une réponse rapide -->
    ~Les_attentes-patt / (~ANY / ~Positif-general)? </e>
</concept>
```

We can see in this example that a concept is defined by a name - "Expectations" - and a level, the level 3 (lowest level of priority). The tag "<!-- Je souhaite une réponse rapide -->" is a comment tag with an example of the type of sentence that we want to extract. The sign "~" indicates that only the concept "Les attentes-patt" is called but not its sub-concepts, while the sign "~~" allows us to call the concept "Positif-general" anywhere it has been defined in the hierarchy (concept, sub-concept). After the cartridge has processed the corpus, the outputs are stored in an xml document including the detected concepts.

### 3. RESULTS AND ANALYSIS

#### 3.1 Detection of concepts on manual vs. automatic transcriptions

We compare here the outputs of the cartridge on the manual vs. automatic transcriptions. 16.3% of speaker turns contain at least one detected concept on manual transcription against 15.7% on automatic transcription with 2288 concepts that are detected on manual transcription and 2163 concepts that are detected on automatic transcription. 1660 concepts are identical between manual and automatic transcriptions. The number of detected concepts is thus a little bit higher in the manual transcription but are of the same order. No manual validation of concept detection is provided so far. At this first stage of our work, we focus on manual vs. automatic transcription comparison and do not carry out any manual concept validation on the corpus.

In order to confirm this observation, we compare the repartition of the different concepts that are detected for manual vs. automatic transcription (see Figure 1).

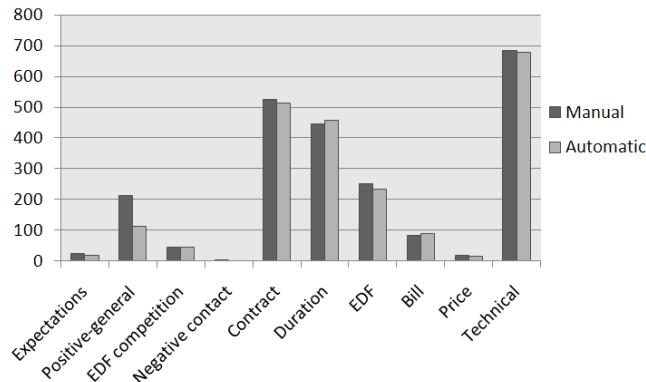


Figure 1 - Concept repartition on manual and automatic transcriptions

The global behavior of the concept detection seems to be quite similar between automatic and manual transcription. The difference is mainly located in the "Positive-general" concept (213 concepts for manual and 111 concepts for automatic). Indeed, this concept is an opinion-related concept, which is modeled by more complex linguistic patterns than topic-related

concepts are. In particular, the modeling of this concept takes into account punctuations, which are so far not present in the automatic transcription. This point has been taken into account for future transcriptions. Other opinion-related concept such as "Expectations", "Positive contact" and "Negative contact" are not well represented in this corpus. A next version of the Skill Cartridge™ will include more opinion-related concepts. This will require to pre-process or to take into account disfluencies in the linguistic patterns by completing or relaxing the rules of the cartridge.

We go then further in our analysis by studying the speaker turns where the concept detection is different between the manual and the automatic transcription. We consider here that the detection concepts obtained on the manual transcription are the reference and compute the recall and the precision. The recall is thus defined as the number of speaker turns containing the same detected concepts on manual and automatic transcriptions divided by the number of speaker turns where at least one manual concept has been detected. The precision is defined here as the number of speaker turns containing the same detected concepts on manual and automatic transcriptions divided by the number of speaker turns where at least one concept has been detected on the automatic transcription. We obtain thus a recall at 74.8% and a precision at 77.7%.

The number of missed speaker turns and the number of speaker turns that are improperly detected when proceeding to the automatic transcription are quite low (rather high recall and precision). These results confirm that the global behavior of the cartridge on the automatic transcription is close to the one obtained on the manual transcription. However, when studying the concepts, differences between the two transcriptions can be highlighted. Table 4 stores the number of concepts that are the identical and those that are different on automatic and manual transcriptions. An in-depth study of these differences is detailed in Section 3.3.

Table 4 – Number of detected concepts on automatic (A) vs. manual (M) transcription

Number of concepts that are in M and not in A	Number of concepts that are in A and not in M
277	207

#### 3.2 Detection errors on manual transcriptions

We analyze here the behavior of the detection concept on manual transcriptions. We want to know if the spontaneous speech features that are transcribed in the manual transcription may provoke problems in the concept detection through the Skill Cartridge™.

First, we focus on the impact of filled pause marker "uh" which is the most frequent disfluency in the manual transcription. We delete every "uh" present in the transcription and process the new transcription with our cartridge. The results obtained are quite the same. However, we found some examples where the hesitation can inhibit the detection of concept: "Le compteur euh euh à poser" ("The meter uh uh to install"). The concept "technical" can't be detected. Second, we analyze the impact of discursive markers (e.g. "ben"). We find that this impact is quite weak. Indeed, discursive markers are link words and they frequently occur between sentences and clauses.

### 3.3 Detection errors on automatic transcriptions

We analyze here the behavior of the detection concept on automatic transcription. We want here to identify what are the types of ASR errors, which provoke wrong concept detections.

A first problem concerns the error in the transcription of named entities such as the name or the city of customer. Such errors provoke incorrect detections, such as in the following example:

Manual: Madame Maubert ("Mrs Maubert")  
Automatic: Madame ampere ("Mrs ampere")  
(Concept = technical)

In this example, the concept "Technical" is improperly located through the word "ampere". Besides, in this example the error is localized but in some cases the transcription error on named entity may spread in the entire sentence transcription:

Manual: donc si j'ai aucun contrat avec zone industrielle Carpiquet  
Automatic: donc les clients comme contrat avec index avec appliquer

The incorrect detections could be reduced by filtering words or phrases which have a low ASR confidence score.

A second problem concerns a frequent phenomenon in French language that is homophony. For example, the system has some difficulty to distinguish the following words: "moi / mois" ("me / month") and "an / en" ("year / in"), which provokes mistakes in the detection of the "Duration" concept.

A third problem, which we identify, is caused by speech disfluencies. For example, the word "long" followed by "hein" is transcribed by the word "loin" ("far"). These errors come from the recognition errors like this example.

Manual : leur euh (" their uh ")  
Automatic : l'heure ("the hour") (Concept = Duration)

These two last problems could be solved by carefully completing the rules of the Skill Cartridge™ with contextual rules and dictionary extension.

Besides, we notice that back channels are rarely transcribed. When they are transcribed, they are most of the time subjected to ASR errors. For example, the back channel "hm" is frequently transcribed by "donc / nous / même / mais" ("Therefore / we / same / but"). Such ASR errors do not lead to improperly detections of concept but could be problematic when dealing with dialogic information.

## 4. CONCLUSIONS AND PERSPECTIVES

Concept detection on automatic transcriptions of call-centre data gives quite similar results to concept detection on manual transcriptions. The recall and precision, which are obtained by comparing the concept detection method on automatic vs. manual transcription, are respectively at 74.8% and 77.7%. However, we notice some local different behaviors especially for the detection of opinion-related concepts relying on more subtle linguistic patterns.

The next step of our work will consist on analyzing the behavior of other EDF cartridges relative to opinion analysis and propose a new strategy for EDF cartridge adaptation on call-centre conversations by taking into account spontaneous speech features

and dealing with speech recognition errors. Besides, we would like to explore other types of information than topics addressed during the call and customer satisfaction such as describing quality and efficiency of the conversation by analysing linguistic or acoustic manifestations of emotions such as in [5] or studying conversation profiles. In this context, a focus on spontaneous speech features detection could be of interest.

## 5. REFERENCES

- [1] Blanche-Benveniste, C. (1990). *Le français parlé : Etudes grammaticales*. Paris.
- [2] Bozzi, L., Suignard, P. (2009). *Segmentation et classification non supervisée de conversations téléphoniques automatiquement retranscrites*. TALN, Senlis.
- [3] Caillau, F., Giraudel A. (2008). *Enhanced Search and Navigation on Conversational Speech*. Actes de Searching Spontaneous Conversational Speech (SSCS 2008), SIGIR Workshop. Singapour.
- [4] Cheng-Tao Chu, Y.-H.S. (2003). *Detection of Word Fragments in Mandarin Telephone Conversation*. Stanford
- [5] Clavel, C., Richard, G., *Reconnaissance acoustique des émotions*, in Systèmes d'Interaction Emotionnelle, 2010, chap. 5, Hermès.
- [6] Coden, A. (2001). *Information Retrieval Techniques for Speech Applications*. Book based on the workshop "Information Retrieval Techniques for Speech Applications", ACM SIGIR
- [7] Galliano, S., Gravier, G., Chaubard, L. (AFCP). *The ESTER 2 evaluation campaign for the rich transcription of French radio broadcasts*. Interspeech, Brighton
- [8] Garnier-Rizet, M., Adda G., Caillau F., Gauvain J. L., Guillemin-Lanne S., Lamel L., Vanni S., Waast-Richard C. (2008). *CallSurf - Automatic transcription, indexing and structuration of call center conversational speech for knowledge extraction and query by content*, LREC 2008.
- [9] Kuznick,L., Guénet A.-L., Peradotto, A., Clavel, C. (2010) L'apport des concepts métiers pour la classification des questions ouvertes d'enquête, TALN, Montréal.
- [10] Lavalley, R., Clavel, C., El-Bèze, M., Bellot, P. (2010) Finding topic-specific strings in text categorization and opinion mining contexts, DMIN, Las Vegas.
- [11] Shriberg, E. (1994). *Preliminaries to a theory of speech disfluencies*. University of California.
- [12] Vasilescu, I., Rosset, S. and Adda-Decker, M. (2010) *On the Role of Discourse Markers in Interactive Spoken Question Answering Systems*. LREC, Valletta.
- [13] Zellner, B. (1992). *Le be - begayage et euh...., l'hésitation en français spontané*. PARIS 7
- [14] GuoDong Zhou, Jian Su (2002). *Named Entity Recognition using an HMM-based Chunk Tagger*. ACL, Philadelphia
- [15] Humphreys,K. , Gaizauskas,R., Azzam,S., Huyck,C., Mitchell,B., Cunningham,H., Wilks,Y.. *Description of the LaSIE-II system as used for MUC-7*. MUC-7, Fairfax, Virginia.199