

La longueur des tours de parole comme critère de sélection de conversations dans un centre d'appels

Philippe Suignard¹ Frederik Cailliau² Ariane Cavet²

(1) EDF R&D, 1, avenue du Général de Gaulle, 92141, Clamart

(2) Sinequa, 12 rue d'Athènes, 75009 Paris

Philippe.Suignard@edf.fr, cailliau@sinequa.com, cavet@sinequa.com

RESUME

Cet article s'intéresse aux conversations téléphoniques d'un Centre d'Appels EDF, automatiquement découpées en « tours de parole » et automatiquement transcrites. Il fait apparaître une relation entre la longueur des tours de parole et leur contenu, en ce qui concerne le vocabulaire qui les compose et les sentiments qui y sont véhiculés. Après avoir montré qu'il y a un intérêt à étudier ces longs tours, l'article analyse leur contenu et liste quelques exemples autour des notions d'argumentation et de réclamation. Il montre ainsi que la longueur des tours de parole peut être un critère utile de sélection de conversations.

ABSTRACT

Turn-taking length as criterion to select call center conversations

This article focuses on telephone conversations collected in an EDF Call Center, automatically segmented in "turn-taking" and automatically transcribed. It shows a relationship between the length of the turns and their content regarding the vocabulary and the feelings that are conveyed. After showing that there is an interest in studying these long turns, the article analyzes their content and lists some examples around the notions of argumentation and claim. It shows that the length of turns can be a useful criterion for selecting conversations.

MOTS-CLES : Centre d'appels, Conversation, Tour de parole, Reconnaissance de Parole.

KEYWORDS : Call Center, Conversation, Turn Taking, Automatic Speech Recognition.

1 Introduction

Avec plus de 30 millions de clients et plusieurs milliers de conseillers en ligne, les Centres d'Appels constituent, pour EDF, un maillon important de la Gestion de la Relation Client et font l'objet d'un suivi permanent avec un focus sur la « professionnalisation des conseillers », consistant à améliorer leur pratique professionnelle afin de toujours mieux répondre aux clients. Cette amélioration passant par des analyses qualitatives, seul un faible pourcentage peut être retenu pour écoute, d'où l'importance des critères de sélection.

C'est dans cet esprit qu'EDF R&D a participé aux projets Infom@gic/Callsurf et Voxfactory. Le projet Callsurf (Garnier *et al.*, 2008 ; Bozzi *et al.*, 2009) consistait à enregistrer et transcrire automatiquement les conversations entre clients et conseillers, pour ensuite les analyser. Le projet Voxfactory y ajoute la détection automatique de l'émotion véhiculée, à partir du texte (Cailliau et Cavet, 2010), et par le signal (Devillers *et al.*, 2010).

Cet article s'intéresse à la notion de « tours de parole » (Sacks *et al.*, 1974) et à leur longueur mesurée en seconde, une information qui ne semble pas encore avoir fait l'objet d'étude,

contrairement à la longueur des phrases, en nombre de mots, dans des données textuelles plus conventionnelles comme l'étude du théâtre du XVII^e siècle par (Labbé et Labbé, 2010).

La suite de l'article s'intéresse à la relation entre la longueur des tours de parole et le vocabulaire qui les compose, ainsi qu'à la relation entre la longueur de ces tours et les sentiments qu'ils véhiculent. Elle montre que lorsque la longueur du tour a tendance à augmenter, les informations trouvées semblent plus chargées émotionnellement et le vocabulaire employé a tendance à devenir plus pertinent d'un point de vue métier.

La partie 2 présente la notion de tour de parole et la partie 3 présente le corpus. La partie 4 s'intéresse à la relation entre longueur du tour de parole et vocabulaire, tandis que la partie 5 décrit la relation entre longueur et sentiment. La dernière partie analyse leur contenu et liste quelques exemples autour des notions d'argumentation et de réclamation.

2 Notion de tour de parole

La notion de « tour de parole » (TDP) correspond à la prise de parole par un locuteur et désigne le temps pendant lequel il garde cette parole. Au final, la suite ordonnée des TDP va constituer une conversation (Vincent, 2002). Le tour de parole semble donc être une notion assez simple, mais bien que beaucoup utilisé en « analyse conversationnelle », il reste sujet à interprétation et beaucoup d'interrogations à son sujet subsistent. Différents travaux se demandent encore « Qu'est-ce que vraiment un TDP ? » (Laforest, 2011).

Sur la figure 1, l'exemple (1) semble présenter trois TDP, mais l'intervention de « l'agent » n'est qu'un simple « back-channel » pour manifester son attention au client. En analyse conversationnelle, on pourrait regrouper ces trois tours en un seul et les considérer comme étant une seule unité. Dans l'exemple (2), le client a du mal à trouver le nom d'un terme métier, que lui souffle l'agent. Ici aussi on peut considérer qu'il s'agit d'un seul tour de parole et qu'il est co-construit par les deux locuteurs.

(1) Client : « ... oui, je vous appelle... » Agent : « oui » Client « ... pour un problème... »	(2) Client : « ... et quand je vais la recevoir, la ... » Agent : « la facture rectificative ? »
---	---

FIGURE 1 - Deux exemples de tours de parole

Sans vouloir éluder la difficulté de définir précisément ce « tour de parole », nous nous en affranchissons un peu dans cette étude, dans la mesure où nous nous intéressons à des conversations automatiquement découpées en TDP par un segmenteur et retranscrites par un procédé de transcription automatique de la parole (Adda *et al.*, 2011). Au final, le découpage en TDP est imparfait mais reflète la réalité des conversations (répétitions, disfluences, parole superposée, etc.) ainsi que les conditions d'enregistrement (voix sur un seul canal, bruit de fond, téléphone portable, etc.).

Plusieurs raisons peuvent expliquer la variation de longueur des TDP : un client peut monopoliser la parole pour exprimer un problème, une réclamation, une insatisfaction, etc., mais un conseiller peut également monopoliser la parole et prendre du temps pour répondre au client (parce qu'un problème peut être compliqué à résoudre par exemple). Enfin, une discussion peut aussi contenir des passages « serrés » ou « tendus » (énervements, émotions...), ce qui peut empêcher le segmenteur de détecter le changement

d'interlocuteur. D'un certain point de vue, cela peut être considéré comme un point faible du système, mais de l'autre c'est également un marqueur du fait qu'il se passe quelque chose de particulier et qu'il s'agit donc d'un phénomène intéressant à étudier.

3 Présentation du corpus

Pour réaliser notre étude, un corpus de conversations téléphoniques a été enregistré dans le Centre d'Appels Bleu Ciel d'Aix-en-Provence. Les enregistrements ont eu lieu entre janvier et février 2010, auprès d'une quinzaine de conseillers volontaires. Comme la plupart des enregistreurs du marché, l'enregistreur de conversation utilisé est mono-canal, ce qui signifie que les deux signaux de parole du client et du conseiller se superposent quand ils parlent en même temps. Une fois enregistrés, les appels font l'objet d'une série de traitements. D'abord intervient le « segmenteur », qui a pour but de séparer le signal en segments qui, idéalement, correspondent aux tours de parole du client et du conseiller. Ensuite, les TDP sont transcrits par un processus de transcription automatique (avec un taux d'erreur d'environ 30 %).

Au total, ce corpus est constitué de 8 551 conversations, composées de 800 596 tours de parole. La durée moyenne d'un TDP est de 3,7 s, le plus long dure plus de 2 min. La répartition du nombre de tours en fonction de leur durée est présentée en figure 2 (échelle linéaire à gauche et logarithmique à droite) :

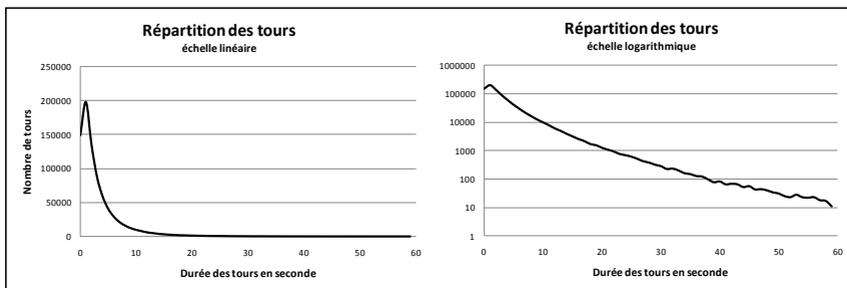


FIGURE 2 - Distribution des tours de parole en fonction de leur durée

Ces courbes montrent que la très grande majorité des TDP dure moins de 10 s et que leur nombre chute fortement en fonction de la durée. Néanmoins, il y a quand même des tours longs, voire très longs. Si l'on considère les TDP dont la durée est supérieure à 20 s, il n'y en a « que » 1,2 % en nombre mais 9 % en durée. Pour un seuil de 40 s, on en trouve 0,12 % en nombre et 1,7% en durée.

Quand on regarde ces données au niveau des conversations, on s'aperçoit que le nombre de conversations ayant au moins un tour de plus de 20 s est de 3 904 soit 45 % et que le nombre de conversations ayant au moins un tour de plus de 40 s est de 674 soit 7,9 %. En conclusion, on peut dire que les longs TDP sont peu fréquents par rapport à l'ensemble des TDP, mais non négligeables si on prend en compte leur durée totale.

4 Longueur des tours et vocabulaire : une relation ?

De manière assez intuitive, les TDP très courts semblent peu porteurs d'information. Dans ceux-ci, on trouvera beaucoup de : « oui », « non », « EDF bleu ciel bonjour », etc. Par contre, dans des TDP plus longs, une conversation peut s'installer, le client peut présenter la raison de son appel, le conseiller va répondre à sa problématique, faire des propositions, etc. La question qu'on se pose ici peut se résumer de la manière suivante : « Est-ce que l'on parle des mêmes choses dans les tours longs que dans les autres ? »

Pour répondre à cette question, la méthode va être la suivante. Pour une durée d donnée, on va constituer deux corpus : $T_{inf}(d)$, le corpus constitué des TDP dont la durée est inférieure à d et $T_{sup}(d)$, le corpus constitué des TDP dont la durée est supérieure à d . Puis, nous calculons la distance entre ces deux corpus comme décrit dans (Labbé et Labbé, 2003) :

$$dist(d) = \sum_{m \in T_{inf}(d) \cup T_{sup}(d)} |f(m, T_{inf}(d)) - f(m, T_{sup}(d))| \quad \text{et} \quad f(m, T(d)) = \frac{occ(m, T(d))}{|T(d)|}$$

Avec $f(m, T(d))$ la fréquence du mot m dans le corpus $T(d)$, $occ(m, T(d))$, le nombre d'occurrences du mot m dans le corpus $T(d)$ et $|T(d)|$, le nombre total de mots dans le corpus $T(d)$. Pour le calcul des mots m , deux alternatives sont retenues : les unigrammes et les bigrammes étendus.

Pour les unigrammes, on constitue la liste de tous les mots du corpus sauf ceux faisant partie d'une « stop liste », comme « le », « la », « les », etc. Pour les bigrammes étendus, on constitue la liste de toutes les suites de deux mots dont aucun des deux n'appartient à la « stop liste » et de toutes les suites de trois mots composées d'une préposition au milieu et dont les deux mots extrémités sont absents de la « stop liste ». On détecte ainsi les bigrammes comme « heures pleines », « relevé de compteur », « pompe à chaleur », etc.

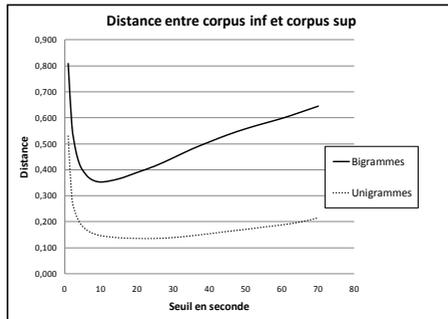


FIGURE 3 – Distance entre corpus inférieur et supérieur à une certaine durée

La figure 3 montre que les deux courbes de distance (pour les unigrammes et les bigrammes) ont globalement la même allure. Quand le seuil de séparation entre les deux corpus est très petit, la distance entre les deux corpus est élevée ce qui semble normal, puisque dans les tours les plus courts, on trouve les « oui », « non », « bonjour », etc. Ensuite cette distance décroît pour atteindre un minimum situé entre 10 s et 20 s. Puis l'écart croît

de nouveau.

Ce graphique montre qu'il y a bien une relation entre la longueur des TDP et le vocabulaire de ceux-ci. C'est vers 15 s que se situe une sorte d'équilibre, c'est-à-dire que c'est pour un seuil de 15 s que le corpus des TDP inférieurs à cette durée ressemble le plus au corpus des TDP supérieurs à cette durée. Cela justifie également l'idée d'aller étudier les tours ayant une durée supérieure à 20 s pour y trouver des particularités ou spécificités de vocabulaire.

5 Longueur des tours et sentiment : une relation ?

5.1 Trouver du sentiment dans un tour de parole

Notre analyse du sentiment est faite en trois phases : détection et normalisation des expressions, calcul d'un poids positif et négatif pour chaque tour de parole et calcul des zones positif et négatif. Les deux premières phases sont détaillées dans (Cailliau et Cavet, 2010) et seront brièvement rappelées ici.

Nous appliquons tout d'abord un ensemble de grammaires faites manuellement à partir d'une fouille approfondie du corpus. Elles détectent les mots et expressions porteurs de sentiment dans le cadre typique de la conversation téléphonique pour un total de 30 types d'entités. A chaque type nous avons attribué un poids déterminé de façon empirique, ainsi qu'une orientation positive, négative ou neutre. Ensuite chaque tour de parole obtient un score de polarité positif et un score de polarité négatif correspondant à la somme des poids des entités repérées dans le tour. Les poids des extractions neutres sont ajoutés au plus haut score positif ou négatif. L'exemple suivant d'un tour de parole contient plusieurs entités, avec le calcul des poids positif et négatif, illustré en figure 4.

« Oui oui mais il non mais c'est ça **OK tant mieux** ailleurs parce que **sinon** ça serait **dur** à sortir »

Extraction	Classe et sous-classe	Poids positif	Poids négatif
OK	Acceptation – Refus : acceptation	2	-
tant mieux	Appréciation : favorable (émotif)	4	-
sinon	Accord – Désaccord : rectificatif	-	1
dur	Appréciation : défavorable (émotif)	-	4
Poids total du TDP		6	5

FIGURE 4 – Poids des entités repérées dans le tour de parole

5.2 Du tour de parole au passage

Comme le montrent indépendamment (Cailliau et Cavet, 2010) et (Danesi et Clavel, 2010), le taux d'erreur de la transcription automatique d'environ 30 % impacte directement les extractions. Pour minimiser cet impact, nous privilégions le passage au TDP, en partant du principe qu'un TDP sentimentalement marqué apparaît rarement isolé. Pour ce faire, nous avons mis en place un lissage par une fenêtre coulissante de 5 tours de parole.

Un algorithme de zonage nous permet de transformer la courbe en zones neutre (□), positive (▤), négative (■), et très négative (■). Les seuils ont été définis de façon empirique et la barre colorée est obtenue par projection des zones sur l'échelle temporelle.

La figure 5 présente les courbes positive et négative obtenues sur une conversation après lissage sur l'échelle du tour de parole et sur l'échelle de temps, avec la barre colorée correspondante. Le pic négatif apparaissant au début de la conversation occupe relativement peu de tours de parole (17 TDP sur 55, soit 30,9 %), mais il occupe beaucoup de temps sur la conversation (environ 300 s sur 590 s, soit 50,8 %).

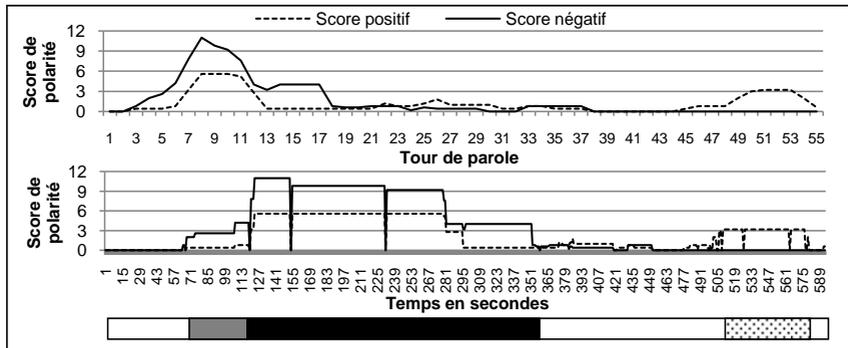


FIGURE 5 – Courbes positive et négative d’une conversation et zones correspondantes

5.3 Longueur des tours de parole et couleur associée

Si on conserve les seuils de 20 s et 40 s utilisés en partie 3, on constate, en figure 6, que les TDP courts apparaissent très majoritairement dans des zones non marquées (95 %). Le nombre de TDP dans des zones non neutres augmente fortement avec la longueur de ces TDP, au détriment du non marqué. Les TDP longs contiennent donc généralement plus d’expressions émotionnelles que les TDP courts. Ces résultats ne sont pas si surprenants : notre méthode de calcul privilégie les TDP plus longs car nous utilisons la somme des poids des entités extraites pour calculer les poids positif et négatif des TDP. On remarque également qu’il ne suffit pas d’avoir un TDP long pour avoir du sentiment : 45 % d’entre eux sont neutres, et moins de 8 % sont marqués comme « très négatifs ».

Durée des TDP	Neutre	Positif	Négatif	Très négatif	
< 20 s	95,3 %	3,0 %	1,6 %	0,1 %	100 %
entre 20 s et 40 s	74,8 %	13,1 %	10,4 %	1,7 %	100 %
> 40 s	45,0 %	25,5 %	21,6 %	7,8 %	100 %

FIGURE 6 – Longueur des TDP et émotion détectée

6 Trouver la spécificité de vocabulaire par régression linéaire

Les deux parties précédentes ont permis de montrer qu’il y avait une spécificité des longs TDP par rapport aux autres : ils sont plus « chargés » en émotion et ont une spécificité en termes de vocabulaire. Dans cette partie, nous appliquons une méthode basée sur la fréquence des mots pour identifier lesquels sont à l’origine de cette spécificité.

On commence par éliminer les tours très courts (inférieurs à 10 s, parce que très nombreux et très peu porteurs d'informations) et très longs (supérieurs à 80 s, parce que très peu nombreux et pouvant donc perturber fortement les résultats) pour constituer 7 sous-corpus : T_{10-20} , T_{20-30} , T_{30-40} , T_{40-50} , T_{50-60} , T_{60-70} , T_{70-80} , T_{10-20} étant le corpus constitué des TDP dont la durée est comprise entre 10 s et 20 s. Pour chaque sous-corpus, on calcule les fréquences des mots. On pratique une régression linéaire pour trouver les mots pour lesquels l'augmentation de fréquence est la plus forte. Une analyse des mots présentant la plus forte augmentation permet de discerner deux sphères : la sphère de l'argumentation et la sphère de la réclamation.

La sphère de l'argumentation est caractérisée par une forte présence d'adverbes comme « effectivement », « exactement », « normalement », « maintenant », « directement » et « justement ». Cette présence des adverbes s'explique par le fait que les longs TDP ont tendance à être d'avantage consacrés aux problèmes les plus complexes et qu'ils donnent lieu à des échanges argumentés à la fois du côté du client pour expliquer son problème mais également du côté du conseiller pour justifier une réponse. Voici quelques extraits :

Agent : « Avant, vous étiez dans une période où **effectivement** le client pouvait choisir... **effectivement** ses heures creuses... »

Agent : « L'intervention coûtera xx €, somme que vous n'aurez pas à payer, elle sera prélevée **effectivement** sur la facture... »

Agent : « Une somme de x €, **effectivement** ce n'est pas négligeable... »

Client : « ... je suis propriétaire mais la locataire a dû elle avait téléphoné **justement** à edf pour résilier le contrat et moi en fait c'est quelqu'un qui a signé de le jour que j'ai emménagé et **justement** pour couper edf j'avais dit que **justement** j'avais fait l'ouverture sur internet... »

La seconde sphère est celle de la réclamation, caractérisée par la présence de mots comme « problème », « réponse », ou directement « réclamation », comme par exemple :

Agent : « ... d'accord avec tous les points où il y a eu un **problème** d'accord d'accord donc j'ai j'ai bien noté que vous souhaitiez une **réponse** par écrit. »

Ces réclamations vont porter à la fois sur des problèmes techniques et sur des problèmes relationnels. Concernant les aspects techniques, on trouve des mots comme : « technicien », « technique », « raccordement », « disjoncteur », « énergie », « chauffage », « hiver », « index », « chantier », « câble », etc. Ces extraits, concernant l'installation ou le raccordement d'un nouveau compteur, peuvent donner lieu à des passages assez longs.

Agent : « ... elle est divisé en 2 parties l'installation donc le réseau edf qui va jusqu'à la partie haute du **disjoncteur** donc tout ce qui est **compteur** et **disjoncteur** c est edf par contre... »

Concernant les aspects relationnels, on trouve des mots comme : « courrier », « fournisseur », « mail », « réponse », « client », « rendez-vous », « montant », etc.

Client : « ... donc premièrement on a on a on a insisté d'avoir un **rendez** de **rendez-vous** téléphonique premièrement... mais quand même bon alors on a attendu 2 semaines pour ce **rendez-vous** téléphonique le monsieur il m'a appelé... »

Client : « ... ça suffit je vais faire une **réclamation** concernant le le **rendez-vous** qui n'a pas écoulez je n'habite pas sur place ... »

Agent : « ... en attendant la **réponse** de notre service national consommateurs... »

7 Conclusion

Comme les Centres d'Appels représentent un maillon important de la Gestion de la Relation Client pour EDF, l'amélioration continue de leur performance est un enjeu majeur. Cette amélioration passant par des analyses qualitatives, seul un faible pourcentage peut être retenu pour écoute, d'où l'importance des critères de sélection. En s'appuyant sur un corpus de conversations enregistrées dans un Centre d'Appels EDF et transcrites automatiquement, nous avons montré une relation entre la longueur des tours de parole et leur contenu, à travers le vocabulaire qui les compose et les sentiments qui y sont véhiculés. Par conséquent, la longueur des tours de parole est un critère utile de sélection de conversations. Il peut s'ajouter de façon complémentaire aux autres stratégies de sélection que sont les mots clés (nom d'offre, entreprise concurrente, etc.), les thématiques, les sentiments, etc.

Références

- ADDA, G., CAILLIAU, F., DAQUO, A-L, GARNIER-RIZET, M., GUILLEMIN-LANNE, S., SUIGNARD, P. et WAAST-RICHARD, C. (2011). La transcription automatique et la fouille de données conversationnelles pour l'analyse de la relation client. In M. Campedel et P. Hoogstoël (Ed.), *Sémantique et multimodalité en analyse de l'information*. Hermes Lavoisier, Paris.
- BOZZI, L., SUIGNARD, P., WAAST-RICHARD, C. (2009). Segmentation et classification non supervisée de conversations téléphoniques automatiquement retranscrites. In *Actes de TALN*, Senlis.
- CAILLIAU, F., et CAVET, A. (2010). Analyse des sentiments et transcription automatique : modélisation du déroulement de conversations téléphonique. In *TAL*, 51-3, ATALA.
- DANESI, C. et CLAVEL, C. (2010). Impact of spontaneous speech features on business concept detection in call centers: a study of call-center data. In *Proc. of SSCS '10*, ACM, New York, NY, USA.
- DEVILLERS, L., VAUDABLE, C et CHASTAGNOL, C. (2010). Real-life emotion-related states detection in call centers: a cross-corpora study. In *INTERSPEECH-2010*, pages 2350-2353.
- GARNIER-RIZET, M., ADDA, G., CAILLIAU, F., GUILLEMIN-LANNE, G. et WAAST-RICHARD, C. (2008). CallSurf - Automatic transcription, indexing and structuration of call center conversational speech for knowledge extraction and query by content. In *Actes de LREC 2008*. Marrakech.
- LABBE, C. et LABBE, D. (2003). La distance intertextuelle. In *Corpus*, décembre 2003, <http://corpus.revues.org/index31.html> [consulté le 20/12/2011].
- LABBE, C. et LABBE, D. (2010). Ce que disent leurs phrases. In *Proceedings of 10th International Conference Statistical Analysis of Textual Data*, Rome, Italie.
- LAFOREST, M. (2011). Trois petits tours et puis s'en vont ou qu'est-ce qu'un tour de parole ? In *Langues et linguistique*, numéro spécial Journées de linguistique, pages 34-42.
- SACKS, H., SCHEGLOFF, E.A. et JEFFERSON, G. (1974). A simplest systematics for the organization of turn-taking for conversation. In *Language*, 50 (4), pages 696-735.
- VINCENT, D. (2002). Les enjeux de l'analyse conversationnelle et les enjeux de la conversation. In *Revue québécoise de linguistique*, 30-1, pages 177-198.