Web Chat Conversations from Contact Centers: a Descriptive Study

Géraldine Damnati, Aleksandra Guerraz, Delphine Charlet
Orange Labs
2 avenue Pierre Marzin, 22307 Lannion Cedex, France
E-mail: {geraldine.damnati,aleksandra.guerraz,delphine.charlet}@orange.com

Abstract

In this article we propose a descriptive study of a chat conversations corpus from an assistance contact center. Conversations are described from several view points, including interaction analysis, language deviation analysis and typographic expressivity marks analysis. We provide in particular a detailed analysis of language deviations that are encountered in our corpus of 230 conversations, corresponding to 6879 messages and 76839 words. These deviations may be challenging for further syntactic and semantic parsing. Analysis is performed with a distinction between Customer messages and Agent messages. On the overall only 4% of the observed words are misspelled but 26% of the messages contain at least one erroneous word (rising to 40% when focused on Customer messages). Transcriptions of telephone conversations from an assistance call center are also studied, allowing comparisons between these two interaction modes to be drawn. The study reveals significant differences in terms of conversation flow, with an increased efficiency for chat conversations in spite of longer temporal span.

Keywords: chat conversations, contact center, interaction, language deviation, typographic expressivity marks

1. Introduction

Even if phone remains the predominant contact channel in companies' interactions with their customers, interaction modes are much more diverse nowadays and Customer Relationship Management has to integrate multi-canality in their analytics measurements. In particular, chat conversations are growing very fast and are appreciated by both customers and agents who consider it as a way to remain in direct contact with customers while preserving a certain distance and avoiding disrespectful exchanges.

From a company's point of view, these conversations, easily available through massive logs, constitute a considerable wealth of information in order to better understand customers' needs. From a language processing perspective, they constitute a new area of study which has been only very little explored.

In this article we propose a descriptive analysis of such a chat conversation corpus. Some aspects are described through a contrastive analysis with a phone call-center conversation corpus.

Most studies in the literature refer to multiparty chat conversations from chatrooms. (Falaise, 2005) constituted a corpus in French. (Martel & al., 2007) and (Shaikh & al. 2010) describe similar corpora in English. (Cadilhac et al., 2013) have studied the relational structure of such conversations through a deep discursive analysis of chat sessions in an online video game. Among the few works that have been published on contact centers chat conversations, (Dickey et al., 2007) propose a study from the perspective of the strategies adopted by agents in favor of mutual comprehension, with a focus on discontinuity phenomena, trying to understand the reasons why miscomprehension can arise. (Wu et al., 2012) propose a typography of communication modes between customers and agents through a study on a conversation interface.

The media cannot be the only prism through which these conversations are studied. The domains, as well as the degree of familiarity between interlocutors are dimensions that must be taken into account. The adaptation of one's expression mode to the interlocutor is described in the literature in terms of sociolinguistic awareness. This is the reason why we propose here to specifically study customer contact chat conversations, which may share common characteristics with chatroom conversations but which have the particularity of being held in a more formal and institutional way. By choosing the domain angle, we can also propose some comparisons with telephone conversations collected from similar domain call-centers. Telephone conversations have been studied with more advanced information extraction experiments on French call-center data from EDF (Garnier-Rizet et al., 2008) or RATP (Béchet et al., 2012), with the main difficulty being the need to handle spontaneous speech phenomena (intrinsic phenomena or noise induced by automatic speech transcription).

The descriptive analysis presented in this paper will ground research work that will be carried out in the DATCHA project[1].

The paper is organized as follows. First, we describe the chat corpus in Section 2. Section 3 provides a comparative analysis of interactions from chat and telephone. Then we analyze the characteristics of this written corpus through the analysis of misspelling errors in Section 4 and the analysis of typographic expressivity marks in Section 5.

2. Chat corpus description

The corpus has been collected from Orange online assistance for Orange TV customers who contact the assistance for technical problems or information on their offers. In certain cases, the conversation is developed linearly as in the example in Figure 1. In other cases, the agent can conduct distant tests on the line, or the customer is asked to perform manipulation on his installation (unplug, reconnect, reboot,…) which also induce

[1] Funded by the French ANR, from 2015 to 2019, aiming at performing knowledge extraction from large chat corpora.

latencies in the conversation. In all cases, the corpus has the following form where timestamps are given at the beginning of each message, corresponding to the moment when the agent or the customer presses the enter key, and thus to the moment when their message becomes visible on the other participant's screen.

```
[12:04:20] Vous êtes en relation avec _AGENT_.
[12:04:29] _AGENT_: Bonjour, je suis _AGENT_, que puis-je
pour vous ?
[12:05:05] _CUST_: mes enfant ont perdu la carte dans le
modem et je nai plus de tele comment dois je faire?
[12:05:27] _AGENT_: Pouvez vous me confirmer votre numéro
ligne fixe afin que je sois sûr d'avoir le bon dossier ?
[12:05:56] _CUST_: _NUMTEL_
[12:07:04] _AGENT_: Si je comprend bien vous avez perdu
la carte d'accès de votre décodeur.
[12:07:27] _CUST_: oui ces bien sa
[12:07:47] _CUST_: code erreure S03
[12:09:09] _AGENT_: Pas de souci, je vais vous envoyer une
autre carte par voie postale à votre domicile.
[12:09:38] _CUST_: est ce que je peux venir chez orange
la chercher aujourdui
[12:10:36] _AGENT_: Vous ne pouvez pas récupérer une carte
depuis une boutique Orange puisque vous n'avez pas une.
[12:11:02] _AGENT_: Car dans une boutique Orange, ils
peuvent seulement faire un échange.
[12:11:33] _CUST_: ok merci de me lenvoyer au plus vite
vous avez bien mes coordonnée
[12:11:57] _AGENT_: Oui je les bien sur votre dossier.
[12:12:51] _CUST_: ok tres bien dici 48h au plus tard 72h
pour la carte
[12:14:06] _AGENT_: Vous la recevrez selon les délais
postaux à l'adresse figurant sur votre dossier (entre 3
et 5 jours).
[12:14:25] _CUST_: ok tres bien en vous remerciant a
bientot
[12:15:20] _AGENT_: Je vous en prie.
[12:15:29] _AGENT_: Avant de nous quitter avez-vous
d'autres questions ?
[12:17:23] _CUST_: non merci
```

FIGURE 1: EXTRACT FROM A CHAT CONVERSATION

Data have been anonymized: the names of customers and agents have been replaced by a single symbol (_CUST_ and _AGENT_ respectively) as well as the phone numbers, contract references, addresses and email addresses. Hence, the lexical analysis excludes any personal data.

3. Interaction analysis

In this section we propose to analyze the temporal course of conversations and to present the results in a contrastive way, by comparing a chat corpus and a call center phone corpus in the same domain. Telephone conversations have been collected from a technical assistance call center on a similar perimeter, with similar properties (linear conversations or latencies due to manipulations from both sides). If chat data are available in large quantity, as they are directly saved in the system's logs, it is not the case for telephone data that have to be processed by a long and costly manual annotation process.

Due to the limitation in terms of available phone data, we have chosen to select an equivalent number of words in both corpora.

The chat corpus contains 230 conversations for a total amount of 6879 messages and 76839 words and the telephone corpus is composed of 56 conversations for a total of 6870 breath groups and 76463 words.

Manual transcription has been performed with Transcriber (Barras et al., 2001) which allows temporal information to be inserted. Synchronization points have been inserted when perceptible pauses were made by speakers, allowing breath groups to be extracted. For the sake of comparison, we consider that a *message* in chats corresponds to a *breath group* on the phone. Several consecutive messages by the same writer constitute a *keyboard turn* while consecutive breath groups constitute a *speech turn*. The main difference is that in chats, the writer himself decides to segment in messages (by voluntarily pressing the Enter key) while for phone conversations, it is the transcriptor who places the segmentation marks. With all these reservations, we will keep this similarity between messages and breath groups (BG) in the rest of the paper.

In Table 1, we propose an analysis of interactions along their duration, their lengths in turns and messages/BG. Messages/GS are divided into two categories:

– Beginning of turn (B): the preceding message/BG comes from another writer/speaker.
– Inside a turn (I): the preceding message/GS comes from the same writer/speaker

Several observations can be drawn:

- Chat conversations are twice as long in duration as phone conversations. For phone conversations, the speech ratio over the total duration is 64.5%, however it is not possible to establish an equivalent ratio for chat as the "inactivity" time is not directly logged. Besides, chat conversation present a larger diversity with a standard deviation of 901s on duration against 306s for phone conversations.
- Phone conversations are 4 times as long in terms of turns (83.3 speech turns against 21.2 keyboard turns).
- The number of messages per turn is sensibly the same in both conditions, with the same difference between CUST and AGENT: in both cases, the number of messages per turn is more important for agents than for customers.

		Chat (230 conversations)			Telephone (56 conversations)		
		Total	**CUST**	**AGENT**	**Total**	**CUST**	**AGENT**
Av. duration (sec)		1185.7	549.3	636.4	594.5	162.0	221.8
#turns by conversations		21.2	10.3	10.9	83.3	41.5	41.8
#messages by turn		1.41	1.27	1.54	1.47	1.33	1.62
#words per message	all	11.2	8.6	13.2	11.1	10.0	12.1
	B-	11.1	8.7	13.3	9.6	8.4	10.8
	I-	11.4	8.1	13.0	14.3	14.9	14.1

TABLE 1: INTERACTION ANALYSIS

- If the number of words per message is globally comparable between chat and phone, we can observe however a difference between CUST and AGENT. In chats, this number is significantly higher for AGENT than for CUST. This difference is reduced on the phone. This can be explained by the fact that agents have access to sentence libraries that they can automatically include in the dialogue. This is particularly the case when they have to provide long and detailed instructions to solve a problem.
- In chats, we observe no difference between the number of words by B- messages and the number of words by I- messages. However, I- messages are significantly longer in the phone corpus.

If the first two observations can seem paradoxical, it can be explained in several ways. First we can argue that it is faster to speak than to write, especially for customers which are not necessarily familiar with a keyboard. Then, we must take into account the fact that chat conversation is not an exclusive activity, while it is more difficult to follow a side activity during a phone conversation. Chat conversations can be conducted simultaneously with another activity (AGENT can have two conversations in parallel and CUST can have another activity) and it is tolerated to wait for a laps of time before receiving or sending a message. Regarding the last observation, we can formulate the hypothesis that a speaker doesn't provide too much information at the beginning of a turn and develop its speech progressively while the information remanence on the screen can lead the writer to directly type a complete and detailed message, even if it means that the other participant will read it several times to understand everything.

4. Analysis of language deviation

For this study a corpus of 276 chat conversations has been manually corrected by a single annotator. It contains 8455 messages and 94244 words. The annotator was advised to correct misspelled words but it was not allowed to *modify* the intent of a message (adding a missing word or suppressing an irrelevant word). In order to compare the original chat conversations with the corrected ones, punctuation, apostrophe and case have been normalized. The manually corrected messages have then been aligned with the original messages thanks to an automatic alignment tools using the classical Levenshtein distance, with all types of errors having the same weight. A post-processing step was added after applying the alignment tool, in order to detect agglutinations or split.

An *agglutination* is detected when a deletion follows a substitution (`[en->entrain] [train->]`) becomes (`[en train->entrain]`). Conversely, a *split* is detected when an insertion follows a substitution (`[télécommande ->télé] [->commande]`) becomes (`[télécommande ->télé commande]`). Instead of being counted as two errors, agglutinations and splits are counted as one substitution.

We examine SER (Sentence Error Rate which corresponds here to a Message Error Rare) and WER (Word Error Rate); measures frequently used for machine translation and speech recognition system evaluation. The percentage of messages with at least one error is 26.3%. The SER is significatively higher for CUST (39.78%) than for AGENT (15.38%). This can be explained by the fact that agents have access to libraries of sentences and by the fact that they have professional skills. Futhermore, as can be seen from Figure 2, a large proportion of correct messages contain only one word, for example (`oui, merci`).

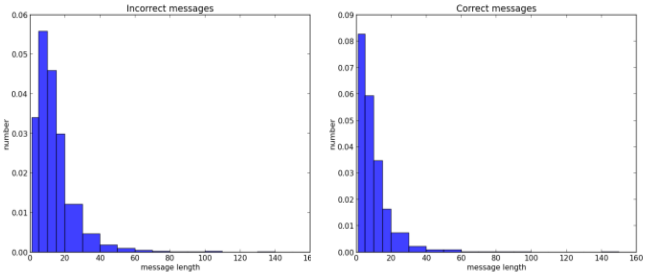

FIGURE 2 MESSAGE LENGTH DISTRIBUTION FOR INCORRECT (LEFT) AND CORRECT (RIGHT) MESSAGES

The WER for all conversations is low (4.3%). Again, CUST messages present a higher WER (about 10%) than AGENT messages (only 1.6%). SER and WER have been calculated for each conversation; the repartition of SER and WER is shown in Figure 3. In terms of orthographic correctness the difference between AGENT and CUST messages is significant. On the left side, we can see that almost three quarters of the conversations have an SER lower than 20% for AGENT where as less than a quarter of conversations have an SER lower than 20% for CUST. On the right side, we can observe that almost 50% of AGENT messages are entirely correct (from 0% to 1% WER), and only 8% of CUST messages have less than 1% misspelled words.

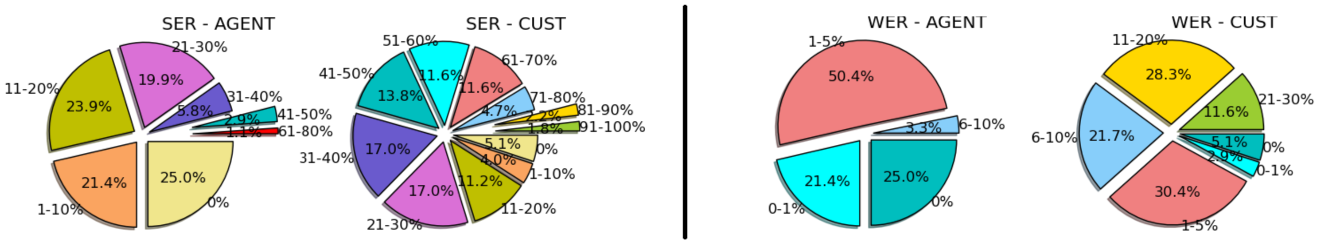

FIGURE 3: SER AND WER DISTRIBUTION

We automatically identified substitution error types in both CUST and AGENT messages.

- Most frequent substitution errors (40%) are *diacritics* errors that are accent differences between the written words and the corrected words `(à ->a, très ->trés, énergie ->énérgie)`.
- 15% of errors are *inflection* errors `(comprends ->comprend, question ->questions, ai ->ais)` among which 2.91% of errors concern specific inflection errors that are very common in French (Véronis and Guimier de Neef, 2006) i.e. past participles replaced by infinitifs for verbs that end with "er" `(essayé ->essayer, changé ->changer)`.These inflection errors are numerous in our corpus and may have a significant impact on syntactic parsing.
- 8.4% are errors due to deletion or "wrong" insertion of an *apostrophe* `(qu' ->qu, est ->est)`.
- *Agglutination* errors constitue 8.2% `(en train ->entrain)`.
- 2.6% of errors are words that were *split* `(livebox ->live box, savoir ->sa voir)`.
- Finally, typographic errors represent 18% of substitution errors. The most frequent case is when one letter is deleted (7.2%), then one letter is substituted (5.4%) and the third most frequent is when one letter is added (3.9%). The remaining typographic errors involve more than one letter.

Texting abbreviations are rare in our corpus and are exclusively used by CUST. This may be explained by the formal dimension of the conversation.

5. Expressivity marks

In this section we present a comprehensive analysis of expressivity marks related to typography: case changes, exclamation marks, suspension marks and smileys.

5.1 Case analysis

In this corpus, agents write with lowercase. When CUST starts to write in uppercase, the agent asks him to switch to lowercase. In only 2 conversations out of 230 the customer exclusively writes in uppercase. In general, the use of uppercase is whether reserved for a certain category of words or for specific purposes. Words that are regularly written in uppercase are technical words (INIT, RESET, BOOT, PLUG,..) acronyms (USB, ADSL, HD, HDMI, TNT, WIFI…), error codes and OK, SVP (for please). Besides that, 77 messages from 33 different conversations contain a case change with a particular intention. 75 are produced by CUST and 2 are produced by AGENT.

The use of uppercase marks the insistence on a particular term:

```
non par PLUG, mais ils fontionnent TRES BIEN
oui, RIEN n'a changé entre avant et aujourd'hui !
mon modele de decodeur est le samsoung shd 85 INTROUVABLE
SUR VOTRE SITE
il nécessaire de brancher le cable du décodeur TV a la live
box SI ON A LE WIFI?
C'est que le code ne permet de bloquer que certaines vidéos
et pas TOUTES LES VIDEOS
```

Some of these examples can be approximately translated as: *"yes NOTHING has changed between before and now!","The code enables to lock only some videos and not ALL THE VIDEOS"*

Some messages can be entirely in uppercase while the rest of the conversation is in lowercase.

```
MAIS POURQOUI CLIQUER SUR PLUSIEURS DECODEUR ORANGE ALORS
QUE J'EN AI QUE 1
IL EST EVIDENT QUE NOUS AVONS SUIVI A LA LETTRE VOS INFOS
```

Switching from lowercase to uppercase translates disappointment or customer misunderstanding.

5.2 Punctuation marks

Exclamation marks

61 messages from 36 different conversations contain at least one "!", among which 41 contain a single "!" while contain duplicated "!" (from 2 to 6 consecutive marks). Only one message was produced by an agent. Most of these messages have a negative polarity. `(« oui j'ai compris faut attendre!!!!! jespere pas trop longtemps!!!!!» [`*`I've understood, I have to wait!!!! I hope not so long!!!!!`*`], « et pourquoi ? je n'ai rien demandé !» [`*`why? I didn't ask anything!`*`])` but some can have a positive polarity when associated to positive words `(« oui merci! » [`*`yes thank you!`*`], « Ça fonctionne ! » [`*`it works!`*`])`.

Suspension points

96 messages from 49 different conversations contain at least one sequence of "…". 68 are produced by the customer and 28 by the agent. 38 conversations contain at least one sequence of "…" from the customer.

For the agent, the use of "…" principally corresponds to an enumeration. For customers, suspension points can express a list continuation (usually in conjunction with *etc…*) but they are most of the time used as an expressivity mark `(« oui car le mien est en train de rendre lame..... » [`*`yes, because mine is dying… `*`], « j etais en relation avec votre service et ....coupure » [`*`I was in relation with your service and ....shut down`*`], « Je vais essayer de les joindre mais c'est vraiment pas facile de les avoir au tel...vous n'avez pas un autre contact? »[`*`I'm going to try to join them but it is really not easy to reach them by phone ....don't you have another contact?`*`])` which expresses weariness, disappointment or even exasperation.

Smileys :

The use of smileys in this formal interaction context is very marginal. Out of 230 conversations, we have only observed three positive smileys and no negative one. The study of a larger corpus of 5000 conversations confirmed this observation, with only 1.6% of these conversations containing a smiley with a negligible proportion of negative ones.

5.3 Synthesis

The following table synthesizes previous observations.

	Messages	_CUST_	_AGENT_	conversations
Case switch	77 (1.2%)	75 (2.5%)	2 (0.05%)	33 (14.3%)
"!"	61 (0.9%)	60 (2.0%)	1 (0.02%)	36 (15.7%)
"…"	96 (1.4%)	68 (2.3%)	28 (0.7%)	49 (21.3%)
smileys	3 (0.04%)	3 (0.1%)	0 (0%)	3 (1.3%)

TABLE 2: Typographic expressivity marks

When related to the total amount of messages, these phenomena are not very frequent, however, when related to conversations these phenomena are more significant. In fact, 44% of the conversations contain at least one expressivity mark expressed by the customer with some of them cumulating several marks (on average, a conversation containing at least one mark, contains 2 marks). Future work on success prediction or satisfaction prediction will allow to evaluate the relevance of these features in the classification process.

6. Conclusion

In this article we proposed a descriptive study of web chat conversations from contact centers. We focused our study on three dimensions. The interactional dimension was analyzed in conjunction with a phone corpus from the same applicative domain, allowing similarities and differences to be highlighted. The two other dimensions are specific to written communication. One is the amount of language deviations (misspelled words) and the other is the presence of expressivity marks through typography. This study is the first study, to the best of our knowledge, which describes such web mediated synchronous communication in a formal institutional context, namely contact centers. The corpus will ground of future research on effective information extraction methods to improve analytics tools for Customer Relationship Management.

7. References

Barras, C., Geoffrois, E., Wu, Z., & Liberman, M. (2001). Transcriber: development and use of a tool for assisting speech corpora production. *Speech Communication*, *33*(1), 5-22.

Bechet F., Maza B., Bigouroux N., Bazillon T., El-Beze M., De Mori R., Arbillot E., (2012). DECODA: a call-centre human-human spoken conversation corpus. Actes de *LREC*.

Cadilhac A., Asher N., Benamara F., & Lascarides A. (2013). Grounding Strategic Conversation: Using Negotiation Dialogues to Predict Trades in a Win-Lose Game. Actes de *EMNLP*.

Dickey M., Burnett G., Chudoba K., & Kazmer, M. (2007). Do you read me? Perspective making and perspective taking in chat communities. *Journal of the Association for Information Systems*, *8*(1), 3.

Falaise A. (2005). Constitution d'un corpus de français tchaté. Actes de *RECITAL*.

Garnier-Rizet M., Adda G., Cailliau F., Gauvain J. L., Guillemin-Lanne S., Lamel L., ... & Waast-Richard, C. (2008). CallSurf: Automatic Transcription, Indexing and Structuration of Call Center Conversational Speech for Knowledge Extraction and Query by Content. Actes de *LREC*.

Martell E., Forsyth N., & Craig H. (2007). Lexical and Discourse Analysis of Online Chat Dialog. Actes de *International Conference on Semantic Computing (ICSC)*.

Shaikh, S., Strzalkowski, T., Broadwell, G. A., Stromer-Galley, J., Taylor, S. M., & Webb, N. (2010). MPC: A Multi-Party Chat Corpus for Modeling Social Phenomena in Discourse. In LREC.

Véronis J. and Guimier de Neef É. (2006). Le traitement des nouvelles formes de communication écrite. In Sabah, Gérard, editor, *Compréhension automatique des langues et interaction,* pages 227– 248. Paris: Hermès Science.

Wu M., Bhowmick A., & Goldberg J. (2012). Adding structured data in unstructured web chat conversation. Actes de *ACM symposium on User interface software and technology*.