Discourse markers for topic change

Paola Herreno Castaneda and **Mathilde Dargnat**
CNRS, Université de Lorraine
ATILF
paola.herreno-castaneda@univ-lorraine.fr
mathilde.dargnat@univ-lorraine.fr

Jonathan Ginzburg
CNRS, Université Paris-Cité
Laboratoire de Linguistique Formelle
yonatan.ginzburg@u-paris.fr

1 Introduction

In conversation, new utterances can address a topic distinct from the current discourse topic (henceforth DT) (or Question Under Discussion (QUD), see Ginzburg, 2012; Roberts, 2012). This phenomenon is usually called *topic change* (henceforth TC). TC is a pivotal issue for both spoken dialogue and written text, but it is quite tricky to offer an account especially for theories derived from text analysis such as RST (Mann and Thompson, 1987; Taboada and Mann, 2006) or SDRT (Asher and Lascarides, 2003). In these frameworks, TC is not *a priori* specified as a genuine discourse relation like Cause or Contrast.[1] However, in some studies on connectives (Roze, 2009; Roze et al., 2012), it is possible to find some items associated with a Digression or a Detachment relation.

In this paper, we focus on the question of TC markers in dialogue, especially with respect to three French Discourse Markers (DMs): *à propos*, *au fait*, and *d'ailleurs*. We conduct an empirical study highlighting three of their properties: anaphoric status, position in the utterance, and the permanent/temporary nature of TC. This was the starting point for what we believe is the first formal account of French topic change DMs (TCDMs).

2 Semantic properties of French TCDMs

The literature on DMs is considerable and there is much discussion regarding which properties to take into account to define MD as a category. Nevertheless, some features are generally agreed on[2] The main functions of DMs can be categorized on the basis of three properties: discourse structuring, manifestation of the speaker's knowledge or emotions, and interaction management. In many discourse theories, DMs are associated with discourse relations which can be, for instance, of a temporal or logical type. Here, we are interested in the discourse relations that the three DMs under investigation can encode.

À propos has been characterised as an enunciative connective (Prévost, 2011; Richard and Drouet, 2016) or as a rupture(-inducing) enunciative marker (Molinier, 2003). Pretheoretically, its meaning indicates a change of DT—introducing a new topic in a way that ensures discourse continuity and avoids an abrupt topic transition.

Au fait as a DM is relatively infrequent in our data, but it seems to work in a similar way to *à propos*. It can be placed in an elaborative context, taking a previous element and elaborating on it (Defour et al., 2010). It also occurs with interrogative structures that can be used as a starting point for a new topic (D'Hondt, 2014).

D'ailleurs is often studied in the argumentation field. In the polyphony-based approach developed by Ducrot et al. (1980), it is characterized by the formula $r : P$ *d'ailleurs* Q, where r is a conclusion for which P and Q are arguments; Q is presented as a non-necessary argument for r, a sort of side comment. As a DM, it introduces a complementary and independent argument or even a digressive comment. It is also seen as possessing an additive function, where the elements it adds are detached from the DT, capable of creating a textual discontinuity or a potential TC (Baider, 2018).

3 Empirical results

In the CODIM project framework, a collection of French corpora was compiled. Here, we use spoken data from six corpora CFPP, CLAPI, ESLO, FRA80, MPF, and TCOF (for details, see the references section). We hoped to find 25 examples for each DM in each corpus, but ended up with some-

---

[1] Asher (2004) does mention in passing that some discourse markers, like *sinon*, *bon* or *au fait*, can trigger a topic shift.

[2] For recent syntheses see a.o., Anscombre et al., 2013; Brinton, 2017; Heine et al., 2021; Dargnat, 2023; Hansen and Visconti, 2024.

what less—238 tokens in total: 150 for *d'ailleurs*, 75 for *au fait*, and 13 for *à propos*.

In order to analyse the anaphoric cues, we distinguish several kinds of semantic relations between utterances that host the DM and previous ones. While explicit anaphora do of course occur, we also find other types of semantic relations such as hypernymy, hyponymy, meronymy, etc.

Regarding the TC span, the possibilities are quite diverse. Some examples show that the new topic continuation span only covers the host utterance (mostly *d'ailleurs*), while in other cases the span is much larger (*à propos* and *au fait*).

We observe that the syntactic position of the DM can lead to differences in meaning or function. When the DM appears at the beginning, its function seems to involve initiating an illocutionary act by capturing the participant's attention, followed by a short-term digression, side comment, or TC, depending on the nature of the DM. However, when they are placed at the end, their use seems to be less essential and more focused on yielding the turn or just closing an idea whose nature is parenthetical (see Koev, 2022).

4 Formalization

Our goal is to propose a formal account of French TCDMs. We follow the basic approach to dialogue coherence detailed in Ginzburg et al. (2022), in particular the characterization of responses that effect TC. Ginzburg et al. (2022) point out that for ChangeTopic the simplest analysis would involve allowing a response specific to an arbitrary question. The obvious problem this would raise is massive ambiguity since many responses from other classes would be analyzable in such terms. To avoid that, Ginzburg et al. introduce the additional restriction that an *irrelevance* relation (IrRel) (Ginzburg, 2012) should hold between the topic changing utterance and the Dialogue Game-Board, creating a lack of coherence with the current context. What would this amount to? Being neither QSpecific concerning q_1 (i.e., a partial answer to or sub-question of q_1) uttered by A to B, nor being co-propositional with a clarification question generated by q_1's utterance[3] nor QSpecific with respect to the issues ?WishDiscuss(B,q1) or

---

³Here *CoPropositionality*, for two questions means that, modulo their domain, the questions involve similar answers: for instance 'Whether Bo left', 'Who left', and 'Which student left' (assuming Bo is a student.) are all co-propositional.

λxKnowAnswer(x,q1) (implicated in metadiscursive and metaepistemic responses, respectively.).

We formulate the basic form of a TCDM in (1a). We treat such DMs as dialogue move indicators, whose force is that the next utterance will involve a new topic. We think that the force of such DMs is not abrupt—here we allow for arbitrary non-related issues, but this might be too permissive —we leave a more empirically based decision to future work. Nonetheless, at least in the case of permanent topic change markers like *au fait* and *a propos*, a downdate is required of the previous MaxQUD, which the speaker feels is exhausted; this is not the case for *d'ailleurs*, so the force it signals will be somewhat different. We assume for now that such DMs are compatible with various kinds of moves (assertions, queries, commands), but we could make the specification more restricted if the data suggested the need. The specification that the DM's complement is verbal and I(ndependent)C(lause):+ means such DMs select for matrix clauses. By specifying information about the LatestMove, we capture the apparent generalization that exophoric triggers are incompatible with such DMs, as shown by our data. The force of such an utterance is $ChangeTopic$, explicated in (1b), with two subcases (i) permanent topic change (applicable to *à propos* and *au fait*) and (ii) temporary topic change (applicable to *d'ailleurs*). In both cases an utterance concerning a question irrelevant to the previous dialogue gameboard is licensed. In case (i) the previous MaxQUD is downdated, whereas in case (ii) both the new and the old issues are maintained as maximal in QUD.

(1) a.
$$\begin{bmatrix} \text{arg-struc} : \left\langle \begin{bmatrix} \text{cat} : \text{v} \\ \text{IC} : + \\ \text{cont} = \text{p1} : \text{IllocProp} \end{bmatrix} \right\rangle \\ \\ \text{dgb-params} : \begin{bmatrix} \text{spkr} : \text{Ind} \\ \text{addr} : \text{Ind} \\ \text{t} : \text{TIME} \\ \text{c0} : \text{addressing(spkr,addr,t)} \\ \text{LatestMove.cont} : \text{IllocProp} \\ \text{MaxQUD=q} : \text{Question} \end{bmatrix} \\ \\ \text{cont} = \text{ChangeTopic(spkr,p1,q)} : \text{IllocProp} \end{bmatrix}$$

b. **Permanent/Temporary ChangeTopic**
$$\begin{bmatrix} \text{pre} : \begin{bmatrix} \text{QUD} = \left\langle \text{q1, Q} \right\rangle : \text{poset(Question)} \end{bmatrix} \\ \\ \text{effects} : \begin{bmatrix} \text{spkr} = \text{pre.addr} : \text{Ind} \\ \text{addr} = \text{pre.spkr} : \text{Ind} \\ \text{r} : \text{Question} \vee \text{Prop} \\ \text{q2} : \text{Question} \\ \text{R} : \text{IllocRel} \\ \text{Moves} = \left\langle \text{R(spkr,addr,r)} \right\rangle \oplus \\ \text{pre.Moves} : \text{list(LocProp)} \\ \text{c1} : \text{Qspecific(R(spkr,addr,r),q2)} \\ \text{(i)} \left\langle \text{q2, Q} \right\rangle \\ \text{QUD} = \\ \text{(ii)} \left\langle \begin{matrix} \text{Max} = \left\{ \text{q2,q1} \right\}, \\ \text{Q} \end{matrix} \right\rangle : \\ \text{poset(Question)} \\ \text{c2} : \text{IrRel(q2,pre)} \end{bmatrix} \end{bmatrix}$$

Corpora

CFPP: Corpus de Français Parlé Parisien. http://cfpp2000.univ-paris3.fr

CLAPI: Corpus de Langue Parlée en Interaction. http://clapi.ish-lyon.cnrs.fr

CRFP: Corpus de Référence du Français Parlé. http://www.up.univ-mrs.fr/delic/corpus/index.html

ESLO: Enquêtes Sociolinguistiques à Orléans. http://eslo.huma-num.fr

FRA80: Corpus de Français des Années 80. CREDIF, ENS de Saint-Cloud.

MPF: Multicultural Paris French. https://www.ortolang.fr/market/corpora/mpf

Scientext: https://lidilem.univ-grenoble-alpes.fr/ressources/corpus/scientext

TCOF: Traitement de Corpus Oraux en Français. http://www.cnrtl.fr/corpus/tcof

Wikiconflits: see (Poudat et al., 2017)

Acknowledgments

References

Jean-Claude Anscombre, Maria Luisa Donaire, and Pierre Patrick Haillet. 2013. *Opérateurs discursifs du français*. Peter Lang Verlag, Lausanne, Suisse.

Nicholas Asher. 2004. Discourse topic. *Theoretical Linguistics*, 2-3(30):163–201.

Nicholas Asher and Alex Lascarides. 2003. *Logics of Conversation*. Cambridge University Press, Cambridge.

Fabienne Baider. 2018. *Par ailleurs* et *d'ailleurs*: marqueurs linguistiques de « rupture » textuelle ou marqueurs de continuation argumentative ? In *6e Congrès Mondial de Linguistique Française*, volume 46, pages 01–14. EDP Sciences.

Laurel J. Brinton. 2017. *The Evolution of Pragmatic Markers in English: Pathways of Change*. Cambridge University Press.

Mathilde Dargnat. 2023. *Lexique et discours*, synthèse d'HDR. Université Paris 8 Vincennes Saint-Denis.

Tine Defour, Ulrique D'Hondt, Anne-Marie Simon-Vandenbergen, and Dominique Willems. 2010. In fact, en fait, de fait, au fait: A contrastive study of the synchronic correspondences and diachronic development of english and french cognates. *Neuphilologische Mitteilungen*, pages 433–463.

Ulrique D'Hondt. 2014. Au fait, de fait et en fait: analyse de trois parcours de grammaticalisation. *Revue Romane. Langue et littérature. International Journal of Romance Languages and Literatures*, 49(2):235–263.

Oswald Ducrot, Danière Bourcier, Sylvie Bruxelles, and [...]. 1980. *Les mots du discours*. Editions de Minuit.

Jonathan Ginzburg. 2012. *The Interactive Stance: Meaning for Conversation*. Oxford University Press, Oxford.

Jonathan Ginzburg, Zulipiye Yusupujiang, Chuyuan Li, Kexin Ren, Aleksandra Kucharska, and Pawel Lupkowski. 2022. Characterizing the response space of questions: data and theory. *Dialogue & Discourse*, 13(2):79–132.

Maj-Britt M. Hansen and Jacqueline Visconti, editors. 2024. *Manual of Discourse Markers in Romance*. De Gruyter, Berlin.

Bernd Heine, Gunther Kaltenböck, Tania Kuteva, and Haiping Long. 2021. *The Rise of Discourse Markers*. Cambridge University Press.

T. Koev. 2022. *Parenthetical Meaning*. Oxford studies in semantics and pragmatics. Oxford University Press.

William C Mann and Sandra A Thompson. 1987. *Rhetorical structure theory: A theory of text organization*. University of Southern California, Information Sciences Institute Los Angeles.

Christian Molinier. 2003. Connecteurs et marqueurs énonciatifs: Les compléments figés formés à partir du nom *propos*. *Lingvisticae Investigationes*, 26(1):15–31.

Céline Poudat, Natalia Grabar, Camille Paoloque-Berges, Thierry Chanier, and Jin Kun. 2017. Wikiconflits : un corpus de discussions éditoriales conflictuelles du wikipédia francophone. In Ciara R. Wigham and Gudrun Ledegen, editors, *Corpus de communication médiée par les réseaux, Construction, structuration, analyse*, pages 19–36. L'Harmattan.

Sophie Prévost. 2011. *A propos* from verbal complement to discourse marker: a case of grammaticalization? *Linguistics*, 49(2):391–413.

Élisabeth Richard and Griselda Drouet. 2016. Confirmer pour mieux détourner : marqueurs d'acceptation et modalités de transition. *Testi e linguaggi*, pages 131–141.

Craige Roberts. 2012. Information structure in discourse: Towards an integrated formal theory of pragmatics. *Semantics and Pragmatics*, 5(6):1–69.

Charlotte Roze. 2009. Base lexicale de connecteurs du français.

Charlotte Roze, Laurence Danlos, and Philippe Muller. 2012. Lexconn: A french lexicon of discourse connectives. *Discours*, 10:01–13.

Maite Taboada and William C. Mann. 2006. Rhetorical structure theory: looking back and moving ahead. *Discourse Studies*, 8(3):423–459.