

Contribution atelier IC2011

Traces, traces numériques, connaissances et cognition

Essai d'approche transversale

Pierre Deransart

INRIA Rocquencourt, BP 105, 78153 Le Chesnay Cedex, France
Pierre.Deransart@inria.fr

Les traces sont au cœur de disciplines aussi variées que la philosophie, la médecine, l'art, la littérature, la sociologie, l'histoire, la biologie, la physique, l'astronomie, la pédagogie, l'informatique et bien d'autres. Les développements récents des sciences et techniques de l'information et de la communication (STIC) s'accompagnent de la numérisation de toutes sortes de données dont une grande partie revêt la forme de traces numériques. Des bases de données de telles traces se constituent et de nouvelles méthodes de traitements et de nouvelles applications apparaissent, créant des champs d'activités en pleine expansion. Par exemple la numérisation de l'image ou de la musique sous forme de traces numériques ouvre des champs de manipulations et transformations inimaginables auparavant [11]. L'élaboration de connaissances passe bien souvent par une première collecte de traces.

Au début il y a une notion de trace qu'il faut essayer de préciser. En informatique l'objet "trace" est relativement bien défini puisqu'il s'agit souvent d'objets construits volontairement sous ce nom. Mais même si, dans ce domaine, on peut en traiter assez formellement, la dénomination peut recouvrir différentes notions pratiques. La première question que l'on veut aborder ici concerne la transversalité de l'idée de trace : dans quelle mesure l'objet trace auquel on s'intéresse ici recouvre des notions communes à différents domaines d'activité ou de recherche.

1 Une ontologie pour les traces

Alexandre Serres [10], traitant de "la question des traces et des corpus dans les recherches en sciences de l'information et de la communication" dénombre pas moins de 6 déclinaisons possibles du concept de trace (avec des intersections non nulles).

1. Trace comme empreinte, marque psychique (mémoire, empreinte, vérité, Paul Ricoeur [12]) traces affective, corticale, écrite (extériorité)
2. Trace comme indice, détail (paradigme judiciaire, Carlos Ginzburg [6]) art, littérature policière (Conan Doyle), psychanalyse (signes, indices, symptômes)
3. Trace comme mémoire [de forme], trace du passé, document (Marc Bloch [2])
4. Trace comme ligne d'écriture (Jacques Derrida [4], Sylvain Auroux [1], grammatisation) discrétisation (écriture, langage, technologie)

- 5. Trace comme outil (scientifique) d'investigation
- 6. Trace comme exemple à suivre (morale)

Une première observation est que, à l'exception peut-être de la dernière interprétation (morale), on est peut-être pas si loin d'une notion assez générale, les différences portant moins sur la forme des objets trace, dont une sorte de syntaxe générale pourrait être reconnue, que sur leur abondance potentielle, leurs traitements associés, et la nature des connaissances recherchées.

Ceci peut se refléter dans une première tentative ontologique esquissée ici.

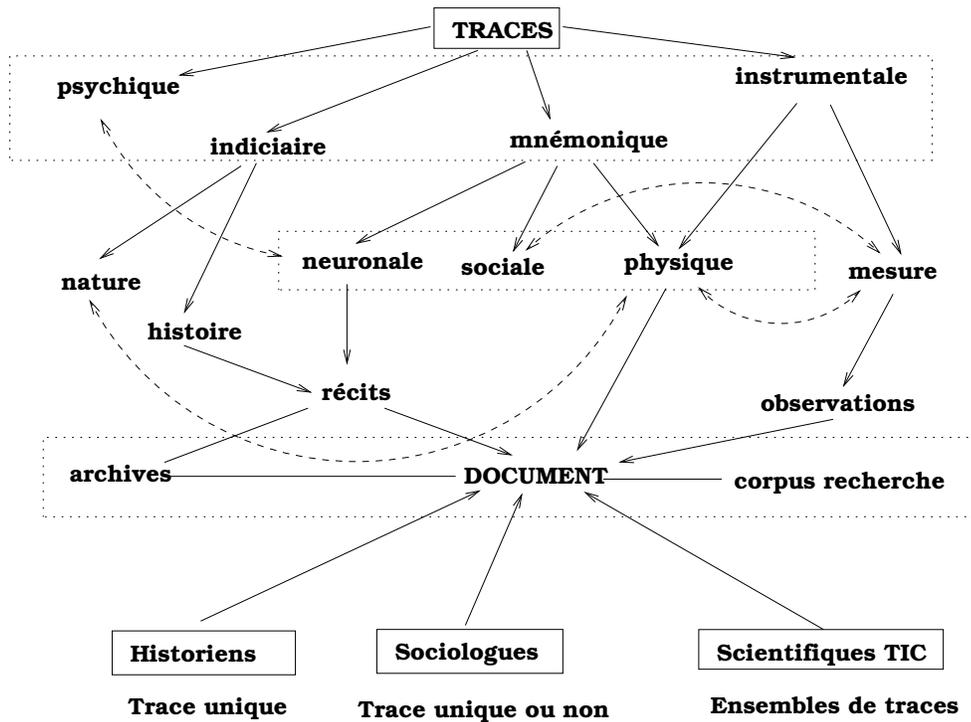


Figure 1. Un essai d'ontologie des traces

De haut en bas. A un premier niveau on trouve 4 grandes catégories correspondant à des marques psychiques, indiciaires, mnémoniques et instrumentales. Les marques mnémoniques peuvent se décliner en 3 sous-catégories : neuronales,

sociales et physiques, correspondant à trois formes de mémoires¹. Ce qui reflète approximativement les 6 catégories identifiées par A. Serres².

Les traces de type indiciaire se retrouvent dans l'exploration de la nature et dans le travail des historiens, et sont génératrices de récits que l'on peut considérer eux-mêmes comme des traces, interprétations d'autres traces. De la même manière, l'approche instrumentale produit des mesures, qui constituent souvent dans la démarche scientifique le point de départ d'observations et de découverte de nouvelles connaissances. D'une manière très générale ces processus (dont l'archivage de faits et de données) se retrouvent sous forme de documents, dont la généralisation de la numérisation accélère l'accumulation. Ces documents appartiennent à des corpus de recherche.

En bas de cette tentative de classement des traces, se trouvent quelques disciplines concernées : histoire (accumulation d'indices, de récits et d'explications), sciences humaines (échanges au sein de groupes sociaux) ou travail en environnements médiés, propre à l'observation dans la démarche scientifique. On met ici l'accent sur ce qui est peut-être une spécificité de chaque domaine, à savoir le nombre de traces examinées à la fois, au moins en première approche. Ainsi l'histoire ne se déroule qu'une fois, une expérience se reproduit difficilement en sciences humaines, enfin la démarche scientifique repose précisément et avant tout sur la reproductibilité des expériences. Il y a bien sûr d'autres manières de voir les choses : l'histoire se construit aussi sur l'examen de récits différents mais concordants, les sciences sociales produisent des expériences reproductibles et les sciences exactes utilisent l'aléa pour améliorer certaines connaissances.

Enfin les liaisons en pointillé mettent en évidence quelques intersections non vides ou interférences.

Notons que la question du document comme trace est une question centrale qui s'applique tant au document utilisé qu'au document produit. Sylvie Leuleu-Merviel [8], analysant le document numérique, le décrit comme résultat d'un processus agissant à 5 niveaux distincts. Sans rentrer dans les détails, disons que divers aspects de forme (trace, signal), de sens (interprétation) et de contexte (légitimité) doivent être considérés pour aboutir au statut de document. Ces observations sont utiles pour parvenir à relier traces et processus cognitifs avec tout ce que cela implique d'états intermédiaires de connaissances.

1. On distingue 3 formes de mémoires propres à l'homme avec différents degrés de persistance : l'ADN dont la durée est de l'ordre de celle de la vie sur terre, la mémoire neuronale dont la durée ne dépasse pas celle de l'individu, et ce que l'on peut appeler mémoire de forme (aussi exo-mémoire), propre à la technologie [7], donc aussi de type instrumentale, et dont la durée semble être de l'ordre de celle de l'homo sapiens et même au delà, en raison de son impacte planétaire.

2. En plaçant la morale dans la mémoire sociale.

2 Questions de sémantiques

Une observation qui peut faciliter une approche transversale est que, dans tous les domaines évoqués, une trace est une séquence d'événements ordonnés (temporellement ou non), sauf dans les cas d'un indice isolé ou d'une caractéristique particulière. Si la trace n'est pas ordonnée, il s'agit plutôt de "traces" (au pluriel) qui constituent un ensemble ou faisceau d'indices traités, soit comme un événement unique temporellement situé ou des événements considérés comme simultanés, soit comme un ensemble de traces qui ont chacune leur propre ordonnancement.

En première approche on pourra donc considérer qu'un objet "trace" est une séquence d'états ou configurations d'un système ou d'un processus plus ou moins bien repéré, comme une suite d'instantanés. Chaque état dans la trace sera nommé "événement" de trace. On peut s'intéresser également à des collections finies ou infinies de traces, elles-mêmes finies ou infinies³. De telles collections sont en général constitués dans le but d'obtenir de nouvelles connaissances. Une question se pose alors de relier ces événements entre eux et on désignera par "action" ce qui peut être repéré comme la cause ou l'explication de l'existence d'un nouvel événement. L'ensembles des actions est ce qui permet en quelques sortes d'"expliquer" l'ordonnancement des événements de trace.

La question qu'il faut aborder maintenant concerne la sémantique, comme manière d'aborder la question du sens ou connaissances associées aux traces.

Plutôt que de distinguer empreinte et trace (A. Mille introduction au débat), nous préférons introduire une dualité sémantique entre événement et action. Il y a en effet difficulté à distinguer les notions d'empreinte de celle de trace, l'empreinte étant en quelques sortes le résultat, la marque ou le souvenir d'une forme d'action "brute", et l'événement de trace quelque chose reconstruit après coup.

Comme en théorie des langages formels, il y a dans toute trace un aspect syntaxique et un aspect sémantique. D'une certaine manière une trace appartient à un langage. Si l'on admet qu'une trace est une suite discrète d'événements d'abord codés par une syntaxe qui sépare et identifie ses événements, les questions qui se posent ensuite sont d'ordre sémantique. Comme en linguistique formelle, on peut distinguer ce qui relève de la génération et ce qui relève de la reconnaissance. La génération c'est le processus qui a produit la trace; ce que l'on tente d'inférer, ce que l'on tente de modéliser, parfois dans le but de prévoir la suite incomplète ou interrompue (météo, analyse du comportement). La reconnaissance, c'est (re)trouver des connaissances, reconnaître dans la suite des événements des concepts avec éventuellement une précision telle que l'on puisse reproduire la trace.

A ces deux approches, on peut associer deux sémantiques énoncées sous forme de modèles implicite et descriptif [3]. Dans l'approche implicite, la trace est confondue avec le processus qui la produit (exemple automate producteur de

3. Une trace infinie correspond à la trace d'un phénomène pour lequel aucune limite temporelles n'a été ou ne peut être donnée.

mots). L'approche descriptive est extensive au sens où elle part d'un ensemble de traces donné auquel on attribue un sens (exemple automate accepteur de mots). L'exemple suggéré laisse penser qu'il peut s'agir de la même chose, mais en fait l'approche implicite tente d'observer un phénomène par ses traces alors que l'approche descriptive tente de retrouver le phénomène à partir de traces. Le choix de l'approche dépendent avant tout du phénomène étudié. On peut étudier un phénomène sans trace a priori, et il ya lieu alors de procéder à une forme d'instrumentation (exemple l'activité cérébrale d'un individu). On peut étudier un phénomène à partir d'un ensemble de traces dont on ne connaît absolument pas le modèle de production (exemple le flot de données du WEB entre deux continents). L'approche implicite est générative (elle contient potentiellement toutes les traces possibles), l'approche descriptive est reconstructive mais ne contrôle par forcément l'ensemble de traces étudié. Les deux approches visent au même but : caractériser toutes les traces possibles. La première approche détermine les actions qui produisent des événements, la seconde retrouve les actions à partir des événements. Du fait que ces sémantiques peuvent recouvrir la même chose, il y a une forme de dualité. C'est cette dualité qui parfois pose quelques difficultés conceptuelles, mais celles-ci ne sont qu'apparente.

Cette dualité en effet se retrouve dans la question récurrente que se posent les chercheurs amenés à spécifier des modèles "Événement/Condition/Action" (ECA : "on observed Event if Cond do Action") où il s'agit de déterminer (en fait choisir) ce qui est un événement et ce qui est une action [9].

Les actions codées dans une trace correspondent bien à des actions qui sont le fait d'un processus connu ou inconnu et qui ont produit les événements de trace observés. Mais chaque événement de trace peut à son tour se comporter comme action. La différence entre événement/action ne provient d'aucune qualité particulière, en tous cas ni dans une durée ni dans une forme. Ainsi par exemple l'action qui a produit l'événement de trace peut être très brève alors que sa trace est rémanente (cas de la sensation d'un long éblouissement laissée par un flash lumineux), ou au contraire l'action peut perdurer alors que sa trace est extrêmement brève (cas du flash lumineux, trace d'une explosion nucléaire observée de très loin). On voit ainsi que le même phénomène (le flash lumineux) peut être à la fois action et événement ; c'est une question de point de vue. On peut donc considérer que tout événement de trace est une action potentielle et que toute action peut produire des événements de trace. Le processus observé produit des événements de trace éventuellement sous l'action d'événements produits par d'autres processus, mais dans tous les cas un événement de trace est la conséquence d'une action.

La dualité mise en évidence par les sémantiques génératives et reconstructives, peut également s'énoncer : un événement est une action réalisée dans le passé et une action un événement pour le futur. Ceci est illustré par la figure 2 où la dualité événement/action est représentée comme une relation entre sémantiques. Un événement de trace décrit également l'action qui le produit. Il est en quelque sorte à la fois événement et action, mémoire d'une action et protagoniste d'une autre. Un événement de trace peut être aussi la cause qui va

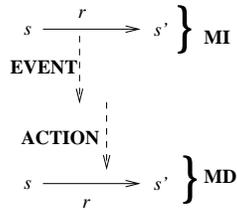


Figure 2. Dualité événement/action : question de point de vue ou de causalité. Dans le modèle implicite une action r produit un nouvel état s' (transition $s \rightarrow s'$) et constitue un événement de trace. Dans le modèle descriptif un événement de trace est interprétable comme une action sur des états (les mêmes s'agissant de la même trace). Le modèle implicite s'intéresse plutôt à la production de traces et le modèle explicite à l'exploitation de celles-ci.

provoquer de nouvelles actions : ce qui n'est au départ qu'un effet se transforme en une action qui va générer un nouvel événement etc.

3 Questionnements

Il semble qu'un concept général de trace susceptible de permettre des rapprochements puisse émerger. Il se peut aussi qu'entreprendre une telle démarche avec des visées de formalisation, puisse mener à autre chose. L'exemple de la linguistique cité par Patrick Greussay (interview dans ASTI Hebdo 2005) est à cet égard significatif. Les recherches en linguistique formelle et calculatoire, la traduction automatique, ont plus contribué à progresser dans la connaissance des langages formels que dans celle de la linguistique et la compréhension de la langue naturelle. Il se peut que l'étude des traces en informatique apporte plus dans des approches formelles de traces portées par la théorie des automates [5]) que dans l'acquisition de connaissances rattachées aux activités humaines. Il paraît tout de même intéressant de poser quelques questions.

- N'y a-t-il pas dans chaque discipline évoquée une (ou plusieurs) notions de traces. Quelles sont alors leurs spécificité? Quels sont leurs point communs?
- Dans un champ d'études travaillant à partir de traces, quelles sont les actions ou quels sont les états correspondant à des événements de trace?
- Quels apports attendre d'une approche multi- ou inter- disciplinaire de la trace, dans quelles disciplines et à quel niveau?
- L'acquisition de connaissances s'appuie sur de multiples étapes plus ou moins hiérarchisées. On retrouve souvent l'idée d'une "trace première" (obtenue par un réseau de capteurs par exemple), inexploitable comme telle mais synthétisée ou "abstraite" en une nouvelle trace analysable, à partir de laquelle de nouvelles connaissances peuvent être découvertes à l'aide d'analyses spécifiques. La qualité des connaissances obtenues dépend de la

qualité des étapes (recueil de données, transformations de traces, analyses, ...). Quelles transformations sont réalisées, comment les caractériser? Les questions de qualité sont-elles les mêmes (confiabilité, sécurité, ...)?
Liste bien sûre non exhaustive ...

Références

1. Sylvain Auroux. *La révolution technologique de la grammatisation. Introduction à l'histoire des sciences du langage*. Editions Pierre Mardaga, Liège, 1994.
2. Marc Bloch. *Apologie pour l'histoire ou métier d'historien*. Armand Colin, Paris, 1974.
3. P. Deransart. Conception de Trace et Applications (vers une méta-théorie des traces). Technical report, Inria Paris-Rocquencourt, February 2010. Working document <http://hal.inria.fr/>.
4. Jacques Derrida. *De la grammatologie*. Editions de Minuit, 1967.
5. V. Diekert and G. Rozenberg. *The Book of Traces*. World Scientific Publishing, Singapore, 1995.
6. Carlos Ginzburg. Traces, Racines d'un paradigme indiciaire. *Mythes, emblèmes, traces. Morphologie et histoire*, pages 139–180, 2004.
7. André Lebeau. *L'engrenage de la technique, essai sur une menace planétaire*. NRF, 2005.
8. Sylvie Leuleu-Merviel. Effets de la numérisation et de la mise en réseau sur le concept de document. *Information-Interaction-intelligence*, 4(1) :121–140, 2004.
9. Wolfgang May, José Júlio Alferes, and François Bry. Towards generic query, update, and event languages for the semantic web. In Hans Jürgen Ohlbach and Sebastian Schaffert, editors, *PPSWR*, volume 3208 of *Lecture Notes in Computer Science*, pages 19–33. Springer, 2004.
10. A. Serres. Quelle(s) problématique(s) de la trace? In URFISTInfo, editor, *CER-COR*, Rennes, 2002.
11. Bernard Stiegler. Désir et connaissance : le mort saisi par le vif. éléments pour une organologie de la libido. *Revue d'Intelligence Artificielle*, 19(1-2) :13–29, 2005. Colloque ARCo'04.
12. Paul Ricœur. *La mémoire, l'histoire, l'oubli*. Seuil, Paris, 2000.