



**HAL**  
open science

# Hypertexte et hypermédia: des liens entre corpus linguistiques

Jean-Marie Pierrel

► **To cite this version:**

Jean-Marie Pierrel. Hypertexte et hypermédia: des liens entre corpus linguistiques. Mokrane Bouzeghoub; Rémy Mosseri Les Big Data à découvert, CNRS Editions, pp.274-275, 2017, 978-2-271-11464-8. hal-01634420

**HAL Id: hal-01634420**

**<https://hal.science/hal-01634420>**

Submitted on 15 Nov 2017

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

## **Hypertexte et hypermédia : des liens entre corpus**

**Jean-Marie Pierrel**

L'hypertexte désigne tout contenu informatif numérique qui n'est pas assujéti à un ordre de lecture préalablement défini. Par extension, l'hypermédia possède la même caractéristique, mais il ne se limite pas à du texte (cartographie, audio, vidéo...). Tedor Holm Nelson forge le terme « *hypertext* » en 1967 en associant au préfixe « hyper », qui signifie « au-dessus, au-delà » le substantif « texte » pour exprimer cette notion de contenu informatif allant au-delà du texte. Mais l'idée même d'hypertexte, aujourd'hui l'un des fondements du Web, nous vient d'encore plus loin. En 1945, Vannevar Bush, dans son célèbre article *As We May Think (Comme nous pourrions penser)*, imagine une machine, « Memex », qui serait le prolongement de la mémoire humaine et qui enregistrerait une somme de connaissances énormes et diversifiées, avec des mécanismes d'accès totalement indépendants de toute classification hiérarchique. Sa démarche part du constat du caractère artificiel des systèmes d'organisation et d'indexation des informations scientifiques – classement purement hiérarchique et accès par une seule rubrique – alors que, selon lui, l'esprit humain fonctionne essentiellement par associations totalement improvisées.

Aujourd'hui, la masse des connaissances existantes sous forme numérique ne cesse de s'accroître et leur consultation nécessite d'accéder à des supports divers et multiples : articles, ouvrages, revues, encyclopédies, dictionnaires, etc. Cette surabondance d'informations est devenue ingérable, et le lecteur n'a plus le temps de suivre les parcours prédéfinis qu'imposait l'écrit. D'où cette recherche d'une rupture de la linéarité de la lecture au profit de parcours intertextuels ou intermédiés libres, comme généralisés sur le Web.

### **Concepts fondateurs de l'hypertexte et de l'hypermedia**

Tout système hypertexte ou hypermédia s'appuie sur trois concepts clés : les nœuds, les liens et les ancres. Les nœuds correspondent à des unités élémentaires d'information d'un ou plusieurs types : texte, graphique, image, son ou vidéo. Les liens, uni ou bi-directionnels, constituent le moyen principal d'organisation d'un ensemble de documents structuré de façon non linéaire et non hiérarchique. Ils permettent à un utilisateur de se déplacer d'un document à un autre, d'une partie d'un document à une autre partie du même document ou d'une partie d'un document à une partie d'un autre document. Enfin, les ancres repèrent dans un document des parties support des liens. Dans le cas d'un texte, ce sera une forme ou une unité sémantique ; pour des objets sonores, graphiques ou vidéo, ce seront des zones ou des régions préalablement définies comme points d'ancrage des liens.

La structure ainsi construite, reliant des ancrs à des nœuds par des liens, constitue un réseau hypertexte ou hypermédia.

### **Différents types d'hypertexte ou d'hypermédia**

D'un point de vue technique, l'implantation d'hypertextes est fortement liée aux choix de réalisation de ces notions de nœuds, de liens et d'ancres. Pour les hypertextes les plus simples, les nœuds correspondent à des documents complets (pages Web), les liens sont unidirectionnels et correspondent à des adresses Web, et les ancrs sont statiques et prédéfinies par le concepteur. Mais le modèle peut aussi se réaliser avec des liens dynamiques (calculés automatiquement) ou associatifs et/ou avec des ancrs non prédéfinies.

Aujourd'hui, la plupart des ancrs sont prédéfinies lors de la construction d'un document avec, le plus souvent dans ce cas, un lien statique. Mais il est possible aussi d'utiliser des ancrs non prédéfinies à l'avance. C'est le cas lorsque, l'utilisateur sélectionne un mot et que, grâce à un clic droit, il provoque une nouvelle recherche sur le Web ou dans un dictionnaire.

Concernant les liens, ils peuvent être de deux types : statiques ou dynamiques. Parmi les liens statiques, on distingue : *i)* des liens associatifs, intégrés à l'intérieur même d'un document ; *ii)* des liens d'annotation permettant, par exemple, d'afficher une note (texte, image, illustration, document sonore ou vidéo) dans une fenêtre spécifique de l'écran ; *iii)* des liens d'inclusion permettant, à partir d'une sorte de document squelette, d'étendre le document initial sur un point précis. Quant aux liens dynamiques, ils ne sont pas codés par le concepteur du document initial, mais calculés automatiquement lors de l'activation d'ancres le plus souvent non prédéfinies, prises en charge par un processus externe au document. La mise en œuvre de ces liens dynamiques peut se faire de différentes façons suivant la manière dont sont exploitées les ancrs. Elle peut se limiter à une recherche stricte de l'ancre (suite de caractères). Mais les dernières évolutions des hypertextes vont bien au-delà : ainsi dans le cas d'une navigation vers un dictionnaire, la recherche du mot « peuvent » nécessite préalablement une lemmatisation (analyse lexicale) de l'ancre permettant de passer de « peuvent » à « pouvoir », voire un étiquetage morphosyntaxique (association des informations grammaticales) pour passer du mot « portes » à « porter » (verbe) ou « porte » (substantif).

Aujourd'hui, on s'oriente de plus en plus vers le développement de liens dynamiques sémantiques au travers de l'exploitation des techniques du Web sémantique (cf. V.4). Le processus de résolution de liens fait alors appel à une requête sur une base de triplets RDF définissant une relation sémantique liant deux ressources ou corpus. Cette nouvelle orientation ouvre sur des modes d'organisation et de recherche d'information beaucoup plus riches que ceux offerts par des systèmes classiques d'indexation. Allant bien au-delà des mécanismes de référencement et de notes de bas de page, elle

présente de plus une qualité cruciale, sa capacité à passer à l'échelle, en s'appuyant sur des chaînes de références potentiellement quasi infinies. Dépassant largement le cadre du Web, cette approche sémantique s'impose aussi dans l'édition suivie de documents ou de ressources électroniques (corpus) que plusieurs auteurs peuvent annoter, corriger et enrichir de façon contextualisée.

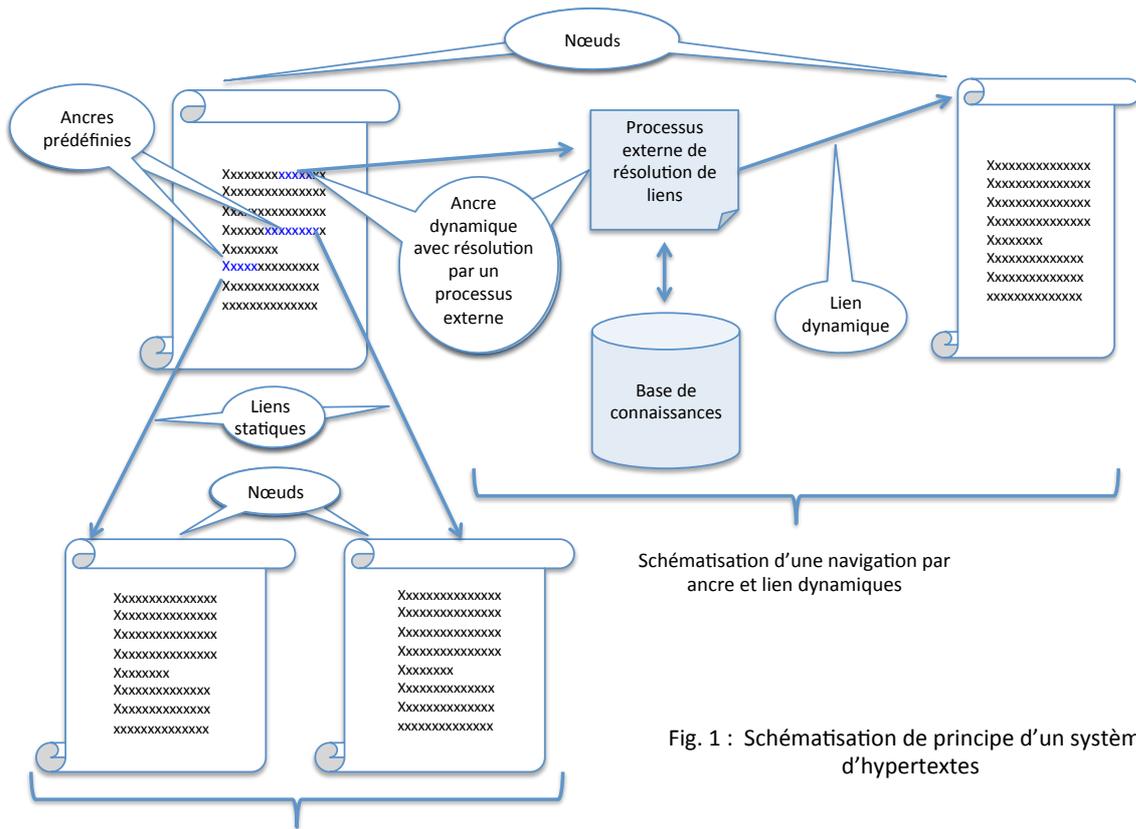
**Jean-Marie Pierrel.** Informatique linguistique, Professeur à l'Université de Lorraine, ATILF UMR CNRS 7118, Nancy.

#### **Glossaire :**

- **Étiquetage morphosyntaxique** : appelé aussi étiquetage grammatical, processus consistant à associer aux mots d'un texte les informations grammaticales telles la partie du discours (Substantif, verbe, adjectif, etc.), le genre, le nombre, etc..
- **Lemmatisation** : opération consistant à regrouper les diverses formes d'un mot sous une forme canonique appelée lemme : infinitif pour un verbe, singulier pour un nom, masculin singulier pour un adjectif.

#### **Références bibliographiques**

- R LAUFER et D SCAVETTA – *Texte, hypertexte, hypermédia*, PUF, coll. "Que sais-je ?", Paris, 1992.
- C. VANDENDORPE – *Du papyrus à l'hypertexte: Essai sur les mutations du texte et de la lecture*. Paris : La découverte, 1999.
- I SALEH – *Les hypermédias: conception et réalisation*. Hermès Science Publications, 2005.



Schématisme d'une navigation par ancre et lien dynamiques

Fig. 1 : Schématisation de principe d'un système d'hypertextes

Schématisme d'une navigation par ancre et liens statiques

■ **CORPUS**, subst. masc.

**Étymol. et Hist.** 1. Fin XII<sup>es</sup>. *corpus Deu* « hostie » (*Mort Garin*, 132 ds T.-L.); 1206 *corpus Domini* (GUIOT, *Bible*, 1223 ds GDF.) – 1584, Benedicti ds *Fr. mod.*, t. 5, 1937, p. 73; cf. encore *corpus* « id. » (1642, Oudin ds *DG* – 1771, *Trév.*); 2. 1863 *corpus juris*, *corpus* « collection du droit romain » (LITTRÉ); 3. 1890 « collection d'inscriptions de l'Antiquité » (*DG*); 4. 1961 « ensemble d'énoncés servant de base à l'analyse linguistique » (*Lar. encyclop.*). Au sens 2, empr. au lat. class. *corpus juris*, v. *corps*. Sens 3 et 4 développés en fr. à partir de 2. Le sens 1 est empr. au lat. chrét. *corpus Domini*, *corpus Christi*, désignant l'Eucharistie.

**MISE À JOUR DE LA NOTICE ÉTYMOLOGIQUE PAR LE PROGRAMME DE RECHERCHE TLF-ÉTYM :**

**Histoire :**

**B. 1. a.** *corpus iuris* subst. masc. « nom d'une collection de droit romain » (droit). Attesté depuis 1826 [in titre de chapitre et table des matières] ([MACKELDEY, Introduction Droit Romain](#), pages 151 et 202, in [GOOGLE, Recherche de Livres](#) : §. XCIV. Additions Mots latins par lesquels on c

Justinien). -

**B. 1. b.** *corpus* subst. m. Terme latin employé po

**A. 1./B. 2.** *corpus* su de leur étude scientifique ([VILLERS, État actuel](#), pag Mr. Schutz], font partie de la fois le terme *corps*\*. -

**A. 2.** *corpus* subst. masc. « ensemble de textes établi selon un critère thématique en vue de leur étude

The screenshot shows a search interface with a purple header 'Informations' and a search bar containing 'lexicographique'. Below the search bar, there are several search results. The first result is from 'MackeldeyIntroductionDroitRomain' with a snippet: 'MACKELDEY, Introduction Droit Romain = Mackeldey (Ferdinand), 1826. Introduction à l'étude du droit romain, traduite de l'allemand par L. Etienne. Revue, augmentée, précédée d'un Précis encyclopédique, et suivie d'une nouvelle restitution de la loi des XII Tables et de l'Édit perpétuel par L.-A. Warnkœnig, Mons, Leroux.' The second result is from 'LITTRÉ : corpus', with a snippet: '1863 (LITTRÉ : corpus, ... d'un certain type (notamment en lettres) réunies en vue alique, dans un contexte faisant référence à l'Allemagne] Livres : Les livres sur la Rhétorique, mis au jour par lui [le savant e Mr. Gœschen). Dans ce rapport le terme corpus côtoie une

Fig. 2 : Types d'ancres et de liens a) ancre prédéfinie « [MACKELDEY, Introduction Droit Romain](#) » et lien statique vers la fenêtre mauve d'annotation. b) ancre dynamique « lexicographique » et liens dynamiques vers Google ou le portail lexical du CNRTL. © CNRTL : CNRTL/ATILF/CNRS <http://www.cnrtl.fr/etymologie/corpus>