

Documents numériques et Documents structurés

Lydia-Mai Ho-Dac

CLLE-ERSS

Master LITL / ECIT-FH

Le rôle des documents dans notre société

(J. Lemarié, Master 2 "ECIL" 2015)

Difficultés liés au traitement des documents – le cas des docs. techniques

- Des documents techniques procéduraux omniprésents dans le travail et le quotidien
- Différentes pratiques d'utilisation d'un document procédural (Schriver 1997) :
 - Action seule (19%)
 - Action et lecture (42%)
 - Lecture, puis action (23%)
 - Action et consultation ponctuelle (17%)

Schriver, K. A. (1997). *Dynamics in document design : creating text for readers*. New York : Wiley

Documents peu, pas ou mal utilisés (e.g. Cellier 2005)

Quelques constats

- 14 000 personnes aux urgences pour non respect des consignes d'utilisation d'un pesticide en 1988 aux USA (Silver & Wogalter 1991)
- Documentations procédurales impliquées dans 45% des incidents rapportés entre 1996 et 1997
- Erreurs de maintenance proviennent de la non exécution d'une instruction (Chaparro et al. 2002)

Cellier, J. M. (2005). Caractéristiques et fonctions des textes procéduraux. Production, compréhension et usages des écrits techniques au travail, 161-180

Silver, N. C., & Wogalter, M. S. (1991, September). Strength and understanding of signal words by elementary and middle school students. In Proceedings of the Human Factors Society Annual Meeting (Vol. 35, No. 9, pp. 590-594). Sage CA : Los Angeles, CA : SAGE Publications.

Enquête auprès de 377 techniciens de maintenance (Chaparro et al. 2002)

- 64% des techniciens de maintenance déclarent réaliser la procédure à leur façon
- 78% déclarent qu'il leur arrive occasionnellement, souvent ou très souvent, de faire appel à un collègue en raison d'un manque de clarté de la procédure
- 47% sont en désaccord avec l'idée que la procédure décrit toujours la façon la plus facile de réaliser une procédure
- 30% sont d'accord avec l'assertion : « J'ai besoin d'un manuel pour comprendre le manuel »

Chaparro, A., Groff, L., Chaparro, B. and Scarlett, D. 2002. Survey of aviation technical manuals : Phase 2 report. User evaluation of maintenance documents, Washington, DC : Federal Aviation Administration, Office of Aviation Research. (Tech. Rep. No. DOT/FAA/AR-02/34)

Documents peu, pas ou mal utilisés

⇒ Manque d'efficacité de la documentation technique...

... qui se manifeste essentiellement par

- non utilisation des documents
- incompréhension de certaines instructions
- exécution incorrecte des actions et/ou volontairement omise

Des améliorations et inconvénients liées à la numérisation des documents

- + des moyens pour consulter en toutes circonstances les documents appropriés (poids, météo, etc.)
- + des outils pour assister la rédaction et la structuration de documents
 - un prolifération de documents plus ou moins officiels et légitimes
 - une grande volatilité (des formats, des outils de lecture, etc.)

...

Définition des documents (numériques)



The screenshot shows the Wikipedia page for 'Charlie Hebdo'. The main article text is partially visible, discussing the newspaper's history and its location. An inset map on the right shows the 'Charleval' district in Paris, with 'les Myrtes' park and the 'Charlie Hebdo' office marked. The map includes street names and a grid system with letters A-F and numbers 1-4.



The image shows a black clapperboard with white text and a film reel with a black and white striped clapperboard attached to it, symbolizing digital video and film production.

The table is a Gantt chart or project management tool. It features a grid with columns representing time intervals and rows representing tasks. The tasks are color-coded (green, red, blue) and have horizontal bars indicating their duration. The word 'Projet' is written in large green letters at the top left of the chart.

Définition des documents (numériques)

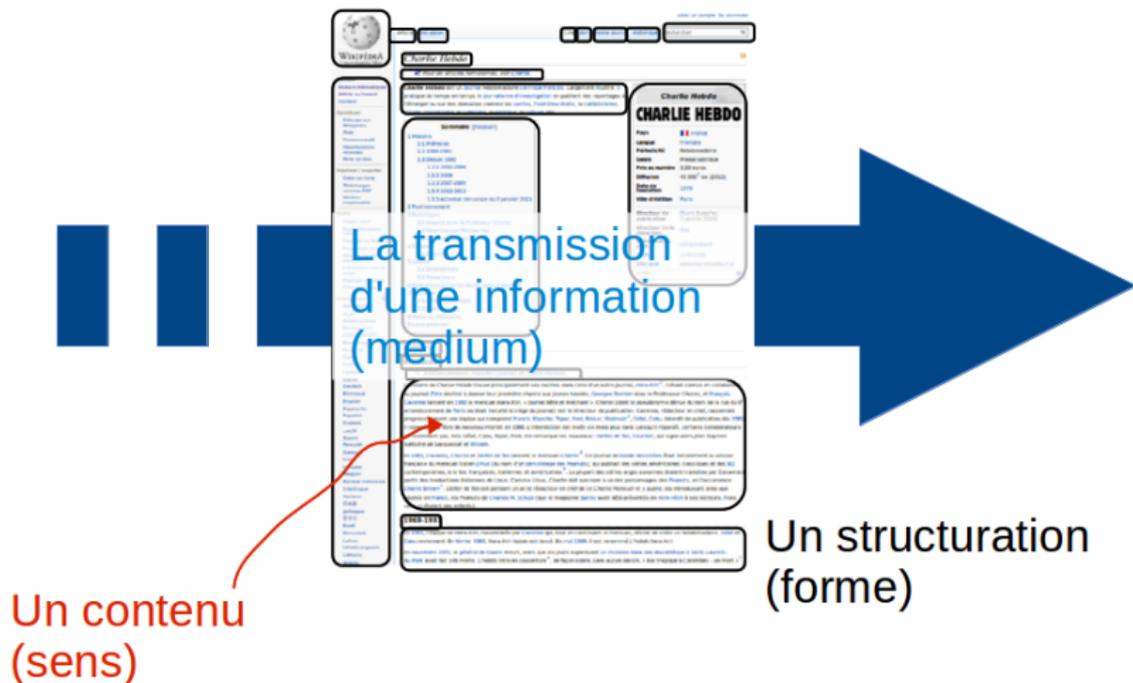
The image shows a screenshot of the French Wikipedia page for 'Charlie Hebdo'. Several elements are highlighted with red boxes to illustrate document structure:

- Navigation and Search:** The top navigation bar and search box.
- Table of Contents:** A vertical list on the left side of the page.
- Infobox:** A box on the right side containing biographical information about Charlie Hebdo, such as birth date, profession, and death.
- Text Content:** The main body of the article, with a red arrow pointing to a specific paragraph.
- References:** A list of sources at the bottom of the page.

Un contenu
(sens)

Un structuration
(forme)

Définition des documents (numériques)



Définition des documents → numériques)

3 points de vues

Pédauque Roger T. (2003) Document : forme, signe et médium, les re-formulations du numérique. <sic 00000511>

- Le document comme **signe**
- Le document comme **forme**
- Le document comme **médium**

Définition des documents → numériques)

3 points de vues

Pédauque Roger T. (2003) Document : forme, signe et médium, les re-formulations du numérique. <sic 00000511>

- Le document comme **signe**

Un document numérique est texte dont les éléments sont potentiellement analysable par un système de connaissance en vue de son exploitation par un lecteur compétent (p.16)

→ Accès au contenu (RI, EI, QA, ...)

PS : système de connaissance nécessaire depuis la numérisation

- Le document comme **forme**
- Le document comme **médium**

Définition des documents → numériques)

3 points de vues

Pédauque Roger T. (2003) Document : forme, signe et médium, les re-formulations du numérique. <sic 00000511>

- Le document comme **signe**
- Le document comme **forme**

Un document numérique est un ensemble de données organisées selon une structure stable associées à des règles de mise en forme permettant une lisibilité partagée entre son concepteur et ses lecteurs (p.10)

→ Architecture textuelle (Virbel 1989) et notions de Patrons en *document design* (Waller et al. 2012)

- Le document comme **medium**

Virbel, J. 1989. The Contribution of Linguistic Knowledge to the Interpretation of Text Structures. In J. André, V. Quint and R. Furuta (eds), *Structured Document*. Cambridge : Cambridge University Press : 181-190

Waller, R., Delin, J. and Thomas, M. (2012) Towards a Pattern Language Approach to Document Description, *Discours*, 10,

DOI : [10.4000/discours.8673](https://doi.org/10.4000/discours.8673)

Définition des documents → numériques)

3 points de vues

Pédaque Roger T. (2003) Document : forme, signe et médium, les re-formulations du numérique. <sic 0000511>

- Le document comme **signe**
- Le document comme **forme**
- Le document comme **médium**

Un document numérique est la trace de relations sociales reconstruite par les dispositifs informatique (p.25)

→ valeur des documents publiés sur internet, données personnelles, nouveaux médias (Twitter, Facebook, forums de discussion, ...)

Documents/Contenu et numérisation

- Volatilité du support et facilités de diffusion
- Conséquences :
 - Imprécision des informations essentielles (qui, quoi, pour qui, pourquoi)
 - Multiplicité des systèmes
- Nécessité de mise en place de connaissances autour du document :
Méta-données (ex. teiHeader, CLARIN)

Documents/Forme et numérisation

- Le support (la forme) devient difficilement identifiable
 - support de stockage ?
 - support logique (fichier et format) ?
- La représentation visuelle n'est plus fixe
- La structure se complexifie → grande variété de formats, de codages, d'organisation
- Nécessité d'homogénéisation des formats
 - mise en place de normes de rédaction
 - outils de diagnostic d'incohérences → *IsChecker, Safety-Data*

Diagnostiquer les consignes sécurité (V. Andréani, Safety-Data, Master "ECIL" 2014, ancienne "ECIL")

Outil *IsChecker, Safety-Data*

- Outil conçu et développé par Safety-Data depuis 2012
- Outil modulaire, adaptable à n'importe quel corpus de documents professionnels, pourvu qu'ils soient rédigés d'après une norme et qu'ils soient dans un format électronique exploitable
 - détection d'incohérence des documents p.r. à la norme définie
 - incohérence de forme
 - incohérence forme / contenu
 - zones de contenu à risque (d'interprétation)
- Corpus traités depuis 2012 : Total, Astrium, Airbus

Repérer les variations de forme (V. Andréani, Safety-Data)

Incohérence de forme dans les "pavés sécurité" (Total)

Pavé correct (texte en rouge, fond blanc, centré, puce < ! >)

SECURITE

< ! > Platinages JP N° 70/53/54

avec port du masque aut

ATTENTION : !!! PRESENCE D'H'S !!!

Ces mouvements de joints pleins doivent être effectués par un p
d'un Appareil Respiratoire de Longue Durée
(2 bouteilles sur chariot + flexible + masque) et so
surveillance d'une troisième personne disposant d'
Appareil Respiratoire Individuel (A.R.I.). Voir aussi P

Problème de couleur de fond.

SECURITE

**< ! > Platinages JP n°74/53/54 avec
port de l'A.R.I**

ATTENTION : !!! PRESENCE D'H'S !!!

Ces mouvements de joints pleins doivent être effectués par un personnel équipé d'un
Appareil Respiratoire de Longue Durée
(2 bouteilles sur chariot + flexible + masque) et sous la
surveillance d'une troisième personne disposant d'un
Appareil Respiratoire Individuel (A.R.I.). Voir aussi PGSEC78

Problème de centrage + puce

SECURITE

Les ROV d'aspiration des pompes
G1008A et G1008B restent ouvertes
(protection flushing des pompes)

Problème de couleur de police + puce.

SECURITE

***Cette opération est donc répétée au
minimum toutes les 2 minutes***

Repérer les zones à risque – incohérence forme / contenu (V. Andréani, Safety-Data)

Incohérence de mise en forme des énumérations

b) Plusieurs véhicules sont bloqués, le mécanicien :

S'assure que ce blocage n'est pas la conséquence de la fermeture d'un robinet d'arrêt de la conduite générale situé avant la partie de train bloquée :

1^{er} CAS : Aucun robinet d'arrêt CG n'est fermé :

Il actionne la commande de la valve de purge le temps suffisant pour provoquer le desserrage sur chaque véhicule bloqué.

2^e CAS : Un (ou plusieurs) robinet d'arrêt est trouvé fermé :

Il ouvre le robinet

Dans les 2 cas, le mécanicien :

- ouvre le robinet d'arrêt CG situé en arrière du dernier véhicule relié à la CG.
- vérifie le serrage des freins du dernier véhicule freiné.
- referme le robinet CG
- vérifie en se dirigeant vers la tête du train :
 - le desserrage des freins de tous les véhicules,
 - que le blocage n'a pas provoqué d'avarie aux roues.

Il applique les mesures concernant le signalement et la reprise de marche (article 385).

Repérer les zones à risque – incohérence forme / contenu (V. Andréani, Safety-Data)

Incohérence de mise en forme des énumérations

b) Plusieurs véhicules sont bloqués, le mécanicien :

S'assure que ce blocage n'est pas la conséquence de la fermeture d'un robinet d'arrêt de la conduite générale situé avant la partie de train bloquée :

1^{er} CAS : Aucun robinet d'arrêt CG n'est fermé :

Il actionne la commande de la valve de purge le temps suffisant pour provoquer le desserrage sur chaque véhicule bloqué.

2^e CAS : Un (ou plusieurs) robinet d'arrêt est trouvé fermé :

Il ouvre le robinet

Dans les 2 cas, le mécanicien :

- ouvre le robinet d'arrêt CG situé en arrière du dernier véhicule relié à la CG.
- vérifie le serrage des freins du dernier véhicule freiné.
- referme le robinet CG
- vérifie en se dirigeant vers la tête du train :
 - le desserrage des freins de tous les véhicules,
 - que le blocage n'a pas provoqué d'avarie aux roues.

Il applique les mesures concernant le signalement et la reprise de marche (article 385).

Repérer les zones de contenu à risque (V. Andréani, Safety-Data)

Zones de flou et d'ambiguïté potentielle, passages complexes

- Appliquer un film **le plus fin possible**
- La valeur correspondant à un tour le corps bielle est égale à une extension d'**environ** 3mm.
- **S'assurer** que la zone de toutes les parties effacées **ou** à **finir sont** propres **et dégraisser**.
- Fermer la vanne V456E **sauf si** la température est supérieure à 80°C **à moins que** la pression soit inférieure à 5 bars.

marqueurs de flou

phrase complexe

phrase conditionnelle

cf. [thèse de Maxime Warnier \(CLLE, sept. 2018\)](#)

Définition des documents → numériques)

3 points de vues

Pédagogue Roger T. (2003) Document : forme, signe et médium, les re-formulations du numérique. <sic 00000511>

- Le document comme **signe**
- Le document comme **forme**
- Le document comme **médium**

Un document numérique est la trace de relations sociales reconstruite par les dispositifs informatique (p.25)

→ valeur des documents publiés sur internet, données personnelles, nouveaux médias (Twitter, Facebook, forums de discussion, ...)

Documents/Medium et numérisation

- Volatilité du support et facilités de diffusion
- Conséquences :
 - Difficulté d'identifier la légitimité d'un document
 - Difficulté de tracer les évolutions d'un document
- Nécessité de mettre en place des procédures de certifications

Normes, Consignes et Certifications – exemples

Normes ISO

Chercher "documents"

Archiving of electronic data

Model document of an amendment

Textes numérisés

TEI-P5 (ex. CNRTL, ortolang)

Pages WEB

Web Content Accessibility Guidelines (WCAG) 2.0

Documentation dédiée

GNOME Documentation Style Guide V1.6

Forums de santé

Health on the Net (HON)

Cycle de vie d'un document

Acquisition	Création, capture, échange
Classement	Indexation, catalogage
Stockage	Support, archivage
Diffusion	Recherche, affichage, sélection
Valorisation	Réutilisation
Destruction	

Quel stabilité de format d'un bout à l'autre ?

Formats des documents (numériques)

Le format est le mode d'organisation des informations (structure + contenu + feuille de style) sous forme numérique, dans un fichier (ou plusieurs)

Différents formats

Cours de Daniel Flipo, Université des sciences et techno. de Lille (2006-2007) : Documents structurés et formats ouverts

- Document structuré = structure + contenu + feuille de style
- Format ouverts ou propriétaires
- Formats normalisés

Les grandes catégories de format

Format brut

texte simple (texte seul, texte nu, etc.)

Formats enrichis

- Formats d'impression : PDF, Postscript, XPS
- Traitements de texte : Word (doc, docx), OpenOffice/LibreOffice (odt), LaTeX (tex), AbiWord (abw), Frame Maker (fm), StarWriter (sxw), Word Perfect (wpd), etc.
- Formats d'échange : RTF, MIF
- Formats pour supports spécifiques :
 - E-books : Epub, Kindle, Mobi, etc.
 - Web : HTML

Principes des formats enrichis

Décrire le contenu du document

Texte et autres éléments

Décrire leur apparence

Taille, couleur, disposition, style

Organiser le contenu

Numérotation, renvois, références, etc.

Un document finalisé a un contenu et une apparence fixée → la Forme du document

Format et structure d'un document

Article wikipedia

- Quelles méta-données peut-on associer au document ?
- Quelle structure caractérise le document ?
- Quelles apparences sont associées aux différents éléments de structure ?
- Quelle est la structure type d'un article de la Wikipedia
- Quelles procédures pour éditer l'article ?

Format et structure d'un document

Documentation de maintenance

- Quelles méta-données peut-on associer au document (contenu) ?
- Quelle structure caractérise le document ?
- Quelles apparences sont associées aux différents éléments de structure ?
- Quelles procédures pour éditer le document ?
- **convertfiles**, pdf2word en ligne : quels problèmes de conversion (style, structure, contenu) ?

Format et structure d'un document

Cette présentation

- Quelles méta-données peut-on associer au document ?
- Quelle structure caractérise le document ?
- Quelles apparences sont associées aux différents éléments de structure ?
- Quelles procédures pour éditer le diaporama ?
- **convertfiles**, pdf2word en ligne : quels problèmes de conversion (style, structure, contenu) ?

Format et structure d'un document

Rapport d'incident aéronautique

- Quelles méta-données peut-on associer au document ?
- Quelle structure caractérise le document ?
- Quelles apparences sont associées aux différents éléments de structure ?
- Quelle est la structure type d'un rapport (Observer d'autres rapports sur le site [bea](#))
- Quelles procédures pour éditer le rapport ?
- [convertfiles](#), pdf2word en ligne : quels problèmes de conversion (style, structure, contenu) ?

Pattern language approach to document description (Waller et al. 2012)

Théorie des *pattern language* (Alexander et al. 1977, cf. [page web dédiée](#))

- Théorie issue de l'architecture qui définit un ensemble de *patterns architecturaux* (253 en tout)
- Un pattern est une configuration reliant une solution (plus ou moins prêt à être appliquée) à un problème (qui revient fréquemment) dans un contexte particulier qui correspond à l'association entre
- Exemple du *pattern* WAIST-HIGH SHELF ([lien](#)) : installer des étagères à mi-hauteur pour y déposer les objets hyper-quotidiens qui sont très souvent égarés (ex. clefs)

Alexander, C., Ishikawa, S. and Silverstein, M. 1977. *A Pattern Language : Towns, Buildings, Construction*. New York : Oxford University Press.

Waller, R., Delin, J. and Thomas, M. (2012) *Towards a Pattern Language Approach to Document Description*, *Discours*, 10,

DOI : [10.4000/discours.8673](https://doi.org/10.4000/discours.8673)

Pattern language approach to document description (Waller et al. 2012)

Adaptation de la Théorie des *pattern language* au *Document Design*

Exemple des formulaires de renseignements où il est demandé de remplir :

- nom, profession, téléphone
 - le format dépend du genre du document (Figure 3a, 3b ou 4)
 - le format contraint le contenu (*extrait du père Noël est une ordure*)
 - ex. pour le numéro de téléphone → plusieurs possibilités :
 - cases – pour OCR – vs. champ libre
 - distinction *indicatif* / *numéro*
 - distinction *domicile* / *mobile*

Waller, R., Delin, J. and Thomas, M. (2012) Towards a Pattern Language Approach to Document Description, *Discours*, 10,

DOI : [10.4000/discours.8673](https://doi.org/10.4000/discours.8673)

Le métier de rédacteur technique

- Présentation du métier de rédacteur technique :
<http://christophe.kerhousse.free.fr/metier.php>
- Quelques aberrations à méditer :
<http://christophe.kerhousse.free.fr/perles.php>