HAL
archives-ouvertes.fr

Correction typographique en édition électronique

Hélène Richy, Jacques André

To cite this version:

Hélène Richy, Jacques André. Correction typographique en édition électronique. [Rapport de recherche] RR-2562, INRIA. 1995. <inria-00074119>

HAL Id: inria-00074119

https://hal.inria.fr/inria-00074119

Submitted on 24 May 2006

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.

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

Correction typographique en édition électronique

Hélène Richy & Jacques André

N° 2562

juin 1995

PROGRAMME 3

Rapport de recherche

ISSN 0249-6399

Correction typographique en édition électronique

Hélène Richy * & Jacques André **

Programme 3 — Intelligence artificielle, systèmes cognitifs
et interaction homme-machine
Projet Opéra

Rapport de recherche n° 2562 — juin 1995 — 50 pages

Résumé : Après avoir rappelé les principes de base de la correction typographique et les problèmes de sa mise en œuvre dans l'édition électronique, nous montrons comment la définition et l'utilisation de modèles de typographie permettent de prendre en compte la flexibilité des règles de typographie.

L'approche présentée ici repose sur un langage de spécification des règles traditionnelles de typographie. Ces règles peuvent avoir une portée générale, être guidées par la structure logique des documents ou bien reposer sur la reconnaissance d'unités lexicales. Ce langage permet de décrire différents jeux de règles de typographie qui tiennent compte des usages nationaux ou éditoriaux, du style des documents ou de la nature de certains textes.

Mots-clé : typographie, édition électronique, documents structurés, correction.

(Abstract: pto)

*. Irisa/CNRS, Campus de Beaulieu, 35042 Rennes cedex – *richy@irisa.fr*
**. Irisa/Inria-Rennes, Campus de Beaulieu, 35042 Rennes cedex – *jandre@irisa.fr*

Unité de recherche INRIA Rennes
IRISA, Campus universitaire de Beaulieu, 35042 RENNES Cedex (France)
Téléphone : (33) 99 84 71 00 – Télécopie : (33) 99 84 71 71

Checking typographic correctness in electronic documents

Abstract: While the structure of documents provides facilities for accessing and presenting electronic multimedia documents, few attention is given to the typography of these structured documents. The objective of this paper is to show how the specification of typography models can help for checking the typography of structured documents.

An approach based on a language for specifying typographic rules is described. The typographic rules are typed and mainly rely on the structure of the documents. When defining various typography models composed with this language, it is now possible to adapt typographic checking with various document style and publishing uses.

Key-words: typography, electronic publishing, structured document, correctness

Table des matières

1 Introduction **4**
1.1 Besoins : correction typographique et multimédia 4
1.2 État de l'art 5
1.2.1 Imprimerie et édition électronique 5
1.2.2 « Savoir-faire »typographique 6
1.2.3 Codes typographiques 7
1.2.4 Typographie et linguistique 9
1.3 Objectifs 10

2 Classification des règles typographiques **11**
2.1 Typologie par signe ou par nature 11
2.1.1 Emploi des capitales et petites capitales 11
2.1.2 Usage de l'italique, du gras, du souligné, etc. 12
2.1.3 Écriture des nombres 12
2.1.4 Abréviations 12
2.1.5 Ponctuation 13
2.1.6 Coupure et séparations des mots 14
2.2 Typologie par contexte 14
2.2.1 Structure logique 14
2.2.2 Structure syntaxique 15
2.2.3 Sémantique 16
2.2.4 Langue 18

3 Langage de typographie **19**
3.1 Structure de document et typographie 20
3.2 Organisation d'un modèle de typographie 20
3.2.1 Alinéa 21
3.2.2 Fonctions de contrôle typographique 21
3.2.3 Règles de typographie 23
3.2.4 Modèles de composition 24
3.2.5 Règles appliquées aux éléments de structure 25
3.2.6 Règles appliquées aux attributs 26
3.3 Spécification d'un modèle de typographie 27

4 Exemple de mise en œuvre **28**
4.1 Principe 28
4.2 Modèles 29
4.3 Fonctions de contrôle 30
4.4 Interface 31
4.5 Résultats 34

5 Conclusion **35**

A Fonctions de contrôle **38**

B Modèle de typographie **40**

1 Introduction

La qualité typographique des documents contribue à faciliter la lecture et à la rendre plus agréable. Avec le développement du multimédia, on assiste à la prolifération de documents électroniques qui sont rarement passés entre les mains des professionnels de l'imprimerie ou de l'édition. Et cela se voit ! Les fantaisies de composition ne suffisent pas à rendre ces documents attractifs. La lecture des documents, y compris ceux que l'on commence à appeler des documents multimédia, est facilitée si certaines règles de composition sont respectées. C'est pourquoi, pour obtenir une qualité professionnelle au niveau de la correction typographique en édition électronique, il est essentiel de partir de l'expérience des professionnels de l'imprimerie.

Ce rapport propose donc de formaliser les règles de typographie en se basant sur les codes typographiques qui font référence en ce domaine afin d'introduire la vérification typographique dans l'édition électronique. Nous proposons ici un langage de typographie que nous avons développé et testé dans l'éditeur de documents structurés Grif [40].

1.1 Besoins : correction typographique et multimédia

S'intéresser aux règles de composition et à la typographie peut paraître anachronique au moment où le texte semble céder le pas aux images et à l'audiovisuel, c'est-à-dire à une époque où l'édition électronique s'ouvre au multimédia. Pourtant dans le cadre du multimédia, la correction typographique se justifie. Il y a plusieurs raisons à cela.

– Les documents multimédia ne contiennent pas seulement de l'image ou du son. Le texte, souvent plus précis que certaines icônes, y occupe une place importante. C'est aussi un moyen de titrage et de mise en valeur des autres média.

– La « lisibilité » des documents (voir en 1.2.1), qu'ils soient imprimés sur papier ou affichés sur des écrans, exige toujours une typographie de qualité. À la différence des documents imprimés qui passaient entre les mains de correcteurs professionnels de l'imprimerie, les documents électroniques échappent désormais le plus souvent à ce savoir-faire.

– De plus en plus souvent, des documents sont engendrés automatiquement en réponse à des requêtes, par exemple, lors de la consultation de bases de données ou de bases documentaires. La production automatique de documents exige non seulement que le contenu des documents ainsi (re)constitués soit pertinent, mais aussi que la présentation et la typographie soient adaptées dynamiquement à ce contenu.

– Les règles typographiques sont mal connues des utilisateurs [1]. Et rares sont les systèmes commerciaux d'édition électronique qui, comme ProLexis ou Hugo Plus, effectuent des vérifications typographiques : les contrôles se limitent généralement à la taille des espaces, la parité des parenthèses ou l'usage des apostrophes.

1.2 État de l'art

1.2.1 Imprimerie et édition électronique

Contrairement à ce que craignaient les professionnels de l'imprimerie depuis une vingtaine d'années, la composition n'a pas été « dévorée » par l'industrie informatique. On a assisté essentiellement à la simplification des tâches de saisie et à l'automatisation des opérations de césure et de justification. Mais la qualité des logiciels et le confort des interfaces cachent encore certaines insuffisances qui concernent, en particulier, la typographie.

Comme le souligne Alan Marshall [33], les activités typographiques sont en train de se reconstituer autour des techniques informatiques. L'informatisation de la composition donne en quelque sorte un nouvel essor à la production typographique. La micro-édition ne diminue pas le besoin des savoir-faire typographiques pour la mise en page des textes. Mais ces savoir-faire doivent pouvoir être transmis [2] à un public plus large. La rencontre de ces deux mondes que sont l'imprimerie et l'informatique permet de redécouvrir la richesse et les traditions [3] de la typographie.

Les typographes ont pour coutume de justifier les règles typographiques par un critère de « lisibilité ». Mais ce critère reste très subjectif. Et si de nombreuses études se disent basées sur la vision (voir notamment [44] [32]), il s'avère qu'on ne sait pas

1. Contrairement à l'orthographe et à la grammaire, la typographie n'est pas enseignée à l'école.
2. Rappelons ici que « tradition » vient du latin *traditere* qui veut dire « transmettre ».
3.

très bien quantifier cette lisibilité. Il vaudrait d'ailleurs mieux parler de lisibilités au pluriel [37]. En effet, la lecture se fait à plusieurs niveaux (rétine, détection des éléments, reconnaissance des lettres, reconnaissance des mots, compréhension). L'interprétation des résultats d'expérimentation doit donc tenir compte de ces divers niveaux et particulièrement du dernier (compréhension, très liée aux habitudes et à la culture de chacun [4]). Comme le fait remarquer Robert Morris [35], il n'est pas rare de voir publiées des opinions contraires, par exemple certains auteurs « prouvant » que les caractères à chasse fixe sont plus « lisibles » que ceux à chasse variable tandis que d'autres « prouvent » le contraire.

La maîtrise de la typographie est donc une question complexe qui a jusqu'ici été réservée aux professionnels. Dans le but de définir le fonctionnement d'un vérificateur typographique, nous proposons donc de partir de l'analyse de leur savoir-faire et des codes typographiques qui font référence en ce domaine.

1.2.2 « Savoir-faire » typographique

Le travail typographique peut être envisagé à trois niveaux :

1. Au niveau le plus bas, celui du signe, on parle généralement de **typographie numérique**. Cela concerne la définition des caractères pris individuellement, comme par exemple l'emploi des *hints* pour les imprimantes à laser ou l'utilisation du *greyscale* pour l'affichage des caractères sur écran [23].

2. Au niveau le plus haut, celui de la mise en page, on parle de présentation graphique ou d'assemblage (maquette) ou encore de **macro-typographie** [42]. Le travail typographique est une « édition visuelle » qui répond surtout à des critères d'esthétique et de lisibilité. Il a pour objectif d'assurer la cohérence graphique du texte, la présentation du texte (marges, titrage, notes, index, bibliographie) et la disposition des éléments graphiques (illustrations, figures). La vérification de cette présentation graphique, qui était réalisée dans les ateliers de composition, est maintenant réalisée en composition électronique, sur écran.

4. Les typographes français justifient par exemple l'emploi d'une fine avant le point-virgule par le fait que sinon, la lisibilité serait moins bonne. Mais l'œil français est-il moins bon que celui des anglo-saxons ? Par ailleurs, Ladislas Mandel [34] fait remarquer que les Allemands ont longtemps employé des gothiques (en fait des fractures), qui sont pour les français complètement « illisibles », mais pas pour eux.

3. Au niveau intermédiaire, le niveau du texte, et indépendamment de toute mise en page, le travail typographique consiste à contrôler la juxtaposition des caractères. C'est ce que François Richaudeau [42] appelle la **micro-typographie**. La vérification de la micro-typographie consiste en une « édition textuelle » visant à rectifier les fautes de grammaire (syntaxe, orthographe, ponctuation, usage des capitales, de l'italique, etc.) et les erreurs typographiques (espacement irrégulier, choix des polices de caractères). Cette correction « micro-typographique » suppose à la fois une compréhension du contenu, des connaissances linguistiques et des connaissances typographiques.

Dans la suite de ce rapport, nous nous intéressons surtout à ce niveau intermédiaire, qui concerne la composition du texte (la micro-édition). Le mot **typographie** désignera donc ce qui appartient généralement au domaine de la micro-typographie.

1.2.3 Codes typographiques

De nombreux ouvrages traitent de typographie. La lecture des principaux codes typographiques en usage aussi bien en France que dans d'autres pays, fait apparaître quelques différences. Les codes les plus significatifs, utilisés pour les textes écrits en français sont les suivants :

– le *Lexique des règles typographiques en usage à l'Imprimerie nationale* [5],

– le *Code typographique* de la fédération nationale des industries polygraphiques [2],

– le *Guide du typographe romand* [4].

Selon les éditeurs ou les pays dans lesquels les documents sont publiés ou lus, les usages diffèrent.[5] De plus, la composition d'un ouvrage technique ne fera pas appel aux mêmes savoir-faire que celle d'un ouvrage littéraire ou publicitaire. Il est donc important de considérer la typographie non pas comme un ensemble rigide de règles, mais plutôt comme des recommandations qui évoluent dans le temps (mais avec une grande stabilité : voir [7] [9]) ou avec certaines modes ou certains styles.

5. Il est d'ailleurs significatif que le titre de [5] stipule « en usage... ». En fait, de nombreux codes typographiques décrivent les règles maison : [21] pour Hachette, [3] pour le Seuil ou [1] pour les journalistes.

Comme le rappelle justement Fernand Baudin [12] [13], certains usages changent brutalement. Tel est le cas de la coupure de mots, par exemple : au cours des trois premiers siècles de l'imprimerie, il était normal de placer des traits d'union sur plusieurs lignes consécutives pour couper les mots des lignes trop longues. Mais, en 1825, Firmin Didot considéra cet usage désagréable et préconisa d'éviter toute coupure. Il fixa même la tolérance à trois coupures consécutives au maximum. Et c'est finalement la règle qui prévaut depuis cette date.

D'autres règles ne font pas le consensus des codes typographiques. Par exemple, le *Lexique des règles typographiques en usage à l'Imprimerie nationale* [5] préconise l'usage des capitales accentuées, alors que le *Guide du typographe romand* [4] recommande, au contraire, de ne pas mettre d'accent à la capitale initiale d'un mot en bas de casse (Emile, Eve). En revanche, le guide préconise de mettre les accents dans les mots ou phrases entièrement en capitales (DÉJÀ, ÉMILE).

Les différents codes typographiques s'accordent cependant sur un certain nombre de points. Ils sont fondés sur les mêmes principes :

Flexibilité et tolérance Ces codes ne stipulent pas des règles rigides, mais plutôt des recommandations. Il est tout à fait possible de déroger à certaines règles pour des considérations techniques ou de mise en page. Par exemple, le *Code typographique* stipule que, dans un texte français, le signe deux-points doit être précédé d'une espace d'un demi-cadratin. Mais cette espace pourra être supprimée si cela évite la coupure du mot qui le précède.

Cohérence L'utilisation cohérente des règles doit permettre d'éviter les ambiguïtés. Par exemple, ces codes préconisent d'utiliser l'italique principalement pour attirer l'attention du lecteur sur un mot, une phrase ou un passage que l'auteur tient à souligner. Mais l'italique est aussi utilisé pour différencier les mots étrangers d'un texte. L'abus de l'usage de l'italique risque d'en diminuer l'efficacité.

Homogénéité Un document doit conserver un aspect homogène. Mais, selon le style du document, les usages peuvent varier. Il importe surtout de respecter une unité de présentation. Par exemple, on veillera à un emploi unifié des expressions synonymes (*cf.* / *voir*, *op. cit* / *ouvr. cité*), lorsqu'elles sont autorisées.

Extensibilité Comme les réactualisations des différents codes ou manuels de typographie le montrent [7] [9], si les usages évoluent peu, ils évoluent tout de même, même si ce n'est que lentement. Il faut donc pouvoir intégrer facilement les nouveaux usages et supprimer les règles tombées en désuétude.

Voici quelques exemples récents des nouvelles tendances françaises. L'usage des points abréviatifs dans les sigles tend à disparaître (S.N.C.F devient SNCF) ; les lettres minuscules (faute de petites capitales) remplacent les petites capitales dans les acronymes prononçables (INRIA devient Inria, mais VAL reste ainsi [18]) ; le point des milliers a disparu dans les nombres écrits en chiffres arabes (1.000 devient 1 000) ; le point remplace souvent la virgule décimale en informatique ou pour les sommes d'argent.

1.2.4 Typographie et linguistique

Bien que la typographie et la linguistique aient toutes deux comme champ d'application le « texte », leur vision en est fondamentalement différente. Si les typographes ont parfois besoin de s'appuyer sur la linguistique ou du moins sur la langue, l'intérêt des linguistes pour la typographie est plutôt rare. Les grands classiques, comme le Girodet [20] ou le Grevisse [22], ne font allusion à la micro-typographie qu'en une courte annexe pour le premier ou à l'occasion de rares notes pour le second. Même si Jacques Drillon [16] termine chaque étude de signe par un alinéa de conventions typographiques, l'ensemble ne doit faire que quelques pages dans un ouvrage qui en compte près de cinq cents. Ce rapport est en revanche meilleur chez Jaquenod [28].

Toutefois, l'écrit, en donnant à ce mot une connotation graphique, n'est pas complètement absent des recherches littéraires. Jacques Anis [10] a publié une importante synthèse sur l'écriture où il propose une approche du système graphique français fondée sur l'autonomie de la forme écrite de la langue, mais finalement il s'intéresse surtout aux rapports entre signifiants et espaces graphiques. De même, le Centre de de Recherche sur l'Écriture de Paris 6 travaille aussi sur les rapports du « dire » et du « voir » [15]. La linguistique structurale, non plus, s'est peu intéressée à la typographie, sauf dans certaines études sur la ponctuation [24].

Mais c'est sans doute dans l'équipe de Jacques Virbel à Toulouse que l'on a trouvé les publications le plus en relation avec notre travail [39] [45] [46] . Mais là encore, la micro-typographie n'a pas été vraiment abordée.

Il ressort finalement de tout celà que les problèmes de micro-typographie ne semblent pas intéresser les linguistes, alors que de notre côté, nous devons nous appuyer sur la linguistique.

1.3 Objectifs

Les travaux décrits dans ce rapport ont pour objet de dégager les concepts de la typographie afin de permettre le développement d'outils de vérification et de correction typographique destinés aux utilisateurs d'éditeurs électroniques. Sans prétendre remplacer un prote (correcteur professionnel) et en évitant de compliquer la tâche des utilisateurs qui ignorent généralement les règles et usages typographiques, de tels outils peuvent contribuer à améliorer la qualité typographique des documents.

Notre approche consiste à définir et à utiliser un **langage de typographie** permettant de décrire des ensembles de règles constituant un modèle de typographie. Cette approche est analogue à celle qui consiste à utiliser un langage pour définir la structure des documents (SGML [25] ou le langage S de Grif [41]) et un langage pour décrire la présentation (DSSSL [27] ou le langage P de Grif [41]).

L'utilisation d'un modèle de typographie pour décrire les règles de typographie à appliquer à un document permet une application uniforme de ces règles dans le document. En outre, l'utilisation d'un même modèle de typographie pour plusieurs documents est une garantie de cohérence et d'homogénéité de ces documents. Mais un modèle de typographie est aussi spécifique à un type de document (rapport, transparents, ...), au style de publication (littéraire ou technique) et à la langue utilisée pour le rédiger. Plusieurs modèles de typographie peuvent donc être définis pour répondre à ces différents besoins. En outre, ces modèles peuvent être modifiés ou évoluer pour tenir compte de nouveaux styles de composition typographique. La flexibilité des règles typographiques et l'absence de norme officielle en ce domaine incitent à définir un correcteur typographique qui soit piloté par les règles définies dans des modèles.

L'analyse des usages typographiques nous a conduit à proposer les deux modes de classification des règles de typographie qui sont présentés dans la section B. Ces classifications servent de base à la définition du langage de typographie décrit dans la section 3. Ce langage permet d'indiquer, pour chaque type de règle, quelle est la règle typographique en vigueur, à quels éléments d'un document structuré et dans quelles conditions elle s'applique : par exemple, à tous les titres, aux paragraphes rédigés en

français, au dernier élément d'une énumération, etc. Un exemple de mise en œuvre dans l'éditeur structuré Grif est présenté dans la section 4. Enfin, en conclusion, nous évaluons les premiers résultats de ce mode de correction typographique, basé sur la structure des documents.

2 Classification des règles typographiques

Les codes de typographie classent généralement les règles selon la nature des signes employés : usage des capitales, de l'italique, de la ponctuation, des abréviations, manière d'écrire les nombres, division des mots et des phrases, etc. Cette classification est rappelée dans la section 2.1.

Cependant, comme de nombreuses règles dépendent de la structure logique, de la structure syntaxique ou de connaissances linguistiques (sémantique ou textuelle), il est intéressant de classer ces règles dans la perspective d'une analyse structurale et textuelle des documents. Cette seconde classification est présentée dans la section 2.2.

2.1 Typologie par signe ou par nature

La classification typographique que nous présentons dans cette section donne un aperçu des règles de typographie selon la nature des signes employés. Comme les codes typographiques ne sont pas d'accord entre eux sur certaines règles (capitales, espacement, etc.), nous restons à un niveau général et n'indiquons pas toutes les nuances mentionnées dans les différents codes (pour plus de détails, on pourra se reporter à la dernière version de *The Chicago Manual of Style* [6]).

2.1.1 Emploi des capitales et petites capitales

La capitale est le signe distinctif du nom propre. Elle s'emploie dans des titres, des institutions, des noms de lieux, etc. Beaucoup d'auteurs abusent des capitales par méconnaissance de ces règles : il faut, par exemple, écrire *Académie française*, *académie de Rennes*, *inspecteur d'académie*. La liste exhaustive de tous les usages de capitale initiale ne figure dans aucun code typographique. Cet usage dépend aussi du sens dans lequel les termes sont employés. S'il arrive qu'un auteur emploie

volontairement la capitale dans certaines expressions, cet emploi doit être « justifié », compris par le lecteur et cohérent.

La capitale est aussi employée en début de phrase, en début d'alinéa, dans certains sigles ou certaines abréviations. Les petites capitales sont employées pour les noms des auteurs (dans les bibliographies et les notes, par exemple), dans certains sous-titres ou pour le ou les premiers mots suivant une lettrine.

2.1.2 Usage de l'italique, du gras, du souligné, etc.

Italique, gras et éventuellement soulignement sont utilisés pour mettre en évidence. L'italique s'emploie en particulier pour les locutions ou mots étrangers, les titres d'œuvres, de journaux, les citations textuelles, etc. Mais alors que certains mots d'origine étrangère intégrés à la langue française restent en romain, d'autres s'écrivent souvent en italique. L'important est l'homogénéité de la composition. Ainsi, si l'on écrit *a priori*, il ne faudra pas, quelques phrases plus loin écrire « à priori », à la française.

2.1.3 Écriture des nombres

Les nombres peuvent être composés soit en toutes lettres, soit en chiffres arabes, soit en chiffres romains. C'est la « sémantique » de ces nombres qui détermine leur écriture : par exemple, ils doivent être composés en toutes lettres pour exprimer la durée (*trois ans*), dans les actes notariés ou en début d'une phrase (sauf s'il s'agit d'une date ou d'un ouvrage statistique). Ils doivent être composés en chiffres arabes pour exprimer des distances (*17 mètres*), des sommes ou certaines heures. Mais les siècles (*XIX*^{e} *siècle*), les arrondissements ou les tomes (*tome VII*) doivent être composés en chiffres romains.

2.1.4 Abréviations

Très fréquentes dans les ouvrages techniques, les abréviations tendent à disparaître des ouvrages généraux. La composition de ces abréviations pose un certain nombre de questions :

– point abréviatif : la tendance actuelle est à la disparition des points, mais ils demeurent lorsqu'il peut y avoir confusion avec d'autre mots, non abrégés (en

anglais : *in.*, *a.*) ou lorsque l'abréviation se fait par retranchement des lettres finales (en français : *chap.*, *cat.*).

– espace entre les éléments abrégés (en anglais, on ne mettra pas d'espace dans *N.Y.*, mais les espaces internes sont conservés dans *Sgt. Maj.*ou *Gov. Gen.)*

– usage des capitales et des exposants (en anglais, *Mr.*, *Mrs.*, *Messrs.* et *Dr.*, en français *M.*[6], Mme, Mlle, *MM.* et D^{r}).

– pluriel des abréviations : en français, on marque le pluriel des mots abrégés par l'adjonction d'un *s* si l'abréviation comporte la dernière lettre du mot (M^{es}pour maîtres, N^{os} pour numéros).

– par contre, on n'ajoute pas, contrairement à l'anglais, de *s* pluriel aux abréviations (les *SGDB* et non les *SGBDs*) et surtout pas aux unités (*12 kg* et non *12 kgs*).

Les réponses à ces questions n'étant pas toujours faciles à déterminer, la plupart des codes typographiques contiennent des listes d'abréviations destinées à aider les auteurs.

2.1.5 Ponctuation

La ponctuation (balancement des virgules, rythme et respiration de la syntaxe, etc.) [16] permet de clarifier la lecture ou d'indiquer certaines nuances propres à un auteur. Les usages peuvent varier d'un auteur à l'autre. Mais ces usages sont très précis dans certaines parties spécifiques de documents tels que les notes, la bibliographie ou les index.

Les règles de composition de la ponctuation étant pratiquement systématiques, certains codes typographiques, par exemple les règles de l'imprimerie nationale [5] (page 149), décrivent la disposition des espaces avant ou après les signes de ponctuation sous la forme de tables indiquant la règles d'espacement de chaque caractère de ponctuation et de parenthésage. Par exemple, il faut mettre une fine avant les symboles : ; ! et ?, il faut mettre une fine avant le guillemet fermant et après le guillemet ouvrant, etc.

6. Parmi les fautes les plus fréquentes, citons l'emploi de « *Mr* » pour abréger « *Monsieur* » alors que « *M.* » est bien plus français.

Ces règles doivent cependant être nuancées : lors de courtes justifications, on pourra, par exemple, coller le point-virgule au mot qui le précède. Enfin, le signe de ponctuation doit être présenté dans le même style que le mot, la lettre ou le symbole qui le précède immédiatement.

2.1.6 Coupure et séparations des mots

La coupure des mots (français) se fait selon l'épellation française : coupure syllabique ou éventuellement coupure éthymologique. Mais il faut éviter de couper les noms propres et les mots-composés. On trouvera dans [19] des références récentes sur ce sujet.

Lors de la mise en ligne du texte, certains mots ne doivent pas être séparés : le jour du mois, l'abréviation du nom qui suit (après *M.*, Mme, D^{r} etc.), les heures, les pourcentages, les signes de ponctuation, les nombres de leurs unités (*12 m*).

2.2 Typologie par contexte

Les règles typographiques citées dans la section 2.1 dépendent souvent du contexte. Il peut s'agir soit de dépendances du contexte structural, du contexte syntaxique ou encore d'un contexte sémantique ou linguistique. Dans les documents structurés, l'analyse du contexte structural est évidente. Par contre, l'identification des autres types de contexte (syntaxe, sémantique, langue) pose des problèmes qui ne peuvent être résolus que par l'utilisation d'outils linguistiques.

2.2.1 Structure logique

Certaines règles ne sont applicables que dans certains types de structure.

> **Exemple** : les titres n'ont pas de ponctuation finale (point, virgule, point-virgule).

L'édition structurée qui repose sur la structure logique des documents permet d'identifier facilement ce type de contexte :

– En premier lieu, elle oblige toujours à identifier la structure d'un document avant de pouvoir le manipuler. Le choix d'une structure générique (DTD

en SGML ou schéma de structure en Grif) reflète le style du document. La typographie (l'usage des ponctuations, des capitales...) sera différente selon que l'on s'intéresse à un article constitué de sections ou à un jeu de transparents pour un exposé.

– Ensuite, chaque document structuré est lui-même composé d'éléments typés, tels que liste, énumération, titre de chapitre, titre de section. À chacun de ces types peuvent être appliquées des règles différentes : l'usage des capitales dans les titres, la ponctuation finale des listes, etc.

– La position de ces éléments dans la structure logique du document est également importante : selon qu'un paragraphe précède une liste ou une énumération, par exemple, la ponctuation finale de ce paragraphe sera différente : soit deux-points (:), si le paragraphe est suivi d'une liste ou d'une énumération, soit une ponctuation finale ordinaire (. ! ?), si ce paragraphe est suivi d'un autre paragraphe.

– Enfin, la structure arborescente des documents permet que des règles qui ont été définies pour un type d'élément soient appliquées à toute l'arborescence sous-jacente par héritage. Inversement, cet héritage peut être limité à certains types de règles ou à certains types d'éléments uniquement. Par exemple, on met généralement une capitale au premier mot de tous les alinéa (titre, sous-titre, paragraphe, légende, ...). Mais, certaines énumérations peuvent commencer par une minuscule.

2.2.2 Structure syntaxique

Certaines règles dépendent de la structure syntaxique du texte, c'est-à-dire de la construction des phrases.

Exemples : Mettre une capitale en début de phrase.
Placer un point final en fin d'une liste, sauf si la phrase englobante n'est pas terminée à la fin de la liste.
Ne pas placer de virgule avant une parenthèse.
Faire précéder « etc. » d'une virgule.

L'usage des ponctuations fait partie des prérogatives de l'auteur et reflète souvent un certain style d'écriture [16] [28]. Mais souvent l'usage des ponctuations est négligé, voire incohérent. Certaines fautes grossières devraient pouvoir être aisément corrigées si l'on est capable d'analyser syntaxiquement ou grammaticalement le texte (virgule entre sujet et verbe ou entre verbe et complément, points de suspension après « etc. », absence d'un point d'interrogation dans une phrase interrogative...).

Avec le concept de documents structurés, la connaissance de la structure des documents s'arrête généralement au niveau des paragraphes. Quelques passages de texte sont repérés pour leur présentation ou leur sémantique par un attribut (emphase ou citation, par exemple). Étendre cette structure à toutes les phrases (structure syntaxique) et à tous les groupes de mots (structure sémantique) provoquerait un morcellement des textes et en rendrait difficile la manipulation interactive. Il faut donc envisager une méthode de reconnaissance des phrases d'un texte ne faisant pas appel à l'utilisation systématique d'attributs.

Lorsqu'on ne dispose pas d'outil d'analyse syntaxique, l'identification des phrases d'un document peut reposer sur la localisation des ponctuations, supposées pertinentes - ce qui n'est pas toujours le cas. Mais la distinction entre un point final et un point abréviatif n'est pas toujours évidente. L'utilisation préalable de lexiques d'abréviations permet de lever certaines ambiguïtés (voir 2.2.3). En conséquence, avec ce type d'identification syntaxique (et lexicale), il est indispensable de procéder au préalable à une identification des abréviations et des noms propres, si l'on veut identifier des phrases dans les meilleures conditions.

2.2.3 Sémantique

Certaines règles dépendent de la nature des mots ou de leur sémantique.

Exemples

L'emploi des capitales : *le Paris de Sauter* (le fabricant de cuisinières) ou *les paris de sauter* (le verbe).

L'emploi des abréviations : utiliser l'abréviation de Madame (Mme) ou Monsieur (M.) dans le texte courant, lorsqu'il est suivi d'un nom de personne ou de qualité (*M. le préfet*), mais pas dans un titre d'ouvrage (*Madame Sans-Gêne*) ou lorsque l'on s'adresse à la personne elle-même, comme au début ou à la fin d'une lettre.

L'emploi de l'italique pour les « noms d'œuvre » : *Le Monde*.

L'espacement autour des points de suspension : mettre un espace avant les points de suspension mis pour un mot unique, mais les coller à la dernière lettre s'il tiennent lieu de fin de phrase ou de mot.

Lorsqu'on ne dispose pas d'outil d'analyse linguistique, la reconnaissance des différentes sémantiques peut reposer sur des lexiques. Cette approche permet d'identifier certaines catégories lexicales dont on a pu dresser une liste. Des **lexiques** spécialisés sont constitués pour permettre l'identification sémantique de certains mots ou groupes de mots. Chaque lexique contient une liste de mots ou groupes de mots usuels appartenant à une même catégorie - au sens des usages typographiques : des titres d'ouvrages, des noms de personne, des lieux géographiques, des noms de logiciels, des sigles, des noms d'universités, des dates célèbres, des termes techniques, etc. L'appartenance à un lexique détermine la catégorie d'unité lexicale : un mot est identifié comme un nom de lieu géographique (resp. abréviation, sigle) s'il appartient au lexique des noms de lieux (resp. abréviations, sigles). Par exemple, on trouvera « Paris » dans un lexique des villes, « Mgr » dans un lexique d'abréviations et « Irisa » dans un lexique de sigles. Chaque lexique typographique peut contenir pour chaque mot ou groupe de mots la description de particularités utiles à la correction typographique :

– des informations typographiques : usage de l'italique, des exposants, des capitales, petites capitales, etc. (GMT s'écrit en majuscules, *ibid.* en italique, D^{r} avec une capitale initiale suivie d'un exposant).

– des informations correctives : un choix de termes invalides peut être associé à un terme valide afin de faciliter la détection et la correction des erreurs typographiques les plus courantes (« c.-à-d. » abréviation de « c'est-à-dire » écrit par erreur « c'est à dire », « etc. » écrit souvent par erreur « etc... »).

– des informations sémantiques. Par exemple, le qualificatif *littéraire* ou *technique* permet de choisir entre la forme au long et la forme abrégée de certains termes.

2.2.4 Langue

Certains usages typographiques dépendent de la langue[7]. Si l'on compare, des textes écrits en anglais et des textes écrits en français, on remarque, par exemple, que l'usage des espaces autour des ponctuations diffère et que l'utilisation des ponctuations pour rythmer les phrases est différente.

Exemples

En anglais, tous les mots « significatifs » d'un titre peuvent porter une capitale initiale.

En anglais, les signes deux-points et point-virgule ne sont jamais précédés d'un espace.

En français, les mots étrangers (latins, anglais, etc.) sont généralement en italique.

Pour un francophone, écrivant un texte en anglais, une erreur fréquente consiste, dans une liste de plus de deux auteurs, à omettre la virgule avant le « *and* » final (on doit écrire « *Vincent Quint, Cécile Roisin, and Irène Vatton* »).

L'usage et les règles d'imbrication des guillemets sont très différents d'une langue européenne à l'autre [11].

Et bien sûr, certains sigles diffèrent en fonction de la langue (USA ou É-U, UN ou ONU).

L'identification de la langue d'un document est donc fondamentale si l'on veut pouvoir appliquer des règles typographiques pertinentes. L'identification peut se faire à plusieurs niveaux :

– La langue du document est souvent connue des logiciels qui disposent de correcteurs orthographiques pour choisir le (ou les) dictionnaire(s) de référence

7. Le plus connu est l'usage des coupures de mots en fin de ligne. Cette question n'est pas abordée ici. Des algorithmes très efficaces existent. en particulier, celui proposé par F.-M. Liang dans [31] à base de motifs de césure est aisément extensible aux langues européennes. Cet algorithme, qui est notamment utilisé par TEX [29], donne d'excellents résultats.

de ces correcteurs. Cette caractéristique peut donc être fournie à un correcteur typographique.

– Lorsqu'il s'agit de documents multilingues, cette information est insuffisante : dans ce cas, il est nécessaire de connaître la langue de chaque paragraphe, de chaque texte, voire de chaque mot, si l'usage de plusieurs langues est intimement mêlé dans le document. La présence d'un **attribut** langue, prévu par certains modèles de documents (SGML [25], LaTeX [30], Grif [41]), facilite cette reconnaissance : la reconnaissance de la langue s'apparente alors à une reconnaissance de la structure logique du document.

– Enfin, pour l'utilisation occasionnelle de termes étrangers, la reconnaissance lexicale (voir en 2.2.3 ci-dessus) peut s'avérer efficace et évite à l'auteur de spécifier explicitement la valeur d'un attribut langue, s'il existe, lors de l'emploi de ces mots dans le texte. Par exemple, il est d'usage de mettre des mots tels que « *a priori* », « *a fortiori* », « *in fine* » en italique, alors que « etc. », « ad hoc » ou « in extremis » qui sont d'un usage plus courant n'en utilisent pas. En ce domaine, comme dans beaucoup d'autres, les usages typographiques évoluent (« *a priori* » s'écrit aussi souvent comme un mot français « à priori »), il est donc intéressant de pouvoir définir et utiliser son propre lexique de mots étrangers auxquels une présentation particulière, en l'occurrence l'italique, sera appliquée systématiquement par un correcteur typographique.

3 Langage de typographie

Comme on l'a vu plus haut, la structure logique des documents est une aide à la correction typographique. On rappelle donc ici brièvement la notion de « document structuré » et d'approche générique (section 3.1). La section 3.2 présente le langage, que nous proposons pour décrire les règles de typographie. Pour plus de précisions, on pourra se reporter à l'annexe A qui décrit la syntaxe de ce langage. Enfin, la section 3.3 aborde les points les plus délicats de la spécification d'un modèle de typographie. Une mise en œuvre de ce langage, appelé TY dans l'éditeur Grif, est décrite dans la section 4.

3.1 Structure de document et typographie

Un document « structuré » est un document dont le contenu est organisé selon une structure logique qui assemble des éléments tels que des titres, paragraphes, sections, chapitres, notes, figures, etc. Certains éléments peuvent porter des attributs (langue, importance, mot-clé) qui précisent leur rôle dans le document. Un attribut a un nom et une valeur. Par exemple, l'attribut « langue », associé à un élément, permet d'indiquer dans quelle langue est écrit le passage correspondant du document. L'élément sur lequel porte l'attribut peut être aussi petit qu'un mot ou aussi grand qu'un chapitre ou une section dans un document multilingue.

Cette structure est principalement arborescente, mais comporte également des liens permettant de représenter des renvois ou des références dans la structure logique du document. Les références peuvent aussi désigner des éléments qui se trouvent dans des documents différents.

L'approche générique (ODA [26], SGML [25], Grif [40]) permet de typer les éléments et de décrire leur mode de construction. Ainsi considère-t-on que des documents construits d'après une même structure générique (*Document Type Definition* de SGML) appartiennent à la même classe de documents. Par exemple, deux articles A et B, bien que différents, sont construits selon la même structure générique qui spécifie qu'un article comporte un résumé suivi d'un nombre variable de sections, chaque section comportant un nombre variable de sous-sections.

Nous proposons de définir la typographie des documents en fonction de leur structure logique. Et, pour respecter l'approche générique des modèles de documents, la typographie sera également définie de façon générique. L'approche générique de la typographie consiste à définir un *modèle de typographie* pour une classe de documents. Les règles spécifiées dans ce modèle peuvent être appliquées à toutes les occurrences de documents de cette classe. Elles s'appliquent uniquement aux parties textuelles des documents. Il est possible d'écrire plusieurs modèles de typographie différents pour la même classe de documents. Cela permet d'offrir aux maquettistes le choix de la typographie qui convient le mieux à leur style de document ou au goût des utilisateurs.

3.2 Organisation d'un modèle de typographie

Le modèle de typographie est associé à une structure générique (DTD pour SGML ou schéma de structure de document Grif). Un modèle précise, pour tous

les types d'éléments et toutes les valeurs d'attributs, quelles sont les règles de typographie à satisfaire et dans quelles conditions elles s'appliquent. Il comporte les déclarations suivantes :

1. la déclaration des alinéas (**ALINEA**),
2. la déclaration des fonctions de contrôle typographique (**FUNCTION**),
3. la déclaration des modèles de composition (**COMPOSITION**),
4. les règles de typographie des éléments structurés (**ELEMENT**),
5. les règles de typographie des attributs (**ATTR**).

3.2.1 Alinéa

La notion d'*alinéa* permet de distinguer, dans un document structuré, les éléments qui seront considérés comme des « paragraphes » par certaines fonctions typographiques. Par exemple, si la règle de vérification de la capitale initiale est définie au niveau de l'élément `Section`, elle sera appliquée, par héritage, à tous les alinéas contenus dans la section. Elle ne sera appliquée aux titres de section que si ceux-ci ont été déclarés comme des alinéas.

3.2.2 Fonctions de contrôle typographique

La classification décrite dans la section 2.1 a mis en évidence plusieurs catégories de contrôles typographiques selon les symboles concernés : signes de parenthèsage, nombres, gras, italique, capitales ou petites capitales, espaces, etc. Les fonctions de contrôle typographique utilisées dans les modèles font référence explicitement à ces catégories : chaque catégorie est représentée par un type dans le langage de typographie.

Le langage de typographie propose les types de base suivants [8] :

– **Abbreviation** : usage des abréviations,

8. D'autres types de règles peuvent être ajoutés pour contrôler l'usage des tirets (ou traits d'union), par exemple.

- **Attribute** : usage des attributs (nom propre, abréviation, sigle, etc.),
- **Capitalization** : usage des capitales,
- **Distance** : règles d'espacement,
- **Exponent** : usage des exposants,
- **InsertPair** : usage des signes de parenthèsage,
- **Punctuation** : usage des signes de ponctuation,
- **SpaceTable** : table d'espacement des caractères,
- **Word** : présentation des mots ou des unités lexicales.

Le fait de classer ces fonctions par type dans les modèles présente plusieurs avantages. Cela permet tout d'abord au correcteur typographique d'ordonner judicieusement les vérifications : par exemple, le contrôle des caractères et des espacements doit être effectué avant la détection des abréviations. En second lieu, cela permet de sélectionner une catégorie de vérification typographique : un utilisateur désireux de n'effectuer que le contrôle de l'usage des capitales, par exemple, n'aura pas à indiquer toutes les fonctions qu'il veut évaluer, mais seulement la catégorie « Capitalization ». Cela revient à extraire du modèle de typographie toutes les règles qui n'appartiennent pas à cette catégorie.

Avant de pouvoir être utilisées comme règles de vérification de la typographie de certains éléments ou attributs en particulier, toutes les fonctions de typographie doivent être préalablement déclarées dans le modèle et leur type doit être précisé. Comme la liste des fonctions n'est pas définie *a priori*, il est possible de définir de nouvelles fonctions sans modifier le langage de typographie. Dans chaque catégorie, on pourra donc disposer d'un nombre variable de fonctions. Certains modèles en utiliseront beaucoup, d'autres peu.

Dans l'exemple ci-dessous (Fig. 1), plusieurs règles du même type (**Punctuation**) sont définies :

- la fonction `PunctEnd` vérifie que les éléments sur lesquels elle est appliquée sont terminés par un signe de ponctuation finale (point, point d'exclamation, point d'interrogation ou points de suspension),

```
{\bf FUNCTION}
Punctuation: PunctEnd;
Punctuation: PunctHighNone;
Punctuation: PunctColon;
```

FIG. 1 - *Règles de ponctuation*

– la fonction `PunctHighNone` vérifie que les éléments sur lesquels elle est appliquée se terminent sans ponctuation finale, ou exclusivement avec un signe de ponctuation haute (? !),

– la fonction `PunctColon` vérifie que les éléments sur lesquels elle est appliquée se terminent par le signe de ponctuation deux-points.

3.2.3 Règles de typographie

La typographie est décrite en fonction de la structure des documents : des règles typographiques sont associées à tous les types d'éléments et à tous les attributs définis dans une structure de document. Ainsi, le contrôle typographique sera effectivement dirigé par la structure des documents.

L'ensemble des règles qui contrôlent la typographie d'un type d'élément ou d'une valeur d'attribut est toujours défini de la même manière. Un tel ensemble associé à un type d'élément s'appliquera aussi, par héritage, à tous les éléments de niveau inférieur dans la structure du document. Il est toutefois possible, à n'importe quel niveau dans la structure, de spécifier une nouvelle règle pour éviter cet héritage. En particulier, on pourra utiliser une fonction « NULL » disponible pour tous les types de fonctions [9].

L'application d'une règle de vérification typographique peut être systématique ou conditionnelle. Si une règle est spécifiée pour un type d'élément donné (un paragraphe ou un titre, par exemple), cela signifie que cette règle doit être vérifiée systématiquement par tous les éléments de ce type. Mais le langage de typographie

9. Cette fonction porte le nom FunctNil dans les exemples cités en annexe.

```
Paragraph:
 IF (Next = List) Punctuation: PunctColon;

Item_list:
 BEGIN
 IF (Last) Punctuation: PunctPeriod;
 IF (NOT IN Item_list AND NOT Last)
 Punctuation: PunctSColon;
END;

Title:
IF (Language = English)
Capitalization: CapAllWord;
```

FIG. 2 - *Règles conditionnelles*

permet également d'indiquer que l'application d'une règle est conditionnelle, c'est-à-dire d'indiquer une ou plusieurs conditions à satisfaire. Ces conditions peuvent porter sur la structure, sur la langue ou sur l'appartenance à une unité lexicale.

Une condition qui dépend de la structure peut, par exemple, indiquer que l'élément auquel elle sera appliquée ne doit pas avoir de successeur ou de prédécesseur à son niveau. Ce type de condition est particulièrement utile pour décrire certaines règles de ponctuation. La Fig. 2 montre quelques exemples de règles conditionnelles : la règle de ponctuation d'un paragraphe est différente si celui-ci est suivi d'une liste ; la règle de ponctuation du dernier élément d'une liste est différente de la règle de ponctuation des autres éléments de cette liste ; l'usage des capitales est différent dans les titres écrits en anglais et dans les titres écrits en français, etc.

3.2.4 Modèles de composition

Pour simplifier l'écriture des modèles de typographie, il est possible de définir des modèles de composition. Un *modèle de composition* permet de grouper un ensemble de règles de typographie et de lui attribuer un nom de modèle. Chaque modèle de composition peut être utilisé ensuite comme une règle de typographie.

```
{\bf COMPOSITION}
 ModTitle =
 BEGIN
 IF (Language = Français)
Capitalization: CapFirstElem;
 IF (Language = English)
Capitalization = CapAllWord;
 Punctuation: PunctHighNone;
 InsertPair: PairNone;
 END;
```

FIG. 3 - *Modèle de composition*

Par exemple, plusieurs types d'éléments peuvent partager le même modèle de composition qui spécifie à la fois l'usage des capitales, de la ponctuation et des espaces. Cet ensemble de règles n'est alors décrit qu'une seule fois dans un modèle de composition appelé, comme dans l'exemple de la Fig. 3, ModTitle, et il peut être ensuite utilisé plusieurs fois, comme le montre l'exemple de la Fig. 5.

3.2.5 Règles appliquées aux éléments de structure

Seuls les types d'éléments sur lesquels portent des règles de typographie dont indiqués dans le modèle. Le nom du type d'élément est suivi de la liste des règles de typographies qui le concernent. Ces règles sont soit des règles systématiques, soit des règles conditionnelles.

Par exemple, à la suite du mot `Paragraph` seront indiquées toutes les règles de typographie à appliquer aux éléments de type `Paragraph`. Dans l'exemple de la Fig. 4 les règles de typographie stipulent que le paragraphe doit commencer par une lettre capitale (`CapFirstElem`), que les seuls signes de parenthèsage autorisés sont les paires de parenthèses (`PairPar`) et enfin que tout paragraphe doit se terminer par un signe de ponctuation finale : point, point d'exclamation, point d'interrogation ou points de suspension (`PunctEnd`).

Un modèle de composition peut remplacer n'importe quelle règle. Plusieurs types d'éléments, tels que `Section_title`, `Caption` ou `Appendix_title` peuvent,

```
{\bf ELEMENT}
Paragraph:
 BEGIN
 Capitalization: CapFirstElem;
 InsertPair: PairPar;
 Punctuation: PunctEnd;
 END;
```

FIG. 4 - *Règles de typographie d'un élément structuré*

```
{\bf ELEMENT}
Section_title:
 ModTitle;
Appendix_title:
 ModTitle;
Caption:
 BEGIN
 ModTitle;
 Punctuation: PunctEnd;
 END;
```

FIG. 5 - *Utilisation d'un modèle de composition*

comme dans l'exemple de la Fig. 5, partager un même modèle de composition (`ModTitle`)

3.2.6 Règles appliquées aux attributs

Seuls les types d'attributs sur lesquels portent des règles de typographie dont indiqués dans le modèle. Pour chaque attribut, désigné par son nom, on peut indiquer la valeur ou l'ensemble des valeurs pour lesquelles les règles de typographie sont appliquées. Cet ensemble peut être réduit à une seule règle. Les règles sont soit des règles systématiques, soit des règles conditionnelles.

```
{\bf ATTR}
Highlight = Definition:
 Word: WordAllCap;
```

FIG. 6 - *Règle de typographie associée à un attribut*

Par exemple, la règle de la Fig. 6 stipule que tous les mots qui portent un attribut `Highlight` dont la valeur est `Definition` porteront une capitale initiale.

L'attribut langue fait l'objet d'un traitement particulier. Les règles typographiques n'apparaissent pas avec celles des autres attributs. En effet, comme on l'a vu en 3.2.3, il est possible d'écrire des règles conditionnelles portant sur la valeur de l'attribut langue en utilisant la condition « `Language= ...` ». Pour définir des modèles de composition pour documents multilingues, il suffit donc d'ajouter dans les règles appliquées aux éléments des conditions qui portent sur la langue (voir le dernier exemple de la Fig. 2).

3.3 Spécification d'un modèle de typographie

Les modèles de typographie sont décrits avec le langage présenté ci-dessus (voir dans l'annexe A la syntaxe du langage TY). Ce langage permet de décrire assez simplement des typographies variées (voir l'exemple de modèle de typographie pour la classe Article présenté dans l'annexe B).

L'efficacité de la correction typographique repose non seulement sur la qualité des modèles, leur définition rigoureuse, mais également sur les supports de cette correction que sont les lexiques et les fonctions de contrôle. La plus grande attention doit être portée aux points suivants :

- La spécification d'un modèle de typographie précis : en particulier, l'utilisation des possibilités d'héritage pour éviter de redéfinir inutilement certaines règles.
- La définition des fonctions de contrôle : par exemple, il est plus efficace de définir une fonction imposant à la fois une capitale initiale et l'usage de petites capitales que d'utiliser plusieurs fonctions pour le faire.

– Le choix des unités lexicales et la constitution des lexiques : de ce choix dépend la pertinence de la reconnaissance sémantique.

– La construction des tables d'espacement : une table d'espacement définit pour tout couple de caractères, si la succession de ces caractères est autorisée et par quel type d'espace ces caractères doivent éventuellement être séparés (espace mot, cadratin, demi-cadratin, fine). Comme les règles d'espacement sont différentes dans les textes écrits en français et dans les textes écrits en anglais, il faut disposer d'au moins deux tables d'espacement. Le rôle de ces tables dans la vérification typographique est important non seulement dans les fonctions de vérification des espaces, mais celles-ci sont également utilisées par les fonctions de vérification de la ponctuation et pour la formulation de certaines propositions de correction.

– Les possibilités de codage des caractères : il est souhaitable de disposer de plusieurs types d'espaces et de tirets de différentes tailles, de capitales accentuées et de petites capitales.

4 Exemple de mise en œuvre

Le langage de typographie dont la syntaxe est décrite en annexe a été défini conformément aux spécifications précédentes. Il est utilisé pour décrire les modèles de typographie qui régissent le fonctionnement du vérificateur-correcteur typographique intégré dans l'éditeur de documents structurés Grif.

4.1 Principe

Le principe de fonctionnement du correcteur typographique de Grif est le suivant. À la manière du correcteur orthographique de Grif [43], ce vérificateur détecte les erreurs de micro-typographie (force des espaces, usage des capitales, des abréviations, etc.) dans les textes d'un document structuré. Et pour faciliter la correction, le vérificateur fait aussi des suggestions de corrections que l'auteur peut appliquer immédiatement ou refuser.

L'approche retenue consiste à décrire différents ensembles homogènes de règles typographiques groupées au sein de modèles s'appuyant sur la structure des do-

cuments (voir 4.2). Ces modèles sont écrits en utilisant le langage TY. Les règles contenues dans ces modèles font appel à des fonctions de contrôle (décrites en 4.3) qui identifient les erreurs typographiques dans le texte des documents. Le dialogue entre correcteur humain et vérificateur électronique (décrit en 4.4) est particulièrement utile pour résoudre certaines ambiguïtés que seule la compréhension du texte permet de lever.

4.2 Modèles

L'éditeur Grif utilise des modèles génériques pour décrire la structure des documents. Ces modèles, appelés *schémas de structure*, sont écrits dans un langage appelé S. L'aspect graphique de tous les éléments d'une classe de document est décrit par un ensemble de règles de présentation. Ces règles sont regroupées dans un modèle, appelé *schéma de présentation*. Les schémas de présentation sont écrits dans un langage appelé P, qui permet aux maquettistes de spécifier leurs propres présentations. Comme pour la présentation, la typographie de tous les éléments d'une classe de document est décrite par un ensemble des règles. Ces règles de typographie sont regroupées dans un modèle, appelé *schéma de typographie*, écrit dans le langage TY.

Plusieurs schémas de présentation et plusieurs schémas de typographie peuvent être définis pour une même classe de documents. Par exemple, les documents de la classe Article peuvent être présentés selon une maquette imposée pour la conférence *USENIX* ou par la revue *Technique et Sciences Informatiques*. Et si les règles de typographie, comme par exemple l'usage des capitales ou des petites capitales dans certains titres ou sous-titres, diffèrent dans ces deux présentations, Grif utilisera deux schémas de typographie différents. Par contre, si les règles typographiques sont identiques, un schéma de typographie unique sera utilisé.

Nous avons adopté pour les schémas de typographie le même principe de modularité que pour les schémas de présentation et les schémas de structure. En effet, si l'on prend l'exemple des paragraphes, la typographie des paragraphes peut être identique dans des documents de classe différente. Ainsi, pour spécifier la typographie d'un article qui contient des paragraphes, des formules de mathématiques et des références bibliographiques, on utilisera les quatre schémas de structure Article, Paragraph, Math et Biblio. On utilisera également les quatre schémas de typographie correspondants à ces structures (ArticleTY, ParagraphTY, MathTY et BiblioTY).

Si l'on prend l'exemple de documents de style très différent, comme les chapitres d'un livre et les transparents d'un exposé, la typographie des paragraphes sera, à coup sûr, différente, l'utilisation des ponctuations étant très réduite dans des transparents - voire inexistante. C'est pourquoi, il est souvent souhaitable de disposer de plusieurs schémas de typographie différents pour les mêmes classes d'objets, en l'occurrence ici pour les paragraphes.

La présence de l'attribut *langue* dans tous les documents manipulés par l'éditeur Grif, permet d'identifier la langue dans laquelle est écrit un texte, un paragraphe, une section ou tout simplement un mot. Dans le cas de documents multilingues, le schéma de typographie utilisé est multilingue : il suffit pour cela de préciser pour chaque règle concernée, une condition qui dépend de la valeur de l'attribut *langue* (voir l'exemple de modèle de typographie d'un Article en annexe).

Le correcteur typographique de Grif s'appuie, pour corriger un document, sur l'ensemble de schémas de typographie qui sont associés aux schémas de structure utilisés pour construire le document. Cette association est analogue à celle qui existe entre les schémas de structure et les schémas de présentation.

Le correcteur procède de la façon suivante : pour chaque élément et pour une catégorie de règle donnée, il recherche d'abord tous les attributs que porte l'élément. Pour chaque valeur d'attribut, il applique les règles de cette catégorie définies dans le modèle de typographie. Si aucune règle ne s'applique, il applique les règles de cette catégorie définies pour ce type d'élément dans le modèle de typographie.

4.3 Fonctions de contrôle

Chaque fonction de contrôle recherche une certaine catégorie d'erreur typographique dans un texte : l'absence d'une capitale initiale, la présence d'une mauvaise ponctuation finale, l'usage d'une abréviation interdite, etc. Ces fonctions comportent deux parties : une première partie chargée de la vérification d'une règle élémentaire de typographie, une seconde partie chargée de la correction de l'erreur identifiée dans la première partie. Le principe de fonctionnement d'une fonction de contrôle est le suivant :

– recherche d'une faute de typographie dans le texte courant,

– sélection du passage ou de l'emplacement suspect : par exemple, sélection d'un mot écrit en minuscules ou d'un élément à la fin duquel manque un signe de ponctuation,

– diagnostic du type de correction (suppression, remplacement, insertion avant ou après la sélection, attribut à poser ou à supprimer),

– préparation d'une ou de plusieurs propositions de correction.

Dans l'éditeur Grif, ces fonctions sont écrites en C [10]. Pour faciliter l'écriture de ces fonctions, un certain nombre de procédures de recherche, de sélection et de correction sont disponibles. L'ajout de nouvelles fonctions ne pose pas de problème puisque le lien entre le nom de ces fonctions (dans les modèles) et le code de ces fonctions est effectué dynamiquement lors de la vérification typographique. À titre d'exemple, l'annexe B présente un modèle de typographie d'un document de style Article utilisant une vingtaine de fonctions de contrôle.

4.4 Interface

Le fonctionnement du vérificateur typographique est analogue à celui du correcteur orthographique de Grif [43]. Tous les textes figurant dans la **zone de correction** sont vérifiés. Chaque « feuille » de texte - au sens de la structure logique du document - subit successivement toutes les vérifications typographiques. Si aucune erreur typographique n'est détectée, le vérificateur contrôle la feuille de texte suivante de la zone de correction . Il s'arrête lorsque toute la zone de correction a été vérifiée.

La fenêtre de dialogue permet de définir la zone de correction (avant la sélection courante, dans la sélection, après la sélection ou dans tout le document). Dès qu'une erreur de typographie est détectée, le vérificateur sélectionne dans le document le caractère ou le passage suspect. Il signale dans la fenêtre de dialogue quel type d'erreur a été détecté. Et, si l'utilisateur a choisi une option « Correction à la demande », il indique une ou éventuellement plusieurs propositions de correction. Par exemple, dans la Fig. 7 ci-dessous, le vérificateur détecte que le mot « Multimedia » porte une capitale initiale injustifiée. Le correcteur propose, soit de remplacer la majuscule par une minuscule, soit d'insérer un point de ponctuation avant ce mot.

10. Il est prévu de pouvoir décrire directement des fonctions de contrôle typographique dans les modèles en utilisant le langage interprété Tcl [38].

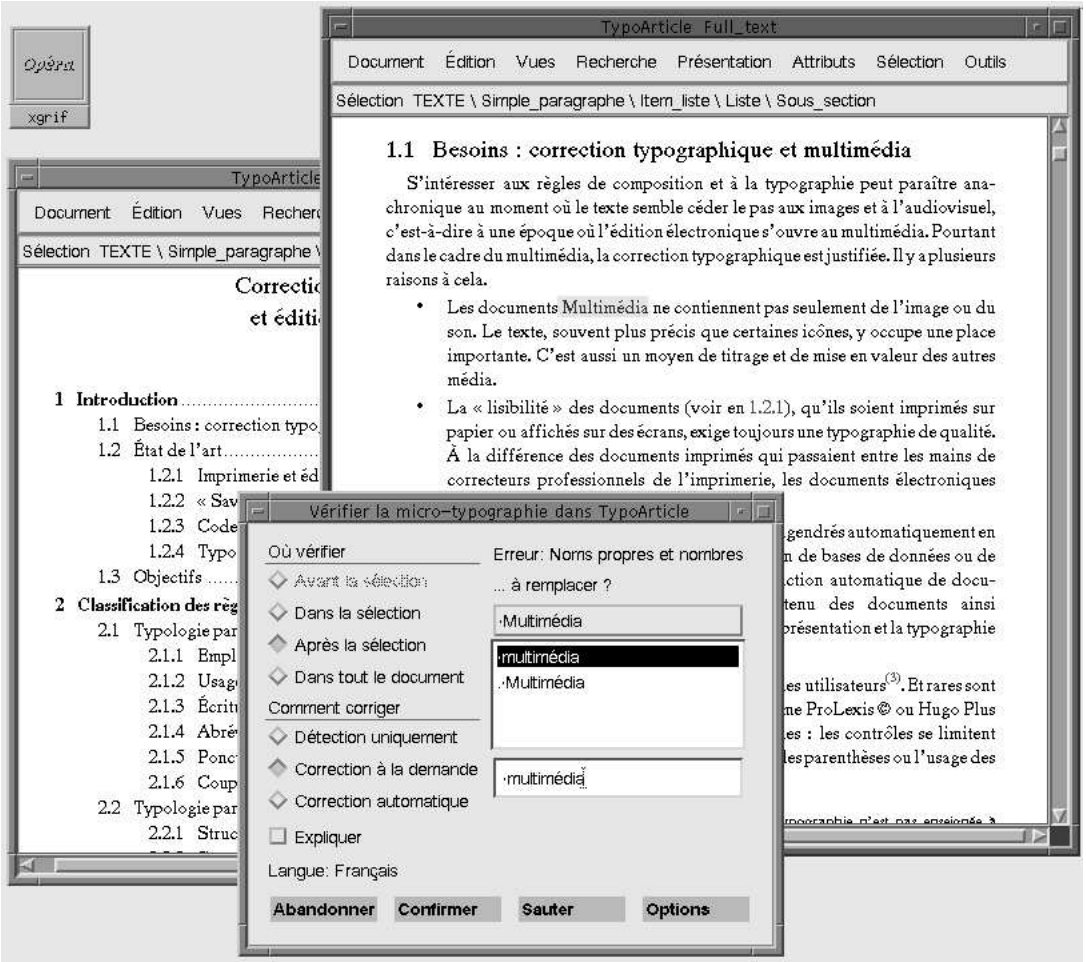

FIG. 7 - *Fenêtres de dialogue avec le vérificateur typographique*

Le correcteur typographique peut proposer l'une des corrections suivantes :

– ajout ou suppression d'un ou de plusieurs caractères : par exemple, ajout d'un espace et/ou une ponctuation en fin de paragraphe, suppression d'une parenthèse, d'un point en fin de titre, etc.

– remplacement d'un caractère ou d'un mot ou d'un groupe de mots : par exemple, remplacement d'une abréviation par le mot complet ou inversement un mot par son abréviation,

– déplacement d'un ou plusieurs caractères, par exemple déplacement du point final de la phrase avant une référence à une note,

– pose ou suppression d'un attribut : par exemple, pour harmoniser la présentation de parenthèses autour d'un mot en gras ou signaler un mot étranger (italique),

– mise en capitales ou en minuscules.

L'utilisateur est libre de choisir parmi les propositions de correction qui lui sont faites, celle qui lui paraît la plus pertinente et la valider (commande « Confirmer »). Il peut, s'il le souhaite, consulter un document d'aide en cliquant sur le bouton « Expliquer » (voir Fig. 7). Il peut aussi éditer directement le passage concerné dans la vue de son choix. Il peut évidemment choisir de n'effectuer aucune correction et poursuivre la recherche de l'erreur suivante (commande « Sauter ») ou tout simplement mettre fin à la vérification (commande « Annuler »).

Bien qu'un mode automatique de correction soit disponible, il est délicat à utiliser. La présence de l'auteur est toujours souhaitable car, en l'absence d'analyse linguistique, il est le seul à pouvoir vraiment « comprendre » le document à vérifier. La qualité des résultats obtenus par la correction automatique dépend de la quantité d'ambiguïtés rencontrées. En l'absence d'ambiguïté et lorsque le contenu des lexiques correspond bien au sujet traité (sigles et vocabulaire technique correctement identifiés), le mode de correction automatique peut être intéressant.

Le formulaire de la Fig. 8 propose un choix d'options qui permettent d'agir sur le fonctionnement du vérificateur typographique. Dans la liste des vérifications à effectuer, les options qui apparaissent en noir correspondent aux types de contrôles typographiques qui ont été définis dans le modèle de typographie du document.

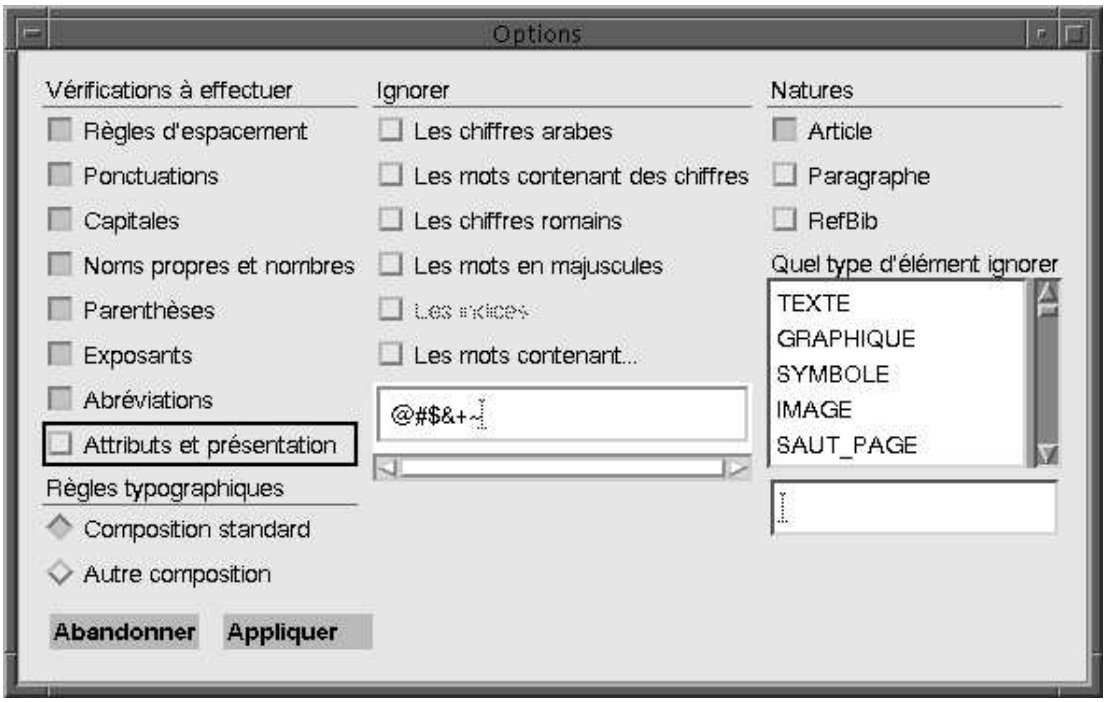

FIG. 8 - *Fenêtre des options de la vérification typographique*

Pour limiter la vérification typographique à certaines catégories de contrôle, il suffit de cocher une ou plusieurs options (en noir dans le formulaire). Pour changer de modèle de typographie, il suffit de cliquer sur « Autre composition ». L'interface permet également d'ignorer certains types d'éléments ou certains mots contenant des symboles particuliers. Toutes ces options sont modifiables à tout moment au cours de la correction.

4.5 Résultats

L'expérimentation en cours montre que les performances d'un système de correction typographique interactif sont satisfaisantes, aussi bien en ce qui concerne le taux de détection des erreurs, la qualité des corrections proposées, que le temps de réponse. La détection des erreurs de typographie se fait à la même vitesse que celle des erreurs orthographiques. La détection est immédiate pour un utilisateur

disposant d'un poste de travail sur lequel est installé l'éditeur Grif. En outre, le vérificateur fournit une aide à la correction en proposant des corrections typographiques. Ces suggestions de correction reposent sur les règles typographiques préalablement définies dans les modèles. La qualité des résultats obtenus dépend donc de la qualité de ces modèles.

À la différence des règles orthographiques qui sont - on peut le supposer, en milieu universitaire - relativement bien connues, les règles typographiques sont souvent méconnues des auteurs. Il s'avère donc très utile de concevoir, comme nous l'avons fait dans Grif, non seulement un « détecteur » de fautes typographiques, mais également un « correcteur » typographique. Le temps de réponse d'un tel correcteur est donc difficile à évaluer puisque le temps passé devant l'écran dépend non seulement du nombre de fautes détectées, mais aussi du temps de réflexion que prend l'utilisateur avant d'accepter ou de refuser certaines corrections délicates.

5 Conclusion

La connaissance de la structure logique des documents offre de nombreux avantages pour les applications qui travaillent sur la présentation des documents et sur leur typographie. La présentation et la typographie peuvent ainsi être adaptées à différents styles de documents. Cependant certains choix typographiques exigent des connaissances linguistiques (compréhension du texte) que le modèle des documents structurés ne peut fournir. C'est donc surtout des résultats des recherches concernant l'analyse des langues naturelles que dépendent les perspectives de la correction typographique des documents électroniques.

Il n'en reste pas moins que l'édition électronique et la PAO ont encore beaucoup à apprendre des typographes et correcteurs professionnels. Ces professionnels ont passé des décennies à améliorer les techniques de l'imprimerie et accumulé expérience et savoir-faire pour imprimer des documents d'excellente qualité. Les utilisateurs d'écrans sont encore loin d'avoir acquis une aussi longue expérience. Et, pour la plupart, ils se soucient peu de qualité, aussi bien pour les impressions papier que pour les images affichées sur les écrans (lisibilité des caractères, présentation des textes et des illustrations, utilisation des couleurs, ...). Ceci explique aussi la médiocre qualité des documents multimédia qui commencent à être diffusés sur le réseau Internet. Mais, avec le développement du multimédia, l'utilisation

des technologies « écran », est promise à un bel avenir. La question de la qualité typographique de tels documents se pose donc avec une particulière acuité.

Syntaxe du langage de typographie Le langage de typographie TY est utilisé pour l'écriture de modèles de typographie qui définissent les règles de typographie à appliquer aux différentes classes de documents et d'objets. Ce langage est décrit ici en utilisant le méta-langage M de Grif [41].

```
SchemaTypo = 'TYPOGRAPHY' IdentType ';'
 [ 'ALINEA' SuiteType]
 [ 'FUNCTION' SuiteFunct ]
 [ 'COMPOSITION' SuiteMod ]
 [ 'ELEMENT' SuiteTypo ]
 [ 'ATTR' SuiteTypoAttr ]
 'END' .
IdentType = NAME .
SuiteType = IdentType < ',' IdentType > ';' .
IdentType = NAME .
SuiteFunct = Funct < Funct > .
Funct = TypeFunct ':' IdentFunct ';'
TypeFunct = 'Abbreviation' / 'Attribute' / 'Capitalization' / 'Distance
'Exponent' / 'InsertPair' / 'Punctuation' / 'Word' .
RegleTypo = [ 'IF' '(' SuiteCondition ')' ] SuiteParam ';' .
SuiteParam = 'BEGIN' ParamTypo < ParamTypo > 'END' ';' / ParamTypo .
ParamTypo = TypeFunct ':' IdFunct / IdentMod .
IdFunct = IdentFunct | 'FunctNil' .
SuiteCondition = Condition < 'AND' Condition > .
Condition = [ 'NOT' ] ConditionElem .
ConditionElem = 'First' /
'Last' /
  [ RelType ] OpType IdentType /
 'IN' IdentType /
 'LANGUAGE' '=' Language /
'IS' LexicalUnit .
OpType = 'IN' / '='.
RelType = 'Next' / 'Previous' .
```

```
LexicalUnit = NAME .
Language = NAME .
SuiteMod = Mod < Mod > .
Mod = IdentMod '=' SuiteRegles ';' .
IdentMod = NAME .
SuiteRegles = 'BEGIN' RegleTypo < RegleTypo > 'END' / RegleTypo .
SuiteTypo = Typo < Typo > .
Typo = IdentType ':' SuiteRegles .
SuiteTypoAttr = TransAttr < TransAttr > .
TransAttr = AttrIdent [ '(' IdentType ')' ] [ AttrRelat ] ':' SuiteRegle
AttrIdent = NAME .
AttrRelat = '=' AttrValue / '>' [ '-' ] MinVal /
'<' [ '-' ] MaxVal / 'IN' '[' ['-'] MinInterval '.' '.' ['-'] MaxInterva
']' . | '>' ['-'] ValeurMin | '<' ['-'] ValeurMax | 'IN' '[' ['-'] Borne
'..' ['-'] BorneSup ']'.
AttrValue = [ '-' ] ValEqual / TextEqual / AttrValIdent .
MinVal =      NUMBER .
MaxVal =        NUMBER .
MinInterval =   NUMBER .
MaxInterval =   NUMBER .
ValEqual =      NUMBER .
TextEqual =     STRING .
AttrValIdent =  NAME .
IdentType =     NAME .
```

Schéma de typographie La vérification typographique d'un document de la classe Article s'appuie sur plusieurs schémas de typographie : le schéma de typographie ArticleTY, associé à la structure Article et le schéma de typographie ParagraphTY, associé à la structure Paragraph. Si un document de la classe Article utilise des références bibliographiques, le vérificateur utilisera en plus un schéma de typographie décrivant les règles de typographie de la bibliographie.

Dans cette annexe, nous ne présentons qu'un seul schéma de typographie, celui d'un Article (section B de cette annexe). Ce schéma utilise une vingtaine de fonctions de contrôle dont le rôle est résumé ci-dessous.

A Fonctions de contrôle

Capitalization Règle appliquée aux alinéas :

- *CapFirstElem* vérifie que le premier caractère du premier mot de l'alinéa porte une capitale initiale ;
- *CapitAll* vérifie que tout le texte contenu dans l'alinéa est en capitale (sauf les textes des exposants ou indices) ;
- *CapAllWord* vérifie que chaque mot « significatif » porte une capitale initiale.

Distance

- *DistAlineaNone* vérifie que l'alinéa ne commence et ne se termine par pas par un ou plusieurs espaces.
- *DistNextHard* vérifie que le dernier caractère de l'élément est un blanc dur.
- *DistNextNone* vérifie qu'il n'y a pas d'espace en fin de cet élément

Exponent

- *ExpNone* vérifie que l'élément ne contient pas d'exposants.

InsertPair

- *PairNone* vérifie que l'élément ne contient pas de caractères de parenthésage.
- *PairPar* vérifie tout d'abord que les seuls caractères de parenthésage utilisés sont les parenthèses (ouvrante et fermante) et ensuite que ces parenthèses sont correctement utilisées (parité, imbrication) et présentées (en romain dans un texte en romain, en italique si le contenu est en italique, etc.)

Punctuation Toutes les fonctions de vérification de la ponctuation s'appuient sur la table d'espacement (généralement spécifique à la langue) définie dans la règle SpaceTable.

Règle appliquée aux alinéas :

- *PunctNone* vérifie que l'alinéa ne se termine pas par une ponctuation,
- *PunctEnd* vérifie que l'alinéa se termine par une ponctuation finale (. ! ? ...),
- *PunctHighNone* vérifie que l'alinéa ne se termine pas par une ponctuation ou exclusivement par ! ou ? (un point abréviatif n'est pas considéré comme une ponctuation finale),
- *PunctPeriod* vérifie que l'alinéa se termine par un point,

SpaceTable Cette règle n'effectue pas de vérification. Elle indique seulement quelle est la table d'espacement qui doit être utilisée par les autres règles, en particulier celles qui concernent les ponctuations, les parenthèses et les espaces : SpaceFrench pour les textes écrits en français, SpaceEnglish pour les textes écrits en anglais. Son contenu précise non seulement quel type d'espace peut ou doit précéder ou suivre chaque caractère, mais aussi quels sont les caractères qui peuvent ou ne doivent pas être consécutifs.

Par exemple, un point ne doit jamais être suivi d'une lettre (sans espace intermédiaire), des parenthèses ne doivent pas être consécutives, un signe de ponctuation ne doit jamais suivre un caractère ouvrant, etc.

Word Ces règles s'appliquent aux mots ou groupes de mots reconnus comme des unités lexicales (par leur appartenance à un lexique) :

- *WordFirstCap* vérifie que le mot porte une capitale initiale,
- *WordSmallCap* vérifie que le mot est en petites capitales.

B Modèle de typographie

Ci-dessous un exemple de modèle de typographie pour un document de style Article, utilisant les fonctions de contrôle décrites ci-dessus.

```
{\bf TYPOGRAPHY} Article
{\bf ALINEA}
Affiliation, Titre, Titre_section,
 Titre_sous_section, Titre_ssous_section,
  Titre_Annexe, Titre_figure, Mots_clés,
 Date_de_mise_a_jour;
{\bf FUNCTION}
 Abbreviation: AbbNone;
 Capitalization: CapFirstElem;
Capitalization: CapAllWord;
 Distance: DistAlineaNone;
 Distance: DistNextHard;
 InsertPair: PairNone;
 InsertPair: PairPar;
 Punctuation: PunctNone;
 Punctuation: PunctHighNone;
 Punctuation: PunctPeriod;
 SpaceTable: SpaceFrench;
SpaceTable: SpaceEnglish;
 Word: WordFirstCap;
 Word: WordAllCap;
 Word: WordNameCap;
{\bf COMPOSITION}
ModTitle =
BEGIN
InsertPair: PairNone;
Punctuation: PunctHighNone;
 IF (Language = Français)
Capitalization: CapFirstElem;
  IF (Language = English)
Capitalization: CapAllWord;
```

```
 END;
ModSubTitle =
BEGIN
Capitalization: CapAllWord;
InsertPair: PairPar;
Punctuation: PunctHighNone;
 END;
 ModSubSubTitle =
BEGIN
 Capitalization: CapFirstElem;
Punctuation: PunctHighNone;
END;
{\bf ELEMENT}
Article:
BEGIN
Abbreviation: AbbNone;
 Capitalization: CapFirstElem;
Distance: DistAlineaNone;
 IF (Next = Ref_biblio) Distance: DistNextHard;
 IF (Next = Ref_note) Distance: DistNextNone;
  InsertPair: PairNone;
Punctuation: PunctNone;
  IF (Language = English) SpaceTable: SpaceEnglish;
  IF (Language = Français) SpaceTable: SpaceFrench;
 IF (IS name) Word: WordFirstCap;
END;
Titre:
 ModTitle;
Titre_Annexe:
 ModTitle;
Titre_figure:
 ModSubSubTitle;
Titre_section:
 ModSubTitle;
Titre_sous_section:
```

```
 ModSubTitle;
Titre_ssous_section:
ModSubSubTitle;
Mots_clés:
 BEGIN
 Capitalization: CapFirstElem;
Punctuation: PunctPeriod;
 END;
Statut:
 BEGIN
 Abbreviation: FunctNil;
 Capitalization: FunctNil;
 Distance: FunctNil;
 InsertPair: FunctNil;
 Punctuation: FunctNil;
SpaceTable: FunctNil;
 Word: FunctNil;
 END;
{\bf END}
```

Références

[1] *Abrégé du code typographique à l'usage de la presse*, éditions du CFPJ (centre de perfectionnement des journalistes), Paris, 1989.

[2] *Code typographique*, Édité par la fédération nationale du personnel d'encadrement des industries polygraphiques et de la communication, Paris, 1981.

[3] *Code typographique à l'usage des auteurs*, Éditions du Seuil, Paris, 1975.

[4] *Guide du typographe romand*, Édité par le Groupe de Lausanne de l'Association suisse des compositeurs à la machine, Suisse, 1994.

[5] *Lexique des règles typographiques en usage à l'imprimerie nationale*, Imprimerie nationale, France, octobre 1990.

[6] *The Chicago Manual of Style,* vol. Fourteenth Edition, , The University of Chicago Press, Chicago, 1993.

[7] Jacques André et Olivier Nicole, “Les Règles typographiques de l’Imprimerie nationale (analyse)”, *Cahiers GUTenberg,* vol. 8, , pp. 51-56, mars 1991.

[8] Jacques André, “Titres : à lire, à voir ou à dire ?”, *Cahiers GUTenberg,* vol. 15, , pp. 33-42, avril 1993.

[9] Jacques André, “Le Guide du typographe romand (analyse)”, *La Lettre GUTenberg,* vol. 3, , pp. 20-22, juin 1994.

[10] Jacques Anis, *L’écriture, théories et descriptions,* De Boeck, Bruxelles, 1988.

[11] Fernand Baudin, “La typographie « entre guillemets »”, *Cahier GUTenberg,* vol. 7, , pp. 33-34, novembre 1990.

[12] Fernand Baudin, “How to Use Ready-Made Alphabets”, *Visual and technical aspects of type,* Roger D. Hersch, ed., pp. 165-196, Cambridge University Press, 1993.

[13] Fernand Baudin, *L’effet Gutenberg,* Éditions du Cercle de la Librairie, Paris, 1994.

[14] Abdel Belaïd et Yolande Belaïd, *Reconnaissance des formes – méthodes et applications,* InterEditions, 1992.

[15] Anne Marie Christin, “Dire, voir, écrire : le texte et l’image”, *Textuel,* num. 6, 1979.

[16] Jacques Drillon, *Traité de la ponctuation française,* Gallimard, 1991.

[17] H. van Dyke-Parunak, “Oral typesetting: some uses of biblical structures”, *Biblica,* vol. 62 , num. 2, pp. 153-168, 1981.

[18] André Eskénazi, “De quelques sigles”, *Lynx (numéro spécial ššLes sigles”),* vol. 30, , pp. 33-90.

[19] D. Flipo, B. Gaulle, K. Vancauwenberghe, “Motifs français de césure typographique”, *Cahiers GUTenberg,* vol. 18, , pp. 35-60, septembre 1994.

[20] Jean Girodet, *Pièges et difficultés de la langue française,* Dictionnaires Bordas, 1988.

[21] C. Gouriou, *Mémento typographique,* Hachette, Paris, 1973 (nouvelle version parue en 1993).

[22] Maurice Grévisse, *Le bon usage – grammaire française,* Éditions Duculot, Gembloux, 1988.

[23] Roger Hersch (éd.), *Visual and Technical Aspects of Types,* Cambridge University Press, Cambridge (GB), 1993.

[24] Lydia Hirschberg, "Lois formelles de la ponctuation", *Linguistics,* num. 19, pp. 21-63, 1965.

[25] I.S.O., *Information processing - Text and office systems - Standard Generalized Markup Language (SGML),* vol. ISO , 8879, octobre 1986.

[26] I.S.O., *Information processing - Text and office systems - Office Document Architecture (ODA),* vol. ISO 8613, , 1989.

[27] I.S.O., *Information processing - Text and office systems - Office Document Architecture (DSSSL),* vol. ISO/IEC DIS 10179, , 1991.

[28] R. Jaquenod, *La ponctuation maîtrisée,* Marabout savoirs, octobre 1993.

[29] Donald E. Knuth, *The TEX book,* Assison-Wesley, Reading, Massachusetts, 1984.

[30] Leslie Lamport, LATEX, *A Document Preparation System,* Addison-Wesley, Reading, Massachusetts, 1994 (nouvelle édition revue et corrigée).

[31] Franklin Mark Liang, *Word Hy-phen-a-tion by computer,* Ph. D., Standford University, Juin 1983 et Report STAN-CS-83-977.

[32] Gordon Legge, "Psychophysics of reading", *Vision research,* pp. 239-252, 1985.

[33] Alan Marshall, "La nature du travail typo", *Caractère : l'événement technique,* vol. supplément au n. 338, , pp. 198- 201, 8 septembre 1992.

[34] Ladislas Mandel, “L’écriture typographique : vers une prise de conscience”, *Communication et langages,* vol. 77, , pp. 5-30, 1988.

[35] Robert A. Morris, Karl Berry, Kathryn Hargreaves, *Towards quantification of the effects of typographic variation on readibility,* num. Technical Report 92-6, University of Massachusetts, Boston, 1992.

[36] Pat Norrish, “Semantic structures of text”, *Structured documents,* J. André, R. Furuta et V. Quint, ed., pp. 143-159, Cambridge University Press, 1989.

[37] J. Kevin O’Regan et Levie-Schoen, “Le concept, pluriel, de lisibilité”, *Communication et langages,* 1985.

[38] John K. Ousterhout, “Tcl: An Embeddable Command Language”, *USENIX Conference,* hiver 1990.

[39] E. Pascual, *Représentation de l’architecture textuelle et génération texte,* Thèse, Université de Toulouse, 1991.

[40] Vincent Quint, Irène Vatton, “Grif: an Interactive System for Structured Document manipulation”, *Text Processing and Document Manipulation,* J. C. van Vliet, ed., pp. 200-213, Cambridge University Press, 1986.

[41] Vincent Quint, Irène Vatton, *Les langages de Grif,* Note technique, Imag, 2 rue de Vignate, 38610 Gières (France), 1994 (ftp://ftp.imag.fr/pub/OPERA/doc/langages.ps.gz).

[42] François Richaudeau, *Manuel de typographie et de mise en page,* Édition Retz, Paris, 1989.

[43] Hélène Richy, Patrice Frison, Éric Picheral, “Multilingual String-to-String Correction in Grif, a structured editor”, *EP’92,* C. Vanoirbek et G. Coray, ed., pp. 183-198, Cambridge University Press, 1992.

[44] Miles Tinker, *Legitibility of Print,* Iowa State University Press, 1963.

[45] Jacques Virbel, “The Contribution of Linguistic Knowledge to the Interpretation of Text Structures”, *Structured Documents,* J.André, R. Furuta et V. Quint, ed., pp. 161-180, Cambridge University Press, 1989.

[46] Jacques Virbel, “ Formalisation d’une classe de relations structurelles de textes”, *Bigre 80 (actes du colloque CNED’92)*, pp. 192-199, juillet 1992.

Projet OPÉRA
p u b l i c a t i o n s

Une liste mise à jour de labibliographie du projet Opéra se trouve dans :
An updated list of Opéra team's references is in:

`http://www-bi.imag.fr/OPERA/BibOpera.html`

Unité de recherche INRIA Lorraine, Technopôle de Nancy-Brabois, Campus scientifique,
615 rue du Jardin Botanique, BP 101, 54600 VILLERS LÈS NANCY
Unité de recherche INRIA Rennes, Irisa, Campus universitaire de Beaulieu, 35042 RENNES Cedex
Unité de recherche INRIA Rhône-Alpes, 46 avenue Félix Viallet, 38031 GRENOBLE Cedex 1
Unité de recherche INRIA Rocquencourt, Domaine de Voluceau, Rocquencourt, BP 105, 78153 LE CHESNAY Cedex
Unité de recherche INRIA Sophia-Antipolis, 2004 route des Lucioles, BP 93, 06902 SOPHIA-ANTIPOLIS Cedex

Éditeur

INRIA, Domaine de Voluceau, Rocquencourt, BP 105, 78153 LE CHESNAY Cedex (France)

ISSN 0249-6399