

Désambiguïsation

Laurent Lyaudet*

29 août 2023

Version initiale : 2023/08/26 Version courante : 2023/08/29

Cet article est à la fois un article de science (facile à comprendre) et un article de religion. Les deux sont entremêlées. Le lecteur intéressé uniquement par la science restera sans doute insensible à la fin de l'article ; tout comme un lecteur peu intéressé par la science trouvera sans doute anecdotique le début. Pourtant les deux comptent et s'entraident. J'espère que certains lecteurs le verront.

J'ai eu la chance quand j'étais petit, à l'âge où toutes les conventions semblent arbitraires, que mon Papa de la terre m'ait expliqué que l'un des buts de l'orthographe est de désambiguïser le langage. Ce n'est pas le seul but, il y a aussi des aides et des rappels à l'étymologie, et d'autres buts sans doute que j'ignore ou que j'ai oubliés. Mais c'est tout sauf arbitraire. Et du haut de mes 7 ans environ (ma mémoire n'est pas assez précise pour que je me souvienne de mon âge avec certitude), je m'étais dit que c'était une bonne chose d'enlever les ambiguïtés quand on s'exprime.

Le fruit immédiat a été de me faire accepter et aimer l'orthographe. Mais depuis quelques années l'arbre s'est développé, tout comme ses fruits.

Depuis de nombreuses années, comme la plupart des scientifiques et des personnes qui doivent travailler avec de l'informatique et des nombres, je suis embêté par les conventions de séparateur décimal. La convention anglaise me semble trop bizarre. Le point est censé marquer la fin de la phrase, pas une séparation décimale. La convention française me semblait naturelle, puisque c'est celle que j'ai apprise à l'école et que la virgule introduit une suite dans une énumération ou une séparation peu marquée. En plus, la norme ISO est pour la virgule, raison de plus pour me sentir chauvin et **suffisant** ;) XD.

Mais du côté de la désambiguïsation, quelque chose me dérangeait. Réutiliser le même symbole typographique ultra fréquent et introduire des branchements conditionnels dans mon cerveau qui doit analyser ce que je lis, ce n'est bon ni pour la clarté, ni pour l'efficacité. Alors pour expliquer un peu de science de manière simplifiée, un branchement conditionnel, c'est, dans du code informatique, quand on a des instructions

```
SI telle condition est vraie
ALORS faire cela
SINON SI telle autre condition est vraie
ALORS faire ceci
```

*<https://lyaudet.eu/laurent/>, laurent.lyaudet@gmail.com

SINON faire autre chose

Imagine que ton cerveau lit une phrase et qu'à chaque symbole il perde du temps à se dire « Alors, en fonction du contexte avant et après, ce symbole-ci veut dire ça. ». Je ne détaillerai pas ici la différence entre les grammaires contextuelles et non-contextuelles en théorie des langages formels, une sous-discipline de l'informatique. Par contre, je vais parler un peu du langage machine avec sa version en assembleur à peine plus différente qu'une simple translittération (pas grave si plusieurs mots dans cette phrase n'ont pas encore de sens pour toi, tu as tout autant de chance de comprendre le propos d'ensemble de cet article, et si tu es curieux, Wikipédia est ton amie :)). Imagine que le processeur qui fait les calculs dans ton ordinateur est en réalité une taupe (oui tout le monde est surveillé;) XD) un peu myope, comme dans les histoires pour enfant. Il range toutes ses instructions dans des boîtes à chaussures, il en a partout, dans l'entrée, le salon, la cuisine, la salle de bain, la chambre et même les toilettes. Mais cette taupe myope est très méthodique, et elle a un ordre total sur ses boîtes à chaussures. Elle lit la numéro 1, fait ce qui est marqué dedans, puis la numéro 2, etc. SAUF quand l'instruction dans une boîte à chaussures, un branchement, lui dit va tout de suite chercher la boîte numéro 111 qui se trouve à l'autre bout du terrier-appartement. Du coup, elle perd du temps à y aller, car elle est un peu lente. Eh bien, pour un processeur d'ordinateur, c'est exactement pareil. Un branchement lui fait perdre du temps, surtout s'il est conditionnel. En gros si l'instruction dans la boîte lui dit toujours de chercher la boîte numéro 111, ce n'est pas conditionnel et la taupe peut demander à un assistant de lire les boîtes avant elle, juste pour repérer les branchements, et d'aller lui chercher les boîtes en avance. S'il y a trop de branchements, l'assistant est débordé, logique; de même, si le numéro de boîte à aller chercher dépend des calculs faits par la taupe, l'assistant utilise des heuristiques pour aller chercher des boîtes. (Une heuristique c'est une procédure qui a raté son examen pour passer algorithmique;). Elle a juste le label « Ça marchera p't'être, ça marchera p't'être pas. ». C'est comme de la manutention, c'est du déplacement de données dans un processeur. Mais c'est la même logique qui fait que quand tu déménages, tu mets tout dans un camion, plutôt que de faire 1000 allers-retours entre l'ancien logement et le nouveau.

Du coup, je ne sais plus trop comment, j'ai commencé à utiliser à l'écrit (manuel) une virgule à l'envers comme séparateur décimal. Je ne sais plus si ça date d'avant ou d'après que j'apprenne que c'est ce que font les Arabes. Je sais juste que les premières traces de cette utilisation dans mes cahiers date de 2020. Mais ils écrivent de la gauche vers la droite contrairement à nous. En fait, mon but c'est d'avoir un séparateur décimal différent de la virgule « texte » qui marque une pause ou une énumération. Si les Arabes ont une virgule « texte » identique à leur séparateur décimal, ils devraient aussi changer pour utiliser la nôtre;) XD. Entre 2020 et 2021, je crois, j'ai une discussion avec Étienne Lemaire; et il me dit que, oui, la virgule comme séparateur décimal l'embête bien car en physique, il note souvent des énumérations de nombres avec des décimales, par exemple
123,11, 123,22, 123,33
pas vraiment mieux que
123.11, 123.22, 123.33.
ou avec des numéros de version en informatique

Firefox 123.2.3.

Mais quand ça s'arrête ?! XD Comme tu peux le voir, il faut faire attention aux espaces, et si tu as l'habitude d'encore écrire à la main, tu sais que les variations de largeur d'espacement sont très fréquentes et que c'est peu fiable. Comment est-ce que tu trouves la lisibilité de ce qui suit ?

123,11, 123,22, 123,33.

Ce n'est pas le caractère arabe que je n'arrive pas à reproduire sur mon vieil ordinateur et sa version de LaTeX. Si tu connais ce logiciel, voilà la commande que j'utilise :

```
\newcommand\sepdecLL{\scalebox{-1}[1]{,}}
```

C'est plus lisible à la main, où nos virgules sont plus marquées que le symbole typographique.

Pour les séparateurs de milliers, je trouve que la convention anglaise avec virgule, ou française avec espace n'est pas la meilleure. Je lui préfère la convention de la plupart des langages de programmation qui ont des séparateurs de nombres littéraux, à savoir le tiret bas, 12.456. (Supplément technique : De plus, j'y vois un moyen d'écrire facilement des nombres en base arbitraire. 4 en base 3 : Oba3.11, 111 en base 12 : Oba12.9.3 Seine Saint-Denis ?;) XD, 118 en base 12 : Oba12.9.10, avec des bases variables Oba3.11_ba5.4 = 4 + 5 + 15 = 24, avec de l'espacement supplémentaire Oba12.1.2.2.11 = 11 + 24 + 288 + 1728 = 2055.)

Toujours dans la même discussion avec Étienne Lemaire, je lui dis que, du point de vue de la désambiguïsation, il y a autre chose qui m'embête c'est le point utilisé en fin d'abréviation. En fait, je pensais que c'était un peu un côté psycho-rigide de ma part. Jusqu'à ce que je voie, dans mon environnement de développement intégré (PyCharm), qu'il ne comprenait pas la structure d'une phrase et m'indiquait une fausse erreur typographique à cause du point d'abréviation. Du coup, c'est du virtuel qui m'a fait réaliser que le problème n'était pas juste une lubie de ma part. J'y ai réfléchi et j'ai fini par me dire qu'une abréviation, c'est comme un mot qui tombe à l'eau, un peu comme une virgule inversée dans le sens de la hauteur. Alors la convention anglaise utilise peu le point d'abréviation mais uniquement des majuscules, et le monde du commerce puis le reste de la société occidentale fait pareil depuis quelques décennies. Exemples : HTML et SNCF, plutôt que H.T.M.L. et S.N.C.F. Mais là encore, je préfère H.T.M.L. et S.N.C.F. Voilà la commande que j'utilise :

```
\newcommand\virabrLL{\raisebox{0.2ex}{\scalebox{1}[-1]{,}}}
```

Toujours sur le sujet des abréviations et des conventions typographiques, je te recommande la lecture du « Lexique des règles typographiques en vigueur à l'Imprimerie Nationale ». Je me dis, il y a un autre truc qui m'embête, c'est l'abréviation intermédiaire avec la fin du mot en petite casse sur-élevée. Par exemple, M^{me}, qui peut-être confondue avec une expression mathématique où M est un majorant, m un minorant et e la base du logarithme népérien, ce qui devient « grand M à la puissance petit m fois e ». Encore une ambiguïté, donc une convention à modifier. De même, la règle devient un peu floue avec Monseigneur abrégé M^{gr}, alors que le g est une lettre intermédiaire et non finale. Du coup, cette fois-ci, je me suis dit, il faudrait une sorte de pont d'abréviation, M^{me} et M^{gr}. Voilà la commande que j'utilise :

```
\newcommand\ponabrLL{\raisebox{-0.2ex}{\scalebox{0.75}[0.75]{\frown$}}}
```

Et ça irait bien avec des abréviations communes en informatique où on met un nombre comptant les lettres intermédiaires, par exemple A11y et I18n, pour Accessibility et Internationalisation. On pourrait imaginer ce nombre, qui aura au plus deux chiffres normalement, en petite casse sur-élevée au dessus du pont d'abréviation.

En fait, sans doute à cause de mon travail sur le principe de première différence et les documentaires à la télévision, je me dis que toutes les conventions que je propose ci-dessus et ci-dessous sont faciles à faire avec un calame en écriture cunéiforme comme les Sumériens. Je me vois en train de graver des tablettes. J'ai une pensée matin et soir pour Noé. Mon Papa de la terre, qui est un peu bizarre parfois avec ses livres d'alchimie ou d'histoires acadabrantiques, me parlait d'une civilisation plus avancée que la nôtre pendant le précédent âge inter-glaciaire. Un peu comme le mythique âge d'or des grecs. Certaines histoires sont fantasmées, mais si tu crois la Bible, certaines sont véridiques. J'ai cette conviction qu'en remontant le courant, j'ai bien des grands frères et des grandes sœurs qui me précèdent et me dépassent, Hénok, etc., tant sur le plan spirituel que matériel. Je me suis dit qu'une des tendances humaines est la paresse, plus court toujours plus court dans le sens du courant. Et ainsi, M_{me} peut devenir M^{me} , puis M^{me}. Si tu peux remonter le courant, tu peux le descendre. Qui peut le plus, peut le moins. L'exemple qui précède est une supposition, mais cela te montre ce que tu peux essayer de comprendre avec tes raisonnements, si tu te souviens que les conventions humaines changent. De plus, peu importe le lieu ou le temps, l'Esprit-Saint y était et peut t'instruire, mais focalise-toi sur ce qu'il y a de bon à savoir.

Toujours sur ce problème d'abréviation, celle des nombres ordinaux : le premier, le deuxième, le troisième, etc., il y a une ambiguïté avec les mathématiques 1^{er} ça donne 1, passons, mais 2^e c'est une constante non triviale. Je lis dans le « Lexique des règles typographiques en vigueur à l'Imprimerie Nationale » que les Allemands utilisent le point pour ça. Je me dis, ils ont raison. C'est une distinction de base entre cardinal et ordinal. La notation des nombres devrait le gérer indépendamment de la langue. (Après les Arabes, les Allemands, il est vraiment bizarre ce gus, et j'aime pas les gens bizarre;) XD.) Pourquoi pas utiliser une virgule inversée horizontalement et verticalement ? Donc au lieu de

1^{er}, 2^e, 3^e (français),

1st, 2nd, 3rd (anglais),

1^o, 2^o, 3^o (latin),

1., 2., 3. (allemand).

on peut utiliser

1., 2., 3.

Voilà la commande que j'utilise :

```
\newcommand\ordLL{\raisebox{0.2ex}{\scalebox{-1}{[-1]{,}}}}
```

Mais j'aurais préféré visuellement pour l'idée d'élévation, une sorte de rampe, un peu comme une virgule sur-élevée,

1., 2., 3.

Voilà la commande correspondante :

```
\raisebox{0.2ex}{,}
```

Il faudrait faire cette virgule sur-élevée plus ouverte comme à l'écrit manuel. Le problème c'est qu'à l'écrit manuel une légère différence de hauteur est un piège majeur.

Voilà des symboles typographiques simples et tous ont pu devenir des points ou des virgules, avec le courant qui uniformise et simplifie au lieu de désambiguïser. Je

considère toutes ces suggestions comme très sérieuses, la graphie est à parfaire, il faut étudier les problèmes d'accessibilité, etc. (Attention à la cohabitation avec les sérifs de certaines polices de caractères.) Néanmoins, il ne faut vraiment pas partir du principe que nos conventions humaines sont les meilleures, par chauvinisme, « humanisme », paresse, etc. C'est une parole de libération si tu réalises que tu peux améliorer y compris des éléments de base. Ne pars pas du principe que, parce que pour une partie des points ci-dessus, ce sont des mathématiques, ce sont des gens sérieux, etc., ce serait parfait. Tu connais le proverbe « Les cordonniers sont toujours les plus mal chaussés. », ça s'applique beaucoup à la typographie. Une autre parabole que j'ai inventée « Est-ce qu'il y a des gaspillages d'eau dans l'agriculture ? Des raccords qui fuient, mal serrés. Rien à voir avec une fuite domestique en termes de volume d'eau. (Je l'ai vu de mes yeux dans l'une des deux serres où j'ai travaillé. Le contre-maître m'a dit que c'était un léger gaspillage quand je lui en ai parlé. Il s'en moquait.) Et pourtant, l'eau est centrale pour l'agriculture. ». De même, les outils pour réfléchir, nos langages communs ou nos jargons techniques nous poussent souvent à gaspiller nos ressources intellectuelles sans le savoir.

Le langage a deux utilités : servir notre communication et servir notre raisonnement. Il faut marcher sur deux jambes, comme le dit le proverbe africain « Seul on va plus vite, ensemble on va plus loin. ». Tu peux avancer seul et, malheureusement, tu devras avancer seul avec Dieu sur certains sujets. Mais, de temps en temps, reviens en arrière, descends le courant et montre ce que tu as trouvé aux autres pour que tout le monde aille plus loin et que ton fruit demeure.

La vraie règle qui s'applique sur nos conventions, c'est « **suffisamment** bon », en anglais « good enough ». Ne te l'applique pas à toi-même, ne fais pas du sur-place. Car sinon, certains ajouteront une virgule « good, enough » et là il y a deux sens, tu peux mal tomber. En vérité, j'essaye de m'appliquer ce conseil à moi-même. Mais clairement, je ne fais pas un sans-faute. Parfois je suis un peu Pharisien (presque comme Parisien, puisque j'y suis né ;)). Mais malgré ça, comme l'a dit le Christ dans l'Évangile selon Saint-Matthieu « Faites donc et observez tout ce qu'ils pourront vous dire, mais ne vous réglez pas sur leurs actes : car ils disent et ne font pas. ». Comme quasi tout le monde, je n'en fais pas **suffisamment**, mais j'essaye.

À un moment dans mon esprit, le principe de désambiguïsation s'est relié à la parole d'Isaïe chapitre 40 versets 3 à 5 « Une voix crie : « Dans le désert, frayez le chemin de Yahvé ; dans la steppe, aplanissez une route pour notre Dieu. Que toute vallée soit comblée, toute montagne et toute colline abaissées, que les lieux accidentés se changent en plaine et les escarpements en large vallée ; alors la gloire de Yahvé se révélera et toute chair, d'un coup, la verra, car la bouche de Yahvé a parlé. ». ». Cette parole est spirituelle, et elle a d'abord une portée spirituelle, sur l'état de nos cœurs, mais elle a aussi porté du fruit dans ma vie à un niveau technique pour des symboles typographiques. Ce sont des doigts pas forcément très beaux qui montrent la lune de la désambiguïsation.

Merci Dieu ! Merci Père ! Merci Seigneur ! Merci Saint-Esprit ! Merci Isaïe ! Merci Hénok ! Merci Noé ! Merci Papa de la terre ! Merci mes Parents pour les corrections de typos !

P.-S. : Un ami m'a fait la blague de la taille des claviers d'ordinateurs qu'il faudrait avoir :). Du coup, ça tombe bien, j'ai plein d'idées concernant nos claviers d'ordi-

nateurs préhistoriques, moins évolués que les claviers d'orgues qui servent tous les dimanches :). Premièrement, un clavier n'a pas besoin d'avoir un nombre de touches qui augmente linéairement avec le nombre de caractères. Avec des combinaisons de touches, on peut obtenir un nombre exponentiel de caractères par rapport au nombre de touches. Le problème c'est d'arriver à appuyer sur plusieurs touches en même temps. Pour ça, la solution ergonomique est de pouvoir maintenir une touche dans un état verrouillé, comme le verrou des majuscules ou du pavé numérique. Une solution peut être d'utiliser des boutons ronds comme ceux des potentiomètres sur les amplis en musique. Un seul de ces boutons ronds peut sélectionner facilement 10 à 20 choix. On devrait déjà avoir des claviers qui gèrent l'ensemble des caractères Unicode, si on s'en donnait la peine (je ne parle pas des séquences de touches qui t'obligent à mémoriser le botin et qui n'affichent pas ce que tu tapes avant d'appuyer). L'Unicode est une norme de caractères qui tente d'unir tous les symboles de toutes les langues de la terre. Ils ont à peu près réussi, mais, comme tous les comités qui finissent en comité Théodule, ils font traîner la fin du travail en refusant certaines icônes universelles (j'en ai croisées dans mon travail ; je n'ai malheureusement pas sauvegardé le lien de la décision de refus d'inclusion de l'icône dans l'Unicode ; quelqu'un de consciencieux devrait pouvoir trouver l'information sur Internet ; ce genre de décisions est peut-être pour garder du travail aux designers d'icônes payantes utiles dans les applications ; mais comparativement aux autres icônes présentes, c'est tout sauf dans l'intérêt de finir Unicode), et en perdant du temps sur 1000 nuances d'icônes « selfies » pour tous les nombrilistes du monde. Personnellement, je n'ai pas la peau jaune, mais dire bonjour aux autres avec un rond jaune qui insiste surtout sur son sourire, ça me va bien :). Mais bon, y'a des modes, et puis les bureaux sont chauffés ou climatisés, y'a une machine à café et des gens qu'on connaît bien avec qui discuter. Alors pourquoi terminer le travail ?;) XD Sinon, l'état courant d'Unicode est correct, pas parfait mais correct et mieux que le reste puisque, de fait, c'est la seule norme universelle et en plus c'est gratuit et ouvert. Elle groupe les caractères en 17 plans de $2^{16} = 65536$ caractères. La plupart des plans sont vides. On est encore loin d'avoir assez d'icônes et de caractères. Dans chaque plan, il y a des blocs plus ou moins grands, car les langues ou autres n'ont pas toujours le même nombre de caractères ou d'icônes. Sur un clavier standard, il y a environ 105 touches. La puissance de 2 la plus proche c'est 128. Imagine que ce soit un clavier avec des touches qui sont des mini-écrans (ça a déjà été fait) ou bien un clavier visuel sur écran où on peut cliquer à la souris ou un écran tactile ou un dispositif de pointage avec les yeux pour les paraplégiques. On peut voir un plan Unicode comme 256 pages de 256 caractères, (j'utilise le mot « page » fréquent en informatique, mais je ne crois pas que ce soit de la terminologie Unicode standard) ou, de manière équivalente, comme 512 demi-pages de 128 caractères, 1024 quarts de page de 64 caractères, etc., selon la taille de clavier que l'on veut obtenir. Donc, pour avoir un clavier Unicode, il suffirait d'une colonne de boutons de potentiomètres à gauche du clavier avec les nombres écrits sur chaque bouton (j'expliquerai les boutons ronds en mode clavier visuel après). Disons que chaque bouton permettrait de choisir de 0 à 9 soit dix choix, puisque l'on travaille en base 10. Une flèche à droite du bouton marquerait la valeur courante visible par la personne qui travaille sur le clavier. Il faudrait 2 boutons pour le choix du plan Unicode (unité, dizaine). On n'utiliserait que 17 choix sur 100, Les choix en plus pourraient servir à avoir les dispositions de clavier standard, ainsi que des palettes de caractères.

Comme un peintre prépare sa palette de couleurs, un utilisateur préparerait une ou des palettes de caractères présauvegardées. Pour des tâches courantes, même évoluées, 5 palettes de 128 caractères seraient largement suffisantes, je pense. Mais aucune raison de le limiter arbitrairement, car le coût additionnel est très faible, voire nul. En plus des 2 boutons pour le choix du plan Unicode, 3 boutons suffisent à choisir les pages ou les demi-pages du plan. Un clavier avec un peu plus d'une vingtaine de touches en plus et 5 boutons suffirait. Si tu veux gagner de la place, un 6. bouton et seulement 64 touches suffisent. Et pour faire marcher le bouton en mode visuel sur un écran, tu peux imaginer que, quand le pointeur de la souris ou un autre dispositif de pointage est dessus, le bouton tourne : lentement au bord du bouton, plus vite au centre du bouton ; je crois que ça existe déjà dans des jeux vidéos ou dans des outils d'accessibilité comme logique ; une recherche d'antériorité te permettra de le savoir. En tout cas, avec les composants électroniques, les micro-contrôleurs, et tout ce qui va avec en 2023, un prototype fonctionnel pourrait largement être fait par des étudiants ou étudiantes motivés lors d'un stage de 6 mois à 1 an, peut-être moins. J'ai l'idée du nombre logarithmique de touches pour Unicode depuis ma thèse je crois (2004-2007), par contre, pour l'idée des boutons de potentiomètres, ça doit dater de 2022 ou de 2023. Là aussi, il s'est passé du temps avant que les fruits poussent dans mon esprit. Vous voyez, pas besoin d'émoticône pour être nombriliste ;) XD. Peut-être qu'un jour les programmeurs utiliseront des symboles mathématiques courants, comme celui de la négation au lieu du point d'exclamation !