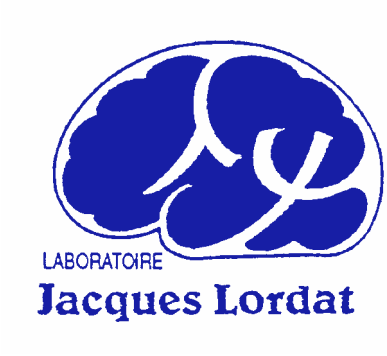


Projet « CALAME ON-LINE »

Prolégomènes à l'étude des erreurs en production écrite

Propositions en vue de la mise au point d'une grille d'analyse

J.-L. Bouraoui[(1)], Ph. Boissière[(1)], F. Vella[(1)], A. Lagarrigue[(1), (2)], M. Mojahid[(1)], D. Laur[(2)], N. Vigouroux[(1)], J.-L. Nespoulous[(2), (3)]

(1) IRIT (Institut de Recherche en Informatique de Toulouse)
UMR 5505 CNRS, INPT, UPS, UT1
(2) URI (Unité de Recherche Interdisciplinaire) OCTOGONE
Laboratoire Jacques Lordat, UPRES 4156
(3) Institut des Sciences du Cerveau de Toulouse
IFR 96

Rapport Interne n° IRIT/RR—2007-7-FR

Mai 2007

Note des auteurs :
Merci d'avance aux utilisateurs de la grille de nous faire part de leurs éventuelles remarques. Ces dernières nous permettront de la faire évoluer en vue de son optimisation.

Table des matières

Introduction

Errare humanum est. Cet adage se vérifie partout, y compris évidemment dans le domaine de la communication et l'expression linguistique. Dans le contexte du dialogue oral, il est fréquent que les erreurs soient repérées par ceux qui les commettent et/ou leur interlocuteur. Quand tel n'est pas le cas, les conséquences sur la transmission de l'information peuvent être importantes, voire graves. Il convient donc de tenter d'optimiser l'identification et l'interprétation des erreurs d'ordre linguistique, et de minimiser leurs conséquences au niveau de la communication interindividuelle. Il s'agit là d'une entreprise scientifique fondamentale puisqu'elle concerne les deux modalités majeures de communication langagière : l'oral et l'écrit. Toutefois, et même si ses soubassements théoriques et méthodologiques sont également valables pour la production orale, l'objectif spécifique de la présente tentative se focalise essentiellement sur la production écrite.

Le travail présenté dans ce document se situe dans une démarche d'ordre neuro-psycholinguistique. Il trouve son origine première dans les travaux menés, il y a plus de vingt ans, par l'un de nous (Jean-Luc Nespoulous, cf. notamment Nespoulous *et al.* 1982), en collaboration avec André Roch Lecours (Université de Montréal ; cf. A. R. Lecours *et al.* 1979, 1982) sur les productions écrites déviantes des patients aphasiques. Il est également complété, sur tel ou tel point, par certains éléments en provenance des travaux sur l'écriture de Nina Catach et de son équipe[1].

Le thème central du présent rapport est la proposition d'une grille d'annotation et d'interprétation des erreurs à l'écrit. L'objectif de cette grille est de rendre compte, de la manière la plus exhaustive possible, des erreurs survenues dans la production écrite, manuscrite ou en saisie « clavier », de personnes présentant divers types de handicaps « centraux » ou « périphériques ».

Plus spécifiquement, nous nous inscrivons dans le cadre du projet ESACIMC[2]. De ce fait, notre premier objectif est d'analyser les erreurs des sujets IMOC/IMC[3] en situation de saisie sur clavier (essentiellement logiciels) de messages écrits. Les corpus utilisés résultent de productions écrites dans ce contexte, même si les conclusions qui sont tirées sont généralisables, comme on le verra plus loin dans le présent rapport.

Dans un premier temps, nous présentons les principaux enjeux et problèmes liés à notre étude. Nous passons ensuite en revue les différentes catégories d'erreurs et de perturbations qui leur sont associées. Nous montrons notamment que lors de la saisie sur clavier, les erreurs peuvent avoir non seulement une motivation « linguistique » (phonétique, graphémique ou même morphémique), mais aussi « spatiale », compte tenu de la configuration du clavier. Enfin, nous décrivons d'un point de vue méthodologique les règles

[1] Cf. notamment (Catach 1980).
[2] Evaluation qualitative de Systèmes d'Aide à la Communication pour les Infirmes Moteurs Cérébraux (http://www.irit.fr/ESACIMC)
[3] Infirmes Moteurs d'Origine Cérébrale/ Infirmes Moteurs Cérébraux

qui ont présidé à l'annotation des erreurs dans un échantillon de 12 énoncés issu d'un corpus[4] particulier. L'objectif est ici de permettre à de futurs annotateurs n'ayant pas encore effectué cette tâche, de travailler sur les mêmes bases que celles du premier annotateur, bases à présent endossées par l'ensemble des auteurs du présent rapport. Il sera ainsi possible de juger de la fiabilité de ces règles et de leur indépendance par rapport à la personne qui effectue l'annotation/interprétation des erreurs.

[4] Corpus n° S7 84 02, recueilli par le centre de Kerpape (Morbihan) dans le cadre du projet ESACIMC financé par l'APETREIMC (Association Pour l'Education, la Thérapeutique et la Réadaptation des Enfants Infirmes Moteurs Cérébraux).

I) Analyse des erreurs
Réflexions théoriques et méthodologiques

De la nécessité (et de la difficulté) d'une analyse « multi-niveaux » des erreurs de la production écrite

Compte tenu de l'existence de différents niveaux d'organisation dans l'architecture structurale des langues naturelles, il apparaît indispensable de rendre compte de l'ensemble des erreurs susceptibles de survenir à chacun de ces niveaux : littéral, graphémique, lexical, morphologique, syntaxique… et portant sur des entités linguistiques allant de la « lettre » à la « phrase » et au « texte ».

Ceci étant, rares sont finalement les erreurs qui (a) ne se situent qu'à un seul niveau et (b) n'ont pas d'impact (même indirect) sur d'autres niveaux. Ainsi, telle omission « locale » d'une préposition entraîne inéluctablement l'agrammaticalité (syntaxique) de la phrase dans laquelle elle intervient (exemple : « Il a posé l'assiette **XXX** la table »). Pareillement, (autre exemple) une erreur qui pourrait n'être qu'orthographique (dans son déterminisme sous-jacent) peut entraîner, secondairement, une violation morphologique (exemple : « il mangeai**s** »).

Il convient par conséquent d'adopter une démarche « multi-niveaux », laquelle requiert, au moins dans un premier temps, d'octroyer à une même erreur « superficielle » plusieurs étiquettes (= annotations). Dans le premier exemple ci-dessus, nous serons ainsi amenés à étiqueter à un premier niveau, l'omission de morphème grammatical en tant que telle, avant d'ajouter une deuxième étiquette, au plan syntaxique cette fois (« agrammatisme »). Cela ne veut certes pas dire que le sujet a commis plusieurs erreurs. Cela veut simplement dire qu'en commettant une erreur « locale » à tel ou tel endroit du message, le sujet a entraîné plusieurs violations aux « conditions de bonne formation » des énoncés (ici écrits). En bref, ceci nous conduira à différencier, (a) des « erreurs locales à impact (simplement) local » et (b) des « erreurs locales à effets secondaires », ces dernières présentant, de toute évidence, un degré de gravité plus important, susceptible de perturber de façon massive l'échange d'informations entre l'émetteur et le récepteur.

Chacun aura bien compris que le point sur lequel nous venons d'insister complique passablement l'analyse dont il est ici question.

1. Le passage de l'oral à l'écrit : autre problème

S'il est certainement envisageable d'analyser les erreurs de la production orale sans tenir compte des représentations écrites des (séquences de) mots[5], il est clairement impossible et inenvisageable d'analyser l'écrit sans prendre en considération l'oral. L'écrit n'est en effet, – et particulièrement dans les langues à script alphabétique – qu'une transcription de l'oral, lequel a été acquis en premier (sauf dans quelques cas particuliers). Dès lors, bon nombre d'erreurs dans la production écrite se trouvent influencées (« contaminées ») par la nature des représentations orales, souvent « co-activées » au moment où le sujet entreprend sa tâche d'écriture !

Ces problèmes oral/écrit sont d'autant plus importants que les règles de conversion phonèmes/graphèmes sont opaques, comme c'est le cas en français ou en anglais, langues dans lesquelles un même phonème peut s'orthographier de *n* manières différentes, au grand dam de l'apprenti scripteur ! L'existence d'homophones non homographes constitue ainsi un problème majeur dont une grille d'analyse complète doit pouvoir rendre compte.

Dans le passage oral ➔ écrit, il conviendra de différencier, (a) les erreurs qui n'entraînent pas de changement de prononciation du mot écrit erroné (exemple : bateau ➔ * bato) et (b) celles qui modifieraient la prononciation du mot si on avait à le produire (exemple: boisson ➔ boison).

2. La chronométrie de la production écrite : une variable importante

Si la chronométrie de la production d'un message oral est aisée à réaliser : un magnétophone suffit, lequel enregistrera aussi bien les plages de production véritables que celles de silence (pauses et hésitations), il n'en va pas de même lorsqu'il s'agit d'analyser la production écrite. De ce fait, la plupart du temps, les erreurs sont analysées « off line », quelques instants (au mieux) après leur production. On mesure bien l'intérêt qu'il y a, depuis l'arrivée des tablettes graphiques, de prendre en considération de tels paramètres temporels. Ceux-ci sont de nature à permettre une analyse plus fine de la production écrite (exemple : arrêt du scripteur entre l'écriture du radical et celle de la désinence d'un verbe… indiquant vraisemblablement que le sujet « n'est pas à l'aise » dans sa gestion de la morphologie flexionnelle verbale !). Il en va, selon nous, de même dans l'écriture sur clavier, et ce, même si bon nombre d'individus utilisant un clavier ne sont pas des experts en dactylographie, (ce qui peut rendre l'interprétation des pauses plus difficile : temps de recherche de la bonne touche, difficulté de pointage dans le cadre de clavier logiciels[6]… indépendamment de ses connaissances linguistiques…).

Même si dans le présent travail, une étude « on line » n'est pas envisagée, il conviendra de prendre en considération, (a) les éventuelles mauvaises segmentations de mots, (b) les problèmes majeurs de ponctuation, voire (c) les tentatives d'autocorrection. Ces dernières, si elles ne sont pas systématiquement relevées peuvent conduire à des erreurs de diagnostic (exemple : analyse de « le plus que je possible » comme énoncé *dysyntaxique* si

[5] Encore qu'il existe des cas où la forme écrite d'un mot vient contaminer la production de sa forme orale ! Il peut être difficile d'interpréter certains phénomènes à partir des seules instructions du présent rapport. Compte tenu de ce fait, il est envisagé d'organiser des séances de « formation » afin d'aider les futurs annotateurs au maniement de la grille d'analyse des erreurs proposée dans ce manuel.

[6] Cf. par exemple (Vella et al. 2006).

nous ne prenons pas en considération le fait que le sujet s'est arrêté un bon moment après le « je » !) parce qu'il y avait conflit sous-jacent entre deux formulations synonymes : « le plus que je peux » et « le plus possible », un conflit ayant conduit, en surface, à un *télescopage* !

3. Erreurs liées à la configuration spatiale des lettres sur un clavier

La saisie « clavier » a ses propres contraintes, et celles-ci peuvent être à l'origine d'erreurs que l'on ne saurait observer en production manuscrite.

En écriture manuscrite, de même qu'en production orale, une erreur « segmentale » trouve son origine, soit dans la confusion entre segments (en général) proches du point de vue de leurs propriétés intrinsèques (phonologiques ou orthographiques) : cf. /p/ VS /b/ ou, pareillement « m » VS « n », soit dans la survenue de « contaminations contextuelles » (ou syntagmatiques), lesquelles ont, par exemple, pour effet la réduplication à courte distance de segments de l'environnement antérieur (= persévérations) ou postérieur (= anticipations). Les dyslexiques en savent quelque chose !

En écriture sur clavier, à ces erreurs (toujours possibles) s'ajoutent celles qui peuvent émaner de la proximité spatiale de certaines lettres, et ce, même si « lettres substituantes » et « lettres substituées » n'ont rien en commun au plan intrinsèque dans le système alphabétique de la langue. Par exemple, les touches correspondant aux lettres « e », « r », « s », « f » sont toutes voisines, sur un clavier AZERTY[7], de la touche « d » ; ce voisinage peut entraîner, par exemple, des substitutions. Mais elles n'ont par contre aucun point commun en termes de graphie (comme par exemple « p » et « q »), ou de nasalité (comme « m » et « n »). Ce point risque fort d'être crucial pour divers types de populations pathologiques présentant des problèmes moteurs importants (quelle qu'en soit l'origine) !

Sur un clavier AZERTY, il y a même pire ! La lettre « s » et la lettre « z » sont spatialement très proches (= la seconde est juste en dessous de la première !). Dès lors comment départager, au plan interprétatif (psycholinguistique), les (trois) possibles déterminismes sous-jacents de la même erreur de surface : (a) /s/ et /z/, en tant que phonèmes de la langue française ne se différencient que par un seul trait phonétique (= la première est « non-voisée » ; la seconde est « voisée ») ; (b) les lettres « s » et « z », graphiquement, et « visuellement », sont très proches (= la première recourt à des courbes là où la seconde recourt à des lignes brisées) ; (c) les deux « touches », très proches spatialement, constituent une troisième source potentielle d'erreurs. De là à proposer une optimisation des claviers afin d'éviter de telles ambiguïtés dans l'interprétation du déterminisme sous-jacent, il n'y a qu'un pas !

Ainsi, si la saisie clavier (physique et/ou logicielle) est susceptible d'aider certains sujets à produire de l'écrit alors qu'ils ne sont pas capables de le faire via leur main dominante, cette même saisie vient, d'un autre point de vue, rajouter une nouvelle source d'erreurs ![1] Ces problèmes sont évoqués en détails dans Vigouroux et al. (2005). On peut ainsi envisager de modéliser ce type d'erreurs en attribuant une « pondération » aux lettres faisant l'objet d'une erreur, en fonction de leur proximité plus ou moins grande sur le clavier[8].

[7] Pour permettre au lecteur de juger facilement ce point, nous plaçons dans l'annexe 2 une représentation d'un clavier AZERTY.

[8] Voir les travaux menés dans le cadre du projet Chatcom : http://www.irit.fr/chatcom, qui a donné lieu à un rapport (Vella et al. 2006). Ceci étant, dans le cadre d'une étude qui prendrait en compte la chronométrie des « frappes de touches » successives, la distance physique des touches (= des lettres sur le clavier) devra être précisément calculée, voire modélisée.

4. Erreurs VS Stratégies

Un dernier point mérite d'être mentionné ici, lequel n'est pas toujours facile à gérer. Il semble facile de qualifier d' « erreur » toute production non canonique. Ceci étant, il arrive que certains phénomènes erronés ne soient pas la conséquence directe d'un « déficit », mais plutôt, la mise en œuvre de stratégies (plus ou moins conscientes et volontaires) susceptibles de faciliter la tâche à l'émetteur. Les écrits SMS (Short Message Service) en fournissent de bons exemples : « g » pour « j'ai »…). Pour plus de détails à ce sujet, le lecteur pourra consulter notamment (Nespoulous & Virbel., 2004), ainsi que (Guimier De Neef et al. 2007) et (Guimier de Neef et Véronis 2006).

5. Stabilité VS Instabilité des erreurs

Il va sans dire que la fréquence des erreurs doit être mesurée. En cas de systématicité, l'erreur traduit en général une « carence de compétence » alors que le caractère épisodique de l'erreur traduit la maîtrise des règles normatives de la part de l'émetteur, celle-ci trouvant alors son origine dans un dysfonctionnement (souvent) de type attentionnel (les « fautes d'inattention » de nos instituteurs !). On parlera alors d' « erreurs de performance ».

Ceci étant, la systématicité d'une erreur peut également indiquer la mise en place d'une « stratégie » (cf. le point précédent), d'où les difficultés déjà mentionnées !

Pour trancher entre « erreur de compétence » et « stratégie » systématique, la seule voie viable consiste à soumettre le même sujet à une tâche de jugement de grammaticalité. On lui demande alors de se prononcer sur le caractère bien formé ou non de messages qu'il n'a pas produits lui-même. Il n'a qu'à se prononcer par « oui » ou par « non », selon que le message lui semble « correct » ou pas.

S'il réussit bien dans cette tâche, alors qu'il recourt spontanément (et systématiquement) à certaines formes erronées, c'est vraisemblablement qu'il a adopté une stratégie, et ce, à nouveau, consciemment ou pas (c'est un autre problème !).

6. Note sur le mode de collecte des données

Nous avons concocté ce qui suit sur la base, (a) de notre pratique antérieure en matière d'étude de productions pathologiques (chez des cérébrolésés) ainsi que (b) sur la base des échantillons que nous avons pu collecter par nos propres réseaux.

Ceci étant, en vue d'un travail à venir, plus contrôlé et rigoureux, il conviendra de veiller à sélectionner diverses tâches de production écrite identiques pour tous les sujets, éventuellement réutilisées, chez les mêmes sujets, à différents moments (cf. étude longitudinale).

Ainsi serait limitée l'inévitable hétérogénéité liée à une collecte écologique de données. Certes, celle-ci séduit toujours par son caractère « naturel » mais elle rend aussi les données glanées auprès d'un sujet difficilement comparables avec celles recueillies auprès d'un autre sujet, dans une situation (presque toujours) différente !

Première suggestion de situation[9] : le *discours narratif* à partir de la présentation d'un matériau iconographique à décrire verbalement (ici à l'écrit). La situation n'est pas totalement

[9] Nous réfléchissons actuellement à d'autres situations …

artificielle. Elle fournit du « texte continu » et les images contraignent les sujets à ne pas trop « prendre la tangente » en traitant de contenus chaque fois différents.

II) Typologie des erreurs

Nous présentons d'abord les différents niveaux auxquels une analyse des erreurs intervenant à l'écrit (avec un focus sur la saisie clavier) peut être menée. Nous nous intéressons plus particulièrement à l'un de ces niveaux, pour des raisons que nous expliquons. Dans une seconde section, nous nous livrons à une catégorisation détaillée des erreurs pouvant survenir au niveau choisi.

1. Approche macroscopique

A l'écrit, les différents niveaux auxquels interviennent les erreurs sont au nombre de cinq. On peut les hiérarchiser, du plus global au plus particulier de la manière suivante :

a) Recours (on non) à de la MFM[10] (au plan spatial essentiellement) ;
b) Problèmes de segmentation en phrases (➔ énoncés phrases, même si celles-ci comportent des erreurs) VS énoncés non-phrases (= « style télégraphique ») ;
c) Problèmes au niveau du substantif : il peut s'agir aussi bien du substantif dans sa globalité que de ses entités constituantes : lettres, morphèmes, etc. ;
d) Problèmes de gestion de la ponctuation : point VS virgule, au niveau phrastique et intra-phrastique ;
e) Problèmes de gestion des blancs inter-mots, voire intra-mots, ces derniers sont appelés « erreurs logogrammiques » par Catach (1980).

Dans l'ensemble du travail présenté dans ce rapport, nous ne nous sommes penchés que sur le niveau c. Il y a plusieurs raisons à cela. Les deux principales sont, d'une part, qu'il se retrouve également, nonobstant certaines différences, dans les productions orales ; d'autre part, il a fait l'objet de nombreuses études et expérimentations, notamment psycholinguistiques, sur lesquelles nous pouvons nous baser.

2. Approche analytique

2.1. Types d'erreurs

Le but est de permettre une description aussi précise que possible des erreurs, du niveau le plus concret au plus abstrait. Pour cela, on utilise une hiérarchie de catégories. Celle-ci est représentée dans la figure 1, que nous explicitons immédiatement après.

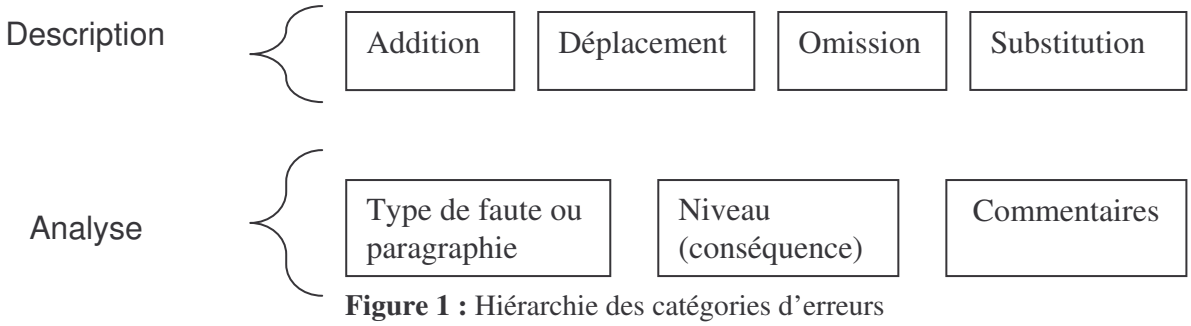


Figure 1 : Hiérarchie des catégories d'erreurs

[10] Mise en Forme Matérielle du texte. Pour résumer, cette théorie (Virbel, 1989) postule notamment la prise en compte de la mise en forme du texte écrit (indentation, mise en gras, etc.) dans la transmission du sens. Par exemple, la mise en italique d'un mot indique un focus particulier porté sur celui-ci.

Le premier niveau correspond à la description de la manifestation de la faute. En effet, toute faute appartient à l'une des catégories de base suivantes : Addition, Déplacement, Omission, Substitution. Nous nous y réfèrerons via l'acronyme *ADOS*. Cette catégorisation a déjà été prise en compte dans l'algorithme de (Levenshtein 1966). Cet algorithme est utilisé en TAP (Traitement Automatique de.la Parole) et en TALN (Traitement Automatique du Langage Naturel), pour calculer les taux respectifs de reconnaissance en graphème ou phonème VS les fautes dans l'écriture (cf. par exemple, (Pérennou et al. 1986)).

Le deuxième niveau, « analyse », correspond à une tâche plus délicate, puisqu'il s'agit de catégoriser de manière plus fine l'erreur, de situer le niveau qu'elle affecte, et d'avancer des hypothèses quant au motif de son apparition. Nous décrivons plus en détail, dans la partie suivante chacune de ces étapes.

Il y a quatre grandes catégories d'erreurs portant sur les différents types d'unités linguistiques :

- Addition ;
- Déplacement ;
- Omission ;
- Substitution.

Lorsqu'une seule erreur de l'une de ces catégories apparaît dans un message, l'analyse ne pose aucun problème. En revanche, lorsque plusieurs erreurs de l'un ou l'autre de ces catégories apparaissent (essentiellement dans le cas des omissions), l'analyse devient plus complexe, surtout s'il s'agit de l'omission de plusieurs mots grammaticaux, auquel cas, il vaudra mieux recourir à l'étiquette d'« agrammatisme » pour caractériser l'énoncé en question (sans chercher à quantifier le nombre d'omissions de morphèmes… ce qui de plus, s'avèrerait quasiment impossible)[11].

2.2. Types d'unités linguistiques perturbées (et leurs interprétations)

S'agissant du langage écrit, un certain nombre d'unités sont pertinentes et sont toutes susceptibles d'être « malmenées » en situation de production écrite. Nous les décrivons ci-dessous.

Reprenons les deux niveaux évoqués dans la figure 1, avec pour chacun les différentes catégories correspondantes :

Niveau « Description » :

Les différentes catégories : Addition, Déplacement, Omission, Substitution, sont assez explicites, et des exemples sont donnés dans la section suivante. Il est ; par contre, important d'apporter deux précisions.

D'une part, l'unité pouvant faire l'objet d'un des phénomènes catégorisés est en général la lettre (ou le caractère qui est son équivalent sur le clavier) ; mais selon les circonstances, il pourra s'agir de plusieurs caractères, ou encore d'un ou plusieurs phonèmes. On notera d'ailleurs qu'en français, lettre et phonème ne sont pas systématiquement dans une relation terme à terme. En effet, des phonèmes peuvent être réalisés orthographiquement par plusieurs lettres pour aboutir à un « graphème ». Un exemple typique est le phonème /ã/, qui, dans le mot « exempt » (/ɛgzã/, correspond à pas moins de 4 lettres (voire 5 au pluriel).

[11] Alors que les omissions, ici ou là, de lettres peuvent, elles, être aisément quantifiées.

D'autre part, les erreurs concernant les diacritiques doivent faire l'objet d'un traitement particulier lorsque c'est un clavier d'ordinateur qui a été utilisé pour produire le texte analysé. On fera, en effet, la distinction entre d'un côté les lettres accentuées présentes « d'un bloc » sur une seule touche de clavier (en français : ç, é, è, à, ù) et de l'autre celles formées par la combinaison de 2 touches (en français, diacritique ^ ou ¨, et une voyelle comme e, par exemple). Dans le premier cas, une seule faute sera comptabilisée. Dans le second, on pourra compter 1 à 2 fautes : une pour le diacritique, et une, éventuellement, pour la lettre accentuée. Illustrons ceci par un exemple : l'emploi du verbe « a » au lieu de la préposition « à » ne comptera que pour une seule erreur (une substitution) ; on comptabilisera par contre deux erreurs dans « hïtre » (au lieu de « hêtre ») : substitution de l'accent circonflexe par le tréma, et substitution de la lettre « e » par la lettre « a »[12]. Enfin, une dernière règle est importante à signaler concernant les accents : on ne prendra pas en compte les cas de substitutions d'un diacritique par un autre qui n'entraînent pas de changement aux niveaux phonologique, mophologique, ou syntaxique[13] (par exemple « é » VS « è » dans « aprém ») Le principe sous-jacent derrière cette règle est, comme nous le disons plus haut (p. 6), nous nous intéressons ici à de l'écrit « oralisé ». Dans ce contexte, la substitution d'un diacritique par un autre est considérée comme une manifestation écrite de comportements oraux « acceptables » linguistiquement.

Au niveau « Description », la classification est faite selon l'unité linguistique perturbée :

- *La « lettre »* : une erreur de ce type sera appelée « *paragraphie littérale* » (**PL**)

  Exemples[14] :
  tornade ➔ *tornadre addition (avec persévération)
  cultivateur ➔ *curvilateur déplacement
  chercher ➔ *checher omission
  fournir ➔ *fourfir substitution (avec persévération)

  Dans les cas de substitutions, on notera scrupuleusement : (a) la « proximité intrinsèque » entre substituant et substitué, (b) de même que l'éventuelle entrée en jeu de contamination contextuelle (persévérations et anticipations ; cf. p.8), sans oublier (c) les erreurs qui viendraient perturber la lecture à haute voix des mots écrits de manière erronée (exemple : boisson ➔ *boison) et débouchant parfois sur la mise à mal d'oppositions lexicales existant dans la langue (exemple: poisson ➔ poison).

  Ces deux derniers phénomènes (persévérations et anticipations) seront également notées dans les cas d'additions.

  C'est à cette première catégorie d'erreurs que viendront se rajouter les paragraphies littérales engendrées par proximité des touches sur le

[12] A noter que cette règle n'est pas applicable dans le cas de la substitution d'une lettre accentuée par un groupe de lettres ou réciproquement (exemple : é ➔er).
[13] Décrits plus bas.
[14] Exemples tirés de Lecours, A.R., Dordain, G., Nespoulous, J-L. & Lhermitte, F. Le vocabulaire de la neurolinguistique, in A.R. Lecours & F. Lhermitte (Eds) *L'aphasie*, Paris, Flammarion, 1979.

clavier (**PLC**). Pour une prise en compte des PLC dans un système d'assistance à l'écriture, cf. Boissière & Dours (1996, p.170).

Appartiennent également à cette catégorie, les erreurs portant sur les diacritiques (accents), dans les conditions présentées p. 11.

Les éventuelles erreurs de majuscules entrent également dans cette catégorie[15].
En ce qui concerne les PLC, il est important de préciser que cette catégorie ne doit être employée que pour les seules erreurs orthographiques d'addition, ou de substitution entre une lettre cible et un périmètre situé immédiatement autour[16]. Pour en juger, on pourra utiliser si besoin est, l'image de clavier donnée en annexe I.
Soit, par exemple, le verbe « trouve ». L'erreur « *troube », sera considérée comme une PLC par substitution de « v » par « b », car les deux lettres sont immédiatement contigües sur le clavier. En revanche, « *trouhe » sera une PL car la lettre « h » est plus éloignée du « V ».

- *Le « graphème »* : une erreur de ce type sera appelée « *paragraphie graphémique* » (**PG**). Il s'agira essentiellement de substitutions.

  Le « graphème » est l'équivalent (ortho)graphique d'un phonème.
  Une « lettre » peut correspondre à un graphème mais fort souvent, surtout dans une langue comme le français, un graphème nécessite la mobilisation de plusieurs lettres (parfois nombreuses). L'exemple de « exempts » déjà donné p. 11 s'applique également ici.
  Cette catégorie d'erreurs est donc surtout utile pour qualifier les erreurs de production écrite de *phonèmes hétérographes*[17] : le sujet utilise une variante graphémique erronée (qui, néanmoins, permet de renvoyer au bon phonème à l'oral).

  Exemples (substitutions) :
  bibliothèque ➔ *bibliotec
  français ➔ *françait
  commencement ➔ *comansement

- *Le « morphème »* : une erreur de ce type sera appelée « paragraphie morphémique » (**PM**).
  Dans sa forme la plus simple à analyser, il s'agira d'une substitution de morphèmes (le plus souvent flexionnels) : « il chantait » ➔ « * il chantais ». Ces erreurs auront déjà été comptabilisées comme erreurs graphémiques et littérales.

  Rentrent également dans cette catégorie les erreurs d'accords morphologiques. Par exemple : « *les petit garçons ». La forme erronée « petit » aura déjà été analysée comme « paragraphie

[15] Cependant, comme nous l'indiquons p.13, nous ne les comptabilisons pas encore dans le cadre du présent travail.
[16] En attendant qu'éventuellement des études plus poussées élargissent le champ d'application des PLC.
[17] On retrouve ici la notion de « phonogramme » de N. Catach.

littérale » (omission de lettre) mais sera ré-analysée à ce niveau comme violation de la règle d'accord morphologique. Même chose pour les accords entre sujet et verbe (exemple : «* Les garçons chante »). A plus forte raison, des erreurs du type « ne pas vouloir faire VITIPI », où le verbe n'est pas fléchi mais apparaît à l'infinitif, ce qui permettra d'utiliser également l'étiquette d'« agrammatisme » pour cet énoncé !
Les omissions de morphèmes grammaticaux entrent également dans cette catégorie ; exemple : « Je ne veux pas… faire exercice »[18] ➔ omission d'article (= « l' ») ou de démonstratif (= « cet »).

Idem pour les additions de morphèmes grammaticaux. Cf. le « célèbre » **pallier à.*

Entrent également dans cette catégorie les erreurs de préfixes. Exemple : impossible ➔ * inpossible.

Il y a enfin une autre catégorie, celle des « Perturbations Morphémiques » (PeM). Elle se divise en deux sous-catégories :

- *« Perturbation Morphémique fusion»* (PeMf) : omission d'une lettre (caractère) séparant deux substantifs (espace, trait d'union, apostrophe), et aboutissant ainsi à la réunion de ceux-ci. Exemple : trait d'union: ➔ * trait dunion

- *« Perturbation Morphémique segmentation»* (PeMs): addition d'une lettre (caractère) séparant (espace, trait d'union, apostrophe) à l'intérieur d'un substantif, aboutissant ainsi à la segmentation de ce dernier en deux unités. Exemple : proposer ➔ * prop oser

Niveau « conséquence » :

Les différents niveaux pouvant être affectés ici sont les niveaux orthographique, phonologique, morphologique et syntaxique, dont les abréviations sont respectivement, « phono », « morpho », et « syntaxe ». Le niveau orthographique n'a pas d'abréviation. En effet, il est par définition présent dans toute erreur : dans le cas où aucun niveau ne se surajoute à celui-ci, on choisit la convention de ne rien mentionner.

En règle générale, les fautes relevant de la PL affectent seulement le niveau orthographique et/ou phonologique ; les PG et les PM peuvent affecter, selon les cas, les 4 niveaux. Voici quelques indications supplémentaires pour chaque catégorie (hors orthographe) :

- *Phonologique* : pour déterminer si une erreur concerne le niveau phonologique, il suffit de comparer la représentation phonétique du mot erroné avec celle du mot cible, par exemple en lisant à haute voix. Si les deux représentations sont identiques, alors il n'y a pas lieu de parler « d'erreur à conséquence phonologique ». Précisons que de ce fait, l'interprétation peut différer selon les variantes locales, telles que la prononciation ou non du e final : par défaut, on se basera

[18] Phrase extraite du corpus S7 84 02.

pour l'interprétation sur le français dit « standard », ce qui implique notamment de ne pas considérer comme une erreur à conséquence phonologique l'oubli du e muet final aussi appelé le « schwa ») ;

- *Morphologique* : il faut faire attention à ne pas confondre les erreurs morphologiques et syntaxiques. Les erreurs d'ordre morphologiques, comme leur nom l'indique, concernent uniquement la morphologie grammaticale du mot : il peut s'agir par exemple d'une faute de flexion (conjugaison, genre, nombre…). Il est important de noter qu'il n'y a pas d'erreur au niveau morphologique si le mot erroné ne ressemble pas à un mot existant de la langue (ainsi, « les » au lieu de « le » est une faute au niveau morphologique, mais pas *« lb », qui est une erreur aux conséquences orthographiques et phonologiques) ;

- *Syntaxique* : le terme « syntaxique » est pris ici au sens de la position et de la fonction des mots dans l'énoncé. De ce fait, on ne parlera d'erreur de syntaxe que dans le cas d'erreurs à ce niveau. Par exemple, l'omission d'un substantif, ou son déplacement à une autre position que celle qui lui est assignée dans la langue, est une erreur de syntaxe (Exemple : * « il y longtemps » au lieu de « « il y a longtemps ». De fait, il faut savoir qu'on constate très peu d'occurrences de ce type d'erreur dans la littérature (psycholinguistique notamment). En cas de doute, il est conseillé de tenir compte de tout l'environnement syntaxique de la phrase.

En règle générale, on mentionnera uniquement l'erreur (la conséquence) de plus haut niveau lorsque cette erreur (cette conséquence) couvre les autres erreurs (conséquences) de niveaux inférieurs. Exemple: PM implique les erreurs PL et PG ; « morpho » implique la conséquence « phono ».

En revanche, dans le cas où on n'observe pas un recouvrement, on adoptera une notation qui montrerait la non-manifestation des autres erreurs (ou conséquences). Par exemple, si on observe une erreur morphologique et pas graphémique on notera : PM - PG.

Niveau « Commentaires » :

Il s'agit de décrire avec encore plus en détails les erreurs, voire d'émettre des hypothèses quant à l'origine psycholinguistique et/ou motrice des erreurs (par exemple, persévération ou anticipation). Dans le cadre d'une première version de l'annotation, les difficultés rencontrées pourront également y être mentionnées pour discussion et mise à jour des règles d'annotation pour une meilleure robustesse et représentativité dans le futur.

III) Application pratique des règles d'annotation

Cette partie décrit les méthodes pratiques à mettre en œuvre pour annoter les erreurs. Nous répondons d'abord à la question de savoir quand catégoriser un phénomène comme étant une erreur ou non, puis nous présentons les conventions d'annotation utilisées.

1. Que considérer comme étant une erreur, dans l'état d'avancement de notre méthodologie ?

La question mérite d'être posée, car la réponse n'est pas aussi évidente qu'il ne paraît. En effet, toute séquence de lettres/caractères ne correspondant pas à l'orthographe standard d'un mot ne sera pas systématiquement prise en compte comme étant une erreur. Plus précisément, on ne prendra pas en compte :

- *Les fautes relevant d'une méconnaissance « commune » de la langue* : par exemple, on ne cherchera pas à catégoriser « *vous avez intérêt **de** » ou « *pallier **à** », respectivement de substitution de « de » par « à », et d'addition. On fera également attention au registre de langue induit par le contexte de production. Ainsi, on ne prendra pas en compte les erreurs relatives à un registre de langage soutenu dans l'analyse de textes écrits dans un cadre plus « relâché », tel que des récits « oralisés ». Par exemple, dans un tel contexte, on ne considérera pas l'absence de « ne » dans « ça marche pas » comme une faute ;

- *Toute faute affectant un nom propre* : Pour les prénoms ou les noms « célèbres », quand ils sont reconnaissables, la prise en considération ou non d'une faute dépendra de la fréquence d'emploi dans la langue quotidienne ;

- *Certains cas particuliers qui peuvent être effectivement une erreur, mais pour lesquels une certitude n'est pas acquise* : Par exemple, dans l'énoncé « le film raconte l'histoire d'un **handicap** ». Il paraîtrait logique de penser que le dernier mot est en fait « handicapé », et résulte donc d'une omission du « é » final. Mais comme il est également possible, dans le contexte de l'énoncé, que « handicap » soit le mot correct, on préférera ne pas se prononcer. Nous ferons ainsi la distinction entre la catégorie « indéchiffrable » (abréviation **IND**), qui correspond à toute séquence de lettres/caractères ne pouvant être identifiée à un mot existant, et « indécidable ». Nous mentionnons cette dernière catégorie pour être exhaustif dans la description de la méthodologie, mais de fait, par définition elle ne compte pas comme une erreur et n'est donc pas annotée ;

- Pour l'instant, les fautes relatives à l'utilisation de majuscules ne sont pas prises en compte.

Il peut arriver que certaines erreurs posent des problèmes d'analyse et d'annotation du fait de leur complexité ou de leur ambigüité. Dans ce cas, on conseille d'étudier d'autres erreurs, plus facilement interprétables, du même corpus/auteur, afin de chercher d'éventuels patterns qui pourraient s'appliquer aux cas plus difficiles.

D'une manière générale, les recommandations décrites dans ce manuel peuvent (et sans doute même devront) être ajustées en fonction des caractéristiques de la population ayant produit le corpus. Par exemple, dans le cas de troubles pathologiques de la production orale[19], on ne tiendrait pas ou peu compte d'erreurs relevant du niveau phonétique. En effet, le sujet souffrant précisément de troubles à ce niveau, il ne serait pas pertinent de se référer dans ce cas à une opposition norme VS erreur.

2. Conventions d'annotation

L'annotation[20] se fait dans un tableau sous tableur (Excel © ou OpenOffice Calc). Deux exemples en sont donnés dans l'annexe 3 : l'énoncé à analyser est présenté dans une ligne du tableau, et chaque substantif faisant l'objet d'une ou plusieurs erreurs est mis en gras. Chaque ligne suivante est consacrée à une erreur en particulier. On appliquera également cette règle dans le cas où un seul substantif présente plus d'une erreur, à l'exception de la colonne « mot en faute », avec utilisation de la fonction « fusion des cellules ». Les colonnes correspondent aux différentes catégories que nous avons décrites dans la partie précédente. Quand tous les mots en faute ont été ainsi présentés et traités, on passe à l'énoncé suivant, et ainsi de suite.

Les conventions de notation pour chaque niveau sont décrites ci-dessous.

Niveau « Description » :

Pour chaque erreur, une et une seule des colonnes est remplie avec l'unité (lettre/caractère(s), phonème(s), graphème(s), ou substantif), faisant l'objet de l'erreur. Le traitement successif des erreurs se fait de gauche à droite, par rapport au mot correct (ne comprenant pas de faute).

Dans le cas de la substitution, on utilise la notation suivante : (unité erronée) → (unité correcte). Par exemple, la substitution finale dans « * foyé » sera annotée en plaçant « é → er » dans la colonne « Substitution ».

Pour marquer un déplacement, on reporte dans la colonne correspondante la lettre déplacée située le plus à droite dans le mot correct[21], De plus, pour permettre au lecteur de l'annotation de reconstituer le mot correct, on rajoutera, également entre parenthèses, le nombre de lettres situées entre la position initiale de la lettre déplacée et celle qu'elle occupe dans le mot erroné. Ainsi, «* utilrise » (pour « utiliser ») sera annoté r(-3) : par rapport au mot correct, le « r » (la lettre située initialement le plus à droite, a subi un décalage de 3 positions (autrement dit, elle a « enjambé » 3 caractères), marquée donc par un « -3 » (cf. annexe 3). La seule exception à cette règle concerne les cas de « double déplacement » (métathèse réciproque) : l'un des déplacements (celui de la lettre située le plus à droite dans le mot correct sera annoté comme nous venons de l'expliquer, le deuxième (lettre initialement le plus à gauche) le sera avec un nombre positif (exemple : « *qarbue » pour « barque » sera annoté

[19] Attestés par exemple dans une fiche clinique du sujet scripteur.

[20] A terme, les annotations se feront au format XML (eXtended Markup Language).

[21] Ceci n'est évidemment qu'une convention d'annotation, qui ne préjuge pas d'une quelconque interprétation : il aurait tout autant été acceptable de considérer que c'est la lettre située « le plus à gauche » dans le mot correct qui a été déplacée.

« q(-2) » et « b(+2)[22] ». Enfin, dans le cas de noms ou mots composés des caractères, tels que l'espace ou le trait d'union, seront également comptés comme une position.

A noter enfin que dans certains cas, le ou les lettres/caractères faisant l'objet de l'erreur, pourront être représentés dans les colonnes, non sous la forme de lettres mais de symboles phonétiques de l'API[23].

Niveau « Analyse » :

- *Colonne « Type de faute ou paragraphie »* : utilisation des abréviations décrites dans « Quelles catégories pour quelles erreurs ? » ci-dessus[24]. Dans le cas de la seule faute à conséquence orthographique, on laissera la case vide. S'il s'agit d'une erreur « indéchiffrable » (cf. plus haut), c'est dans cette colonne que l'on notera « IND » ;

- *Colonne « Niveau (conséquence) »* : idem ;

- *Colonne « Nature / Commentaire »* : la rédaction est laissée à l'appréciation de l'annotateur, avec une préférence pour des énoncés succincts.

[22] Le nombre de caractères « enjambés » est de 2 et non 3 puisque le 3ème caractère (ici le « b » et le « q ») n'occupe plus sa position initiale.

[23] Alphabet Phonétique International.

[24] L'algorithme de calcul est conçu pour laisser une certaine souplesse dans la notation des erreurs (il autorise par exemple « PM » ou « P.M. ». Cependant, pour éviter de se retrouver dans un cas non prévu (et donc indétectable) il est préférable que les annotateurs essaient autant que possible d'appliquer une convention d'écriture fixe : pas d'espace avant ou après l'abréviation, pas de points dans celle-ci, etc.

Conclusion et perspectives

Le travail que nous avons présenté n'en est encore qu'à ses débuts. Nous avons fait annoter un corpus d'une douzaine de phrases par deux annotateurs différents, à partir des règles décrites dans ce document. Les résultats de ce travail sont publiés dans (Boissière et al. 2007).

A court terme, nous envisageons de mener une campagne d'annotation à plus grande échelle, sur plus de corpus (corpus thématiques différents, sujets avec des handicaps langagiers différents, etc.) et avec plus d'annotateurs pour définir le coefficient de Kappa (Carletta 1996). Cela nous permettra d'obtenir une mesure de « l'accord inter annotateurs » afin d'évaluer de manière fiable et objective la stabilité et la robustesse de notre grille d'annotation[25].

Enfin, à plus long terme, on peut envisager de mener le même type de travail sur les autres niveaux présentés dans la deuxième partie de ce document.

Remerciements :

- Cette étude a été réalisée grâce aux financements de l'APETREIMC et de la Fondation Motrice.

- Les auteurs expriment également leur reconnaissance à nos partenaires du *Centre Mutualiste de Rééducation et de Réadaptation Fonctionnelles de Kerpape* qui nous ont gracieusement fait parvenir les textes écrits par certains de leurs patients. Que ces derniers en soient ici remerciés.

[25] Un formulaire de phrases à annoter va bientôt être mis en ligne à l'adresse : http://www.irit.fr/ESACIMC/annotation.html

Bibliographie

Baudot J.A. *Information, redondance et répartition des lettres et des phonèmes en français*, Rapport, Université de Montréal, mars 1968.

Boissière Ph., Dours D. "VITIPI : Versatile Interpretation of Text Input by Persons with Impairments", In *5th ICCHP (International Conference on Computers for Handicapped Persons)*. p.165-172, Linz, July 1996.

Boissière Ph., Bouraoui J.-L., Vella F., Lagarrigue A., Mojahid M., Vigouroux N., Nespoulous J.-L. « Méthodologie d'annotation des erreurs en production écrite. Principes et résultats préliminaires », *TALN'07, Atelier « Reconstruire la langue dans les communications alternatives et augmentées »*, Toulouse, 5-8 juin 2007.

Carletta J., "Assessing agreement on classification tasks: the kappa statistic", *Computational Linguistics*, 22 (2):p.249-254, 1996.

Catach N., *L'enseignement de l'orthographe*, Paris, Nathan, 1980.

Guimier De Neef E., Debeurme A., Park J. « TiLT correcteur de SMS : évaluation et bilan qualitatif », *TALN 07*, Toulouse, 5-8 juin 2007.

Guimier de Neef E. et Véronis J. « Le traitement des nouvelles formes de communication écrite », dans Gérard SABAH (dir.), *Compréhension des langues et interaction*, Hermès-Lavoisier, 2006.

Lecours A. R., "Serial order in writing – a study of misspelled words in "developmental dysgraphia"", *Neuropsychologia*, Vol.4, p. 221-241, 1966.

Lecours A. R., Deloche G., Lhermitte F., « Paraphasies phonémiques – description et simulation sur ordinateur – » dans *Colloque INRIA-Informatique Médicale*, p.311-351, Rocquencourt, 1973.

Lecours A. R., Lhermitte F., "Phonemic paraphasias: linguistic structures and tentative hypotheses", *Cortex*, 5, p.193-228, 1969.

Lecours A. R., Lhermitte F. et al., *L'aphasie*, Paris, Flammarion, 1979.

Lecours, A.R., Dordain, G., Nespoulous, J-L. & Lhermitte, F., « Le vocabulaire de la neurolinguistique », in A.R. Lecours & F. Lhermitte (Eds) *L'aphasie*, Paris, Flammarion, 1979.

Lecours, A.R. & Nespoulous, J-L., « Biologie de l'écriture », *Etudes Françaises*,18/1, p.33-45, Les Presses de l'Université de Montréal, 1982.

Levenshtein V.I., "Binary codes capable of correcting deletions, insertions, and reversals", *Cyber. Contr. Theory*,10 (8), p. 707-710, 1966.

Mounin G., *Introduction à la sémiologie*, Paris, Editions de Minuit, 1970.

Nespoulous, J-L. & Lecours, A.R. « Les troubles de l'écriture dans l'aphasie », *Etudes Françaises*, 18/1, p. 47-59, Les Presses de l'Université de Montréal, 1982.

Nespoulous, J-L. & Virbel, J., « Apport de l'étude des handicaps langagiers à la connaissance du langage humain », *Revue Parole*, N°29-30, p. 5-42, 2004.

Pérennou G., Daubèze P., Lahens F., « La vérification et la correction automatique de textes: le système VORTEX », *TSI (Technique et Science Informatique)*, volume 5, numéro 4, p. 285-305, 1986.

Vella F., Vigouroux N., Amadieu F., Raynal M., Tricot A., « Résultats quantitatifs de l'usage de claviers virtuels avec de nouvelles modalités *», projet ChatCom,* coordinateur B. Oriola, 20 pages, 16 octobre 2006.

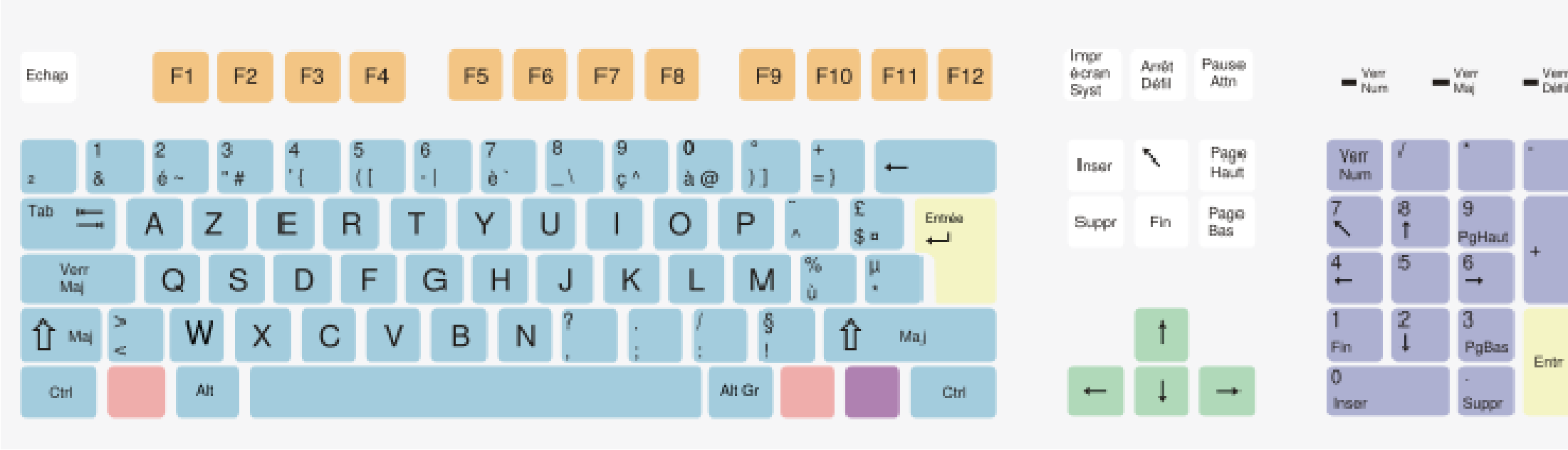
Vigouroux N., Vella F., Raynal M., Boissière P. « Solutions et défis pour une meilleure accessibilité et utilisabilité des communicateurs - Optimisation de la saisie de texte ». *Handicap et Environnment : de l'adaptation du logement à l'accessibilité de la cité - Entretiens de l'Institut de Garches*, Nanterre, 24/11/2005-25/11/2005, p. 209-222, novembre 2005.

Virbel J., "The contribution of linguistic knowledge to the interpretation of text structure". Dans André, J., Quint, V. et Furuta, R., (Eds) *Structured Documents*, p. 161–181, Cambridge University Press, 1989.

Annexe I : Typologie des erreurs (N. Catach, 1980)

TYPES D'ERREURS	SOUS-CATÉGORIES		EXEMPLES
ERREURS PHONÉTIQUES			**faimé (fermé)* **entrare (en retard)* **adrapé (attraper)*
ERREURS PHONOGRAMMIQUES	valeur phonique altérée		**causse (cause)* **fénêtre (fenêtre)*
	valeur phonique non altérée		**lampader (lampadaire)* **consierge (concierge)*
ERREURS MORPHOGRAMMIQUES	morphèmes grammaticaux	axe syntagmatique	*il *ramenée (ramenait)* *un *animaux (animal)*
		axe paradigmatique	**accroché (accrocher)* **envoyer (envoyait)*
	morphèmes lexicaux	Préfixes	**enmena (emmena)*
		Suffixes	**arrer (arrêt)*
ERREURS DISTINCTIVES	homophones grammaticaux		**sont (son) / *se (ce)*
	homophones lexicaux		**fois (foie)*
ERREURS LOGOGRAMMIQUES	segmentation de mots		**mattacher* *le *l'endemain*
ERREURS EXTRA-ALPHABÉTIQUES	Majuscules		**pierre (Pierre)*
	Ponctuation		
ERREURS A-SYSTÉMATIQUES	Historiques		**colit (colis)*
	Étymologiques		**moin (moins)*

Annexe II : Représentation détaillée d'un clavier azerty



© http://fr.wikipedia.org/wiki/AZERTY

Annexe III : Exemples d'annotation d'une erreur

NB : Les deux énoncés présentés ici sont tirés du corpus d'étude. Tous les types d'erreurs décrits dans le présent document sont issus de ce corpus. Par contre, nous avons artificiellement changé la nature, le nombre, et le positionnement des occurrences de fautes qui y apparaissent, à fins de représentativité des différentes catégories que nous avons présentées.

Mot en faute	**Addition**	**Omission**	**Déplacement**	**Substitution**	**Analyse de l'erreur**		
					Type de faute ou paragraphie	**Niveau (conséquence)**	**Nature / Commentaire**
il **estg** super **facip** à **utilisre**							
estg	g				PLC	Phono	
facip				p→l	PLC	Phono	
		e			PL	Phono	oubli [e] final
utilisre			r (-1)e(+1)		PLC	Phono	Métathèse réciproque

Mot en faute	Addition	Omission	Déplacement	Substitution	Analyse de l'erreur			
					Type de faute ou paragraphie	**Niveau (conséquence)**	**Nature / Commentaire**	
les **médcin on sinier poyr** mon **deuxièm** fauteuil **eletrique.** C'est **bein.** **j'ai merai** etre en **musur daprand** un **maitié** pour pas **allé** au **foyé** !								
médcin		e			PL			
		s			PM	morpho	singulier/pluriel	
On		t			PM	morpho	Verbe / pronom	
Sinier		g			PG	phono		
	i				PG	phono		
Poyr				y → u	PLC	phono		
deuxièm		e			PL			
eletrique				e → é	PL	phono		
		c			PL	phono		
Bein			e (-1)		PL	phono		
j'ai merai	"espace"				PM Segmentation		Même si la touche espace n'est pas "très proche" du de la touche m, avec sa forme assez longue elle peut avoir été touchée involontairement	
		s			PG	morpho		
Musur				u → e	PG	phono		
		e			PG			
daprand		'			PM fusion			
		p			PG			
				a → e	PG			
		re			PM	phono		

Mot en faute	Addition	Omission	Déplacement	Substitution	Analyse de l'erreur		
					Type de faute ou paragraphie	**Niveau (conséquence)**	**Nature / Commentaire**
maitiè				ai → é	PG	phono	
				è → er	PG	phono	
Ne		ne			PM	syntaxe	ne …. pas
Allé				é → er	PM		
Foyé				é → er	PG		