

ITI 2015

Table des matières

Session Invites

Dialogue homme-machine : conception et enjeux.....	1
Couplage structurel : une autre façon d'appréhender les interactions langagières homme - machine	2

Session Interprétation

Les modalités linguistiques pour aider à l'interprétation de textes juridiques.....	3-9
Interprétation rapide et non visuelle de pagesWeb : TAL et Tonnerre de mots.....	10-18

Session Handicap

Interfaces textuelles, une difficulté pour l'utilisateur sourd ? Réflexion et éléments de réponse autour d'une étude de cas.....	19-31
Navigation Aveugle des Pages Web sur Dispositifs Mobiles.....	32-43
TALN et IHM : une approche transdisciplinaire pour la saisie de textes de personnes en situation de handicaps.....	44-54

Dialogue homme-machine : conception et enjeux

Frédéric Landragin

LATTICE, CNRS, ENS & Université de Paris 3, 1 rue Maurice Arnoux, 92120 Montrouge
frederic.landragin@ens.fr

Résumé. Le but de cette présentation est de faire un point sur les théories, les méthodes, les techniques, les enjeux impliqués dans la conception de programmes informatiques capables de comprendre et de produire de la parole. Comment une machine peut-elle parler, comprendre ce qu'on lui dit, et entretenir un dialogue proche du dialogue naturel entre deux humains ? Quelles sont les étapes de conception d'un système de dialogue homme-machine ? Quelles sont les capacités de compréhension, de raisonnement et d'interaction attendues pour de tels systèmes ? Comment les implémenter ? Comment s'approcher du réalisme et de la fluidité du dialogue humain ? Ces questions sont à l'origine de mon parcours, qui a oscillé entre linguistique et informatique, entre recherche fondamentale et développement, entre laboratoires de recherche publics et privés : INRIA, puis THALES, et actuellement CNRS. Le but ici est de mentionner les principaux problèmes posés par chaque étape de conception d'un système de dialogue homme-machine, et de présenter quelques pistes théoriques et techniques pour traiter ces problèmes. Le but est aussi de montrer qu'il existe bien une école française du dialogue homme-machine, caractérisée par sa pluridisciplinarité, par son implication dans différents secteurs, qu'il s'agisse de développement de systèmes, de mise en œuvre de méthodes et de campagnes d'évaluation, de conception d'architectures logicielles, de dialogue multimodal, d'ergonomie, d'agents conversationnels animés, ou encore d'application de techniques d'apprentissage automatique évoluées au dialogue homme-machine.

Abstract.

Human-Machine Dialogue: Design and challenges.

The goal of this presentation is to outline the theories, methods, techniques and challenges involved in the design of computer programs that are able to understand and produce speech. How can a machine talk, understand what is said and carry out a conversation close to natural conversation between two human beings? What are the design stages of a human-machine dialogue system? What are the understanding, thinking, and interaction abilities expected from such systems? How should they be implemented? How can we get closer to the realistic and fluid aspect of human dialogue? These questions are at the origin of my path, which oscillated between linguistics and computer science, between pure research and development, between public and private research laboratories: INRIA, then THALES and currently the CNRS. The goal here is to mention the main issues created by each stage of the design of a human-machine dialogue system, and to present a few theoretical and technical paths used to deal with these issues. The goal here is also to show that today there still is a French school of human-machine dialogue, characterized by its multidisciplinary approach, its involvement in different fields, such as system development, implementation of assessment methods and campaigns, software architecture design, multimodal dialogue, ergonomics, embodied conversational agents, and application of up-to-date machine learning techniques.

Mots-clés : Dialogue humain-machine, cycle de conception, généricité, multimodalité, génie logiciel.

Keywords: Human-machine dialogue, design cycle, genericity, multimodality, software engineering.

Couplage structurel : une autre façon d'appréhender les interactions langagières homme - machine

Jacques Labiche

Groupe v, Laboratoire LITIS, Université de Rouen
jacques.labiche@univ-rouen.fr

Résumé. A partir de la notion de couplage structurel tel que proposé par les biologistes Maturana et Varela, je revisiterai la problématique des interactions entre utilisateur et système informatisé. Le développement d'une plateforme informatique dans le domaine du droit du transport sera évoqué. Les traces langagières déposées volontairement par l'utilisateur lors de ses interactions avec la plateforme feront l'objet d'un questionnement approfondi.

Abstract.

Structural coupling: another way of comprehend the linguistic interactions man – machine.

From the concept of structural coupling as proposed by biologists Maturana and Varela, I shall question the problem of interaction between an user and computer system. The development of an IT platform in the area of transportation law will be discussed. Linguistic cyber trail submitted voluntarily by the user when interacting with the platform will undergo a thorough questioning.

Mots-clés : autopoïèse, couplage structurel, interactions homme-machine.

Keywords: autopoiesis, structural coupling, man machine interactions.

Le point de vue adopté pour la présentation sera très en amont dans la mesure où il tente de poser un cadre systémique à la dynamique des interactions entre un utilisateur souhaitant acquérir des connaissances et son système informatisé.

Dans une première partie, j'analyserai la notion de couplage. En partant des systèmes isolés, j'introduirai les systèmes autopoïétiques, seuls aptes à décrire les systèmes biologiques, donc le vivant. En admettant avec (Maturana, Varela) et (Luhmann) que l'on peut étendre la notion de système autopoïétique à l'homme, mais aussi aux sociétés humaines, je m'attarderai sur la nature « structurelle » du couplage ainsi que sur la notion de clôture opérationnelle qui spécifie l'unité du système vivant. Cela permet, dans une perspective phénoménologique, de prendre en compte, l'histoire, la dynamique, donc le vécu de l'utilisateur en interaction avec la machine.

Dans une deuxième partie, après avoir montré quelles sont les conséquences de ce point de vue pour appréhender les notions de communication et de langage, je décrirai l'une des plateformes en développement au sein du groupe v. Les considérations mises en avant dans la première partie montrent qu'une plateforme informatisée se doit d'être centrée utilisateur (Pierre Beust) afin qu'il puisse réellement se produire un couplage structurel entre utilisateur et système. Cette plateforme d'aide à l'acquisition de connaissances dans le domaine juridique, encore à l'état de prototype, présente la particularité de proposer à l'utilisateur de conserver et, si besoin, de réutiliser ses traces.

Dans la dernière partie, j'analyserai plus particulièrement les possibilités offertes à l'utilisateur de déposer des commentaires textuels lors de sa navigation, puis en « rejouant le scénario » de visualiser ses propres commentaires.

Ce sont les nombreuses questions que pose cette démarche appréhendée du point de vue de l'énaction (Varela) qui me permettront de conclure en interagissant avec l'auditoire.

Références

Maturana H.R., Varela F. (1994), *L'arbre de la connaissance*. France : Ed. Addison Wesley.

Varela F. (1989), *Autonomie et connaissance, essai sur le vivant*. Paris : Edition du seuil, La couleur des idées. Trad. de *Principles of Biological autonomy*. 1980, New York, Elsevier.

Luhmann N. (1994). Le droit comme système social. *Droit et Société* 11/12, 53-67.

Les modalités linguistiques pour aider à l'interprétation de textes juridiques

Fadila Taleb^{1, 2}

(1) Laboratoire Dysola EA 4701 (Dynamiques sociales et langagières), Normandie Université,
Université de Rouen

(2) Litis EA 4108 (laboratoire d'informatique, de Traitement de l'Information et des Systèmes),
Normandie Université, Université de Rouen et INSA de Rouen
fadila.taleb1@univ-rouen.fr

Résumé. L'un des objectifs majeur que se fixent les concepteurs des bases de données est celui de permettre un accès facile et rapide à de multiples documents et informations soit dans un domaine ciblé ou dans des domaines variés. Ces ressources numériques doivent aussi prendre en compte et répondre efficacement aux besoins pratiques de leurs futurs utilisateurs. Cependant, l'accès aux documents d'un corpus numérique suscite de nombreuses interrogations, notamment sur la recherche d'informations. Cette dernière se voit souvent réduite à un réseau restreint de mots-clés, ce qui ne saurait satisfaire à tous les besoins des utilisateurs. Nous proposons dans cet article une approche linguistique dans la recherche d'informations en intégrant un module d'aide à l'interprétation qui, permettra à l'utilisateur un accès au plein texte, en lui proposant des scénarii textuels vu sous l'angle des scénarii modaux susceptibles de le guider vers des passages textuels répondant à ses requêtes en fonction de sa pratique en cours.

Abstract.

Study of linguistics Modality to assistant interpretation of legal texts.

One of the main objectives set by the designers of databases is to allow an easy and fast access to multiple documents and information either in a targeted area or in different areas. These digital resources must also consider and effectively address the practical needs of their future users. However, access to documents of a digital corpus raises many questions, including research information. The latter sees often reduced to a restricted network of keywords, which does not respond in any case the requirements of users. We propose here a linguistic approach in the search for information by integrating a help module for interpretation that will allow the user access to the full text by guiding and proposing textual scenarios that meet their requests and taking into account its social profile.

Mots-clés : Scénario modal, sémantique interprétative, traitement de texte, base de données..

Keywords: Modal script, Interpretive Semantic, word processor, database.

1 Introduction

A l'heure où les technologies numériques permettent un accès aux textes et aux corpus, nous questionnerons les interactions entre linguistique et informatique que résume imparfaitement l'appellation : « dialogue Homme-machine » puisqu'il serait plus juste de parler de « modèle de dialogue informatique », laissant perceptible la programmation sous-tendue.

Il devient en effet possible à la fois de ne plus encapsuler l'interprétation textuelle dans des réseaux sémantiques préexistants (réseaux de mots-clés), mais également de prendre pleinement en compte les contextes d'interprétations liés aux vécus des utilisateurs, à leur connaissance et à leur pratique sociale guidant leur parcours interprétatif. Dans ce contexte, nous cherchons à contribuer à une aide à l'interprétation des textes juridiques numérisés dans la base de données de l'Institut du Droit International des Transports (IDIT) basé à Rouen. Cette aide à l'interprétation est appréhendée par des approches linguistiques s'inscrivant dans le domaine de la sémantique temporelle et modale et de la sémantique textuelle. A partir d'un mémoire d'étude de Master 2 (2014), dans lequel nous avons étudié des fiches numériques résumant les arrêts de la cour d'appel dans le domaine du droit du transport, nous avons mis au jour un scénario qui de par ses contraintes modales est susceptible de guider l'interprétation des utilisateurs de la base de l'IDIT. Ce scénario modal inspiré de la théorie modulaire des modalités de (Gosselin, 2010) et de la notion de parcours interprétatif de (Rastier, 2001) : « (...) se manifeste aussi, bien comme arrière-plan de la signification de certains lexèmes, comme comportement stéréotypique des individus dans une société, et comme structure sous-jacente à certains textes (...) mais aussi [comme] trame textuelle et

argumentative des arrêts de cours d'appel» (Gosselin, 0000).

Ce travail se poursuit depuis la rentrée 2014 par une recherche doctorale¹ intitulée : « *Aide à l'interprétation : apport des maqueurs modaux à l'appréhension d'un genre textuel : application au discours juridique* » Nous nous focaliserons ici sur les contraintes modales et temporelles représentant des scénarii réguliers de la pratique juridique à l'œuvre dans ces textes. L'objectif de cet article consistera à montrer l'importance d'une analyse linguistique préalable à la conception d'une plate-forme d'aide à l'interprétation dédiée à cette pratique textuelle. Le but est d'adapter un système d'aide à l'interprétation aux besoins pratiques des utilisateurs lambda sur cette base de données. Nous présenterons tout d'abord le cadre général de notre étude et la problématique centrale de notre recherche, puis la notion du scénario modal telle qu'elle a été introduite par (Gosselin, 0000). Nous illustrerons alors notre démarche à partir d'un exemple. Nous pourrions alors conclure à une appréhension de la modalité comme faisceau de contraintes pouvant peser sur le parcours interprétatif d'un utilisateur et donc sur ses traces de navigation.

2 Contexte de la recherche

Notre étude s'inscrit dans le cadre général du projet PlaIr 2.0 (Plate-forme d'Indexation Régionale) qui se développe en partenariat avec l'IDIT², l'Université et l'INSA de Rouen. Cela fait plusieurs années que des chercheurs informaticiens et linguistes (Ferrari *et al.*, 2012) des deux laboratoires du LITIS et du DYSOLA tentent d'améliorer les interactions entre les abonnés de cette base jurisprudentielle et le système d'information en vue de répondre aux besoins réels des différentes catégories professionnelles (juristes, assureurs, transporteurs, étudiants en droit ...).

Les recherches au sein du projet PlaIr 2.0 ont été jusqu'à aujourd'hui essentiellement menées du point de vue ingénierie des systèmes, mais sans qu'aucun personnel (stagiaire de M2 ou doctorant) ne soit spécifiquement chargé d'une recherche linguistique (à l'inverse des stagiaires, ingénieurs, doctorants qui ont eu pour mission les développements informatiques de cette plate-forme). Notre travail est de ce point de vue sans précédent.

2.1 Le corpus de travail

Notre corpus de travail nous a été fourni par l'institut du droit international des transports qui donne accès à une base de données contenant environ 40.000 références organisées en 4 groupes :

- jurisprudences
- articles
- réglementation
- fonds documentaire

Notre étude porte sur le fond numérisé des arrêts de jurisprudence (tribunal de commerce, cour d'appel, cour de cassation). Ces arrêts sont accompagnés de fiches présentées comme des réécritures plus abrégées de ces arrêts. Elles sont classées par numéro et répondent toutes à la même structuration. Dans la perspective de mener une étude approfondie sur les textes pleins, nous avons choisis d'explorer tout d'abord les fiches, supposées être représentatives du contenu des arrêts. Cette pré-analyse sur les fiches nous a permis d'expérimenter la notion de scénario modal.

3 La notion de scénario modal

Notre hypothèse est que l'enchaînement des actions répétées effectuées par un sujet dans le cadre d'une pratique constituerait une sorte de mémoire actancielle à l'image d'un apprentissage. Afin de rendre compte de l'organisation schématique de cette expérience mémorielle, comme suite organisée de tâches sur le modèle d'un script, une notion s'avère être centrale : le scénario.

Le concept de scénario (script, schéma, ou cadre de connaissance) est apparu dans des travaux de l'intelligence artificielle (Schank & Abelson, 1977) dans le but de modéliser les comportements humains. En informatique, il peut être défini

1. Thèse régionale financée dans le cadre du GRR CSN axe 6 interdisciplinarité et technologies numériques en SHS : apport linguistique au projet PlaIr 2.

2. La base de données de l'IDIT se propose de répondre à toutes les questions d'ordre juridique, économique ou technique, intéressant les transports de quelque nature qu'ils soient, tant sur le plan national qu'international.

comme « *une succession ordonnée d'un ensemble d'opérations (. . .) [ou] un processus qui se déroule suivant un plan défini* » (Saidali, 2002). En ingénierie informatique, cela suppose d'emmagasiner les connaissances des experts liées à la façon d'agencer et d'utiliser des outils pour atteindre un but, notamment dans la construction de plate-forme de traitement d'images comme ce fut le cas pour la thèse de (Saidali, 2002). La notion de scénario est alors nécessairement à visée pragmati

- la définition et la catégorisation du contexte
- la décomposition de l'intention en sous-intentions (à l'instar d'un programme informatique)
- l'élabue. Un scénario contiendra dans ce cas : oration d'une méthode décrite par étapes
- l'enchaînement des étapes
- le choix et le réglage des paramètres pour chaque méthode
- la perception et l'évaluation des résultats (ibid)

En linguistique, l'idée de scénario comme « représentation d'une situation » trouve son origine chez Ch. Fillmore dans sa sémantique des frames (ou sémantique des cadres) (Fillmore, 1982) qu'il introduit pour décrire des enchaînements stéréotypiques d'actions, de situations et d'événements de la vie quotidienne . Pour (Fillmore, 1982), il s'agit d'abord de décrire les propriétés des cadres auxquels renvoient les verbes avant de déterminer leurs contenus lexicaux. Dans cette perspective, ce sont les événements et les situations dans lesquels s'inscrivent les mots qui déterminent leurs significations. Tel est l'exemple d'une scène de transaction commerciale qui évoque un ensemble de lexèmes : vendre, vente, achat, acheter, payer, facture, vendeur, acheteur etc. Chacun d'eux renvoie à un aspect différent du cadre mais s'y rapporte nécessairement. Vendre par exemple met en relief les éléments : vente/vendeur, facture met en avant les lexèmes achat/acheter etc. Cette mise en relief renvoie à ce que R. Langacker ((1987) cité par (Gosselin, 0000) désigne par « profilage » et qui représente un arrière-plan conceptuel où le lexème profile des éléments. Pour comprendre par exemple la signification du concept de l'hypoténuse, il est impératif de connaître le concept de triangle rectangle et qui exige à son tour d'avoir la connaissance de triangle et d'angle droit.

En s'inspirant des travaux de Ch. Fillmore sur les Frames, mais également de la théorie des modèles comportementaux de R. Jackendoff (2007) et de la sémiotique narrative de A.J. Greimas (1976, 1983) cité dans (Gosselin, 0000), L. Gosselin propose d'introduire la notion de scénario dans le cadre de sa Théorie Modulaire des Modalités (2010) sous le nom de « scénario modal ». Ce dernier « (...) se manifeste aussi, bien comme arrière-plan de la signification de certains lexèmes, comme comportement stéréotypique des individus dans une société, comme structure sous-jacente à certains textes (...) » (Gosselin, 0000).

Sans retenir l'idée d'un profilage ni celle d'un *a priori* conceptuel notre étude s'en tiendra à la trame textuelle et l'impact d'un réseau de contraintes modales sur l'interprétation des utilisateurs, En effet, l'intérêt de l'approche de L. Gosselin est qu'elle ne cherche pas à définir une liste de marqueurs spécifiquement dévolue à l'expression de tel ou tel calcul de sens (approche de nature réductionniste) mais contribue à une prise en compte globale (basée sur une conception large des modalités) des phénomènes lexicaux et grammaticaux constitutifs de la cohérence textuelle et de la construction du sens en contexte. De ce point de vue, l'étude des modalités et la formulation de scenarii modaux présentent un intérêt considérable pour la constitution d'une aide à l'interprétation dans un corpus numérique.

4 Démarche de l'analyse du corpus

Dans cette section nous présentons en grands traits la démarche adoptée pour l'analyse du corpus illustré d'un exemple d'une fiche résumant un arrêt de la cour d'appel de Rouen. Nous avons tout d'abord commencé par relever et classer dans un tableau les types de modalités qui caractérisent le texte contenu dans la zone « objet » de la fiche. À l'aide de ses modalités, nous avons retracé le scénario modal qui nous a permis par la suite de mettre en relief la structure interprétative des textes.

4.1 Extrait d'une fiche de la base de données de l'Idit

Référence :

Cour d'appel de Rouen, 2e ch.

17 juillet 2003

Sté P & O Nedlloyd BV et autre c/ Sté Générale de Manutention Portuaire et autres Navire "Bunga Pelangi"

Objet :

Transport d'un lot de cartons de vin empotés dans un conteneur réfrigéré - Température pré-réglée à -18°C au lieu de $+12^{\circ}\text{C}/+14^{\circ}\text{C}$ - Marchandise livrée congelée - Recours du transporteur-affrèteur contre le manutentionnaire et le frèteur - Responsabilité du manutentionnaire (oui) - Preuve de ses manquements aux obligations fixées par le contrat de manutention - Défaut de branchement du conteneur sur le terminal et de vérification de l'indexation de la température - Responsabilité du frèteur (non) - Contrat d'affrètement d'espace - Clauses - Frèteur responsable de l'indexation de la température des conteneurs (non) - Température pré-réglée par l'affrèteur - Connaissance de la température requise par le frèteur (non) - Obligations du frèteur limitées à la fourniture de l'alimentation électrique et à la vérification du bon fonctionnement des conteneurs.

Relevons que les périphrases de la zone objet sont construites sur le principe de l'« Economy Grammar » (Halliday, 1967) principe qui participe d'un mouvement de décontextualisation bien connu en documentation. La qualification juridique des faits, fondée sur un principe dichotomique (oui/non) intervient juste après leur rappel : Responsabilité du manutentionnaire (Oui) » alors qu'elle n'apparaît bien entendu que dans les conclusions de l'arrêt de la cour.

4.2 Relevé des modalités

Afin de pouvoir retracer l'enchaînement des modalités et d'établir un scénario modal pour la situation évoquée, nous devons tout d'abord relever les principaux marqueurs modaux qui jouent un rôle dans l'argumentation. Rappelons que les textes étudiés sont une première réécriture des arrêts de cours d'appel présentés comme une suite de mots-clés. Les textes se caractérisent par des modalités intrinsèques (associées ou dénotées aux lexèmes) et/ou des modalités inférées sur la base de connaissances encyclopédiques (Gosselin, 2010). Le tableau ci-dessous présente les marqueurs modaux mis en gras et les types de catégories modales exprimées, ainsi que les valeurs.

Modalités				
Les marqueurs modaux	catégories	Valeurs	Obtenues par inférence	valeurs
Température pré-réglée à -18°C au lieu de $+12^{\circ}\text{C}/+14^{\circ}\text{C}$.	Appréciative	Négative		
Marchandise livrée congelée .			Appréciative	négative
Recours du transporteur-affrèteur contre le manutentionnaire et le frèteur	Aléthique	positif	Boulque	positive
Responsabilité du manutentionnaire (oui)	Déontique	Positif		
Preuve de ses [manquements aux obligations] fixées par le contrat de manutention.	Aléthique (preuve) Axiologique (manquements) Déontique (obligations)	Positive négative positive		
Défaut de branchement du conteneur sur le terminal et de vérification de l'indexation de la température	Axiologique	négatif		
Responsabilité du frèteur (non)	Déontique	négative		
Frèteur responsable de l'indexation de la température des conteneurs (non)	Déontique	négative		
Connaissance de la température requise par le frèteur (non)	Epistémique	négative	déontique	négative
Obligations du frèteur limitées à la fourniture de l'alimentation électrique et à la vérification du bon fonctionnement des conteneurs	Déontique. Appréciatif	Positive positive		

4.3 Le scénario modal

Légende :

X = La société appelante / Y1 et Y2 = les intimées / Z = représentant de la loi

E1 = Transport d'un lot de cartons de vins empotés d'un conteneur réfrigéré.

Sit1 = température pré réglée -18°C.

Sit2 = marchandise congelée.

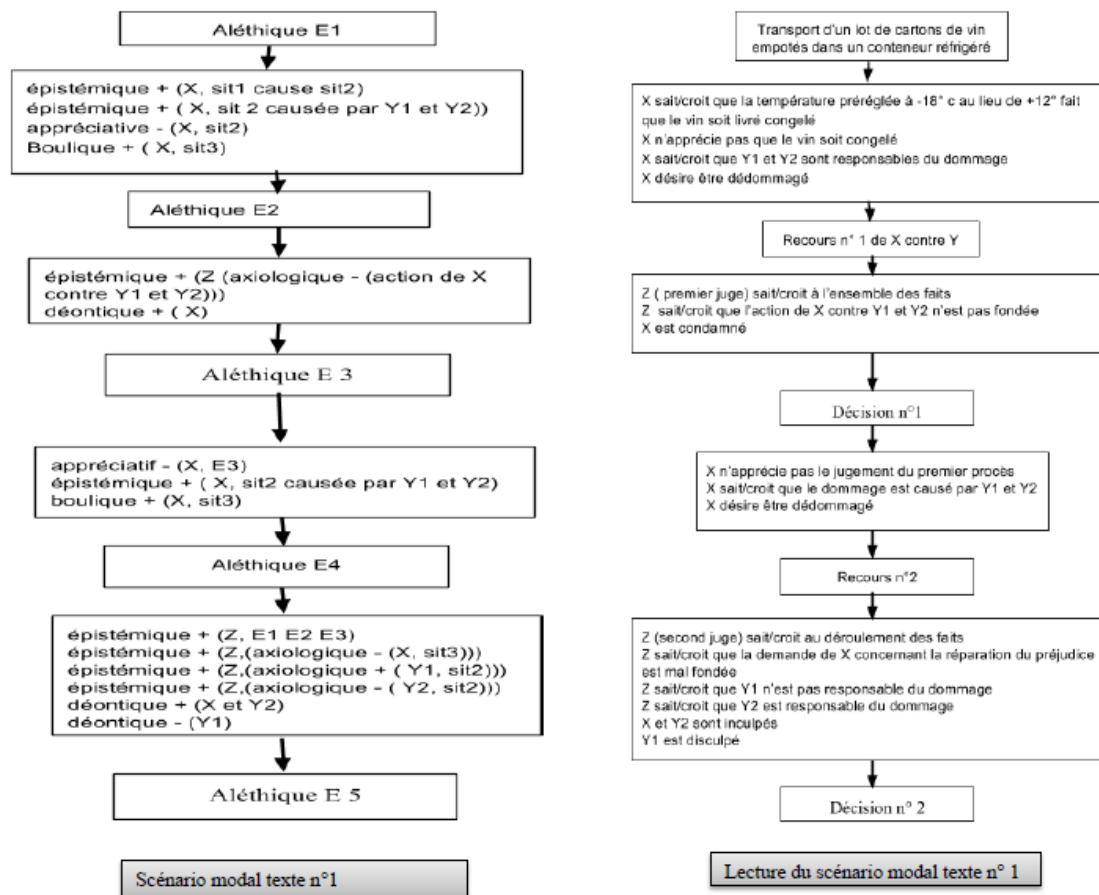
E2 = Premier recours de X contre Y1 et Y2

E3 = Décision du premier procès

E4 = Deuxième recours de X contre Y1 et Y2

E5 = Décision du second procès

La présentation est faite en deux plans. Nous avons dans le côté gauche de la figure le déroulé modal organisé grâce aux marqueurs modaux relevés dans le texte. A droite nous avons mis en correspondance les passages textuels correspondants. Cette démarche nous a permis de relever la structure chronologique et argumentative de la zone objet de la fiche supposée être celle du texte plein qu'elle représente.



4.4 Commentaires

Dans le côté gauche de la figure ci-dessus nous apercevons une présentation chronologique et narrative du contenu des arrêts étudiés. Cette chronologie a été mise au jours grâce à des marqueurs modaux qui témoignent de l'ordre dans lequel les événements sont successivement présentés. A droite nous avons repris les passages textuels figurants sur la fiche et nous les avons organisés selon le modèle du scénario. Ce qui ressort de cet enchaînement modal est le protocole discursif mis en place dans l'argumentation juridique. Dans cet arrêt de cour d'appel, le juge reprend d'abord la description des

faits du litige (illustré dans le scénario par les modalités aléthiques), suivi de leur appréciation (négative ou positive, en fonction de l'orientation du juge (condamnation ou acquittement). Les appréciations sont désignées dans le scénario par les modalités appréciatives. Ces dernières sont suivies de jugements d'ordre légal sur les faits (exprimer ce qui est blâmable ou louable) on parle alors de modalités axiologiques. Ce n'est qu'à ce stade du parcours que le juge prononce sa sentence (dire l'interdit et ou le permis) ce qui renvoie au dernier bloc du scénario, les modalités déontiques.

5 Conclusion et perspective

Nous avons présenté dans cet article la notion de scénario modal avec laquelle nous avons reconstitué le possible parcours interprétatif d'un arrêt de cour d'appel. Les premiers résultats obtenus à partir de ce corpus de fiches laissent entrevoir une possible implémentation dans un système d'aide à l'interprétation. À partir du scénario modal nous formulons l'hypothèse de l'existence au sein des textes pleins d'une structuration en « zones d'isotopies³ modales ». Ces zones, une fois délimitées, sont susceptibles de contenir des informations différentes pouvant répondre à des requêtes précises formulées par les utilisateurs. Afin de conduire un utilisateur vers une zone pertinente, nous envisageons un système d'annotation des marqueurs modaux spécifiques à certaines zones de textes. Nous pensons également possible, avec une étude approfondie sur les textes pleins, de proposer différents chemins de lecture et donc d'interprétation correspondants à la pratique en cours de l'utilisateur. L'un de nos objectifs est de valider cette étude de scénario modal sur un corpus de textes pleins. Pour faire cela, nous procédons actuellement à l'aide d'un logiciel textométrique (TXM)⁴ à une étude statistique afin de relever toutes les caractéristiques linguistiques comparatives entre trois corpus de textes (cour d'appel, cour de cassation et tribunaux de commerce). Cette démarche nous permettra de cerner leur structuration spécifique et d'y adapter notre aide à l'interprétation supposant avoir affaire à des pratiques sociales différentes, malgré la proximité des genres. Une fois ce travail accompli, nous mènerons une enquête qualitative auprès d'un panel d'utilisateurs de la base de l'IDIT dans le but de cerner leur parcours de navigation et de le confronter à notre scénario modal. Nous pourrions alors mieux comprendre leurs besoins réels et apprécier la portée de nos résultats encore théoriques au regard des traces de leur navigation. Le lecteur humain est évidemment le meilleur interprète d'un document, par comparaison à tout traitement automatique et machinal, qui n'est d'ailleurs jamais une véritable lecture.

Remerciements

Les travaux menés dans cet article bénéficient d'un financement du Grand Réseau de Recherche Culture & Société en Normandie (CSN) de la Région Haute-Normandie. Et nous avons le soutien du projet PlaIR 2,0 (2013-2016) pris en charge par le GRR Transport Logistique Technologies de l'Information de la Région Haute-Normandie.

Références

- FERRARI S., MAUREL F., BEUST P., MAUGER S., HOLZEM M., BAUDOUIN N., TRUPIN E., SAIDALI Y., LABICHE J. & DIONISI D. (2012). Pour une recherche d'information et une veille juridique interactives et socio centrées. ENT éactif & veille en droit du transport. *Ingénierie des Systèmes d'Information*, 2, pages 17–40.
- FILLMORE C. (1982). Frame semantics. in linguistics in the morning calm. *Revue Ingénierie des systèmes d'information - RSTI série ISI*, 17, 111–137.
- GOSSELIN L. (0000). à paraître : De l'opposition modus/dictum à la distinction entre modalités extrinsèques et modalités intrinsèques. *Bulletin de la Société de Linguistiques de Paris*, p.35.
- GOSSELIN L. (2010). *Les modalités en français : la validation des représentations*. Rodopi.
- HALLIDAY M. A. K. (1967). Notes on transitivity and theme in English : Part 2. *Journal of Linguistics*, 3(02), 199–244.
- RASTIER F. (2001). *Arts et sciences du texte*.
- SAIDALI Y. (2002). *Modélisation et acquisition de connaissances : application à une plate- forme de traitement d'images*. Thèse d'informatique automatique et systèmes. PhD thesis, Université de Rouen.

3. La notion d'isotopie est définie par F. Rastier (1972) comme l'effet de la récurrence d'un même sème, dit sème isotopant.

4. Un logiciel gratuit de traitement statistique des textes qui se développe dans l'ENS de Lyon.

SCHANK R. C. & ABELSON R. P. (1977). *Scripts, Plans, Goals, and Understanding : An Inquiry Into Human Knowledge Structures*. Psychology Press.

Interprétation rapide et non visuelle de pages Web : TAL et Tonnerre de mots

Jean-Marc Lecarpentier Fabrice Maurel Stéphane Ferrari Alexandre Beudin
GREYC, CNRS, Campus Côte de Nacre, Boulevard du Maréchal Juin, 14032 CAEN cedex 5
{firtsname.lastname}@unicaen.fr

Résumé. L'accès rapide à un contenu précis dans une page web est généralement réalisé par une étape dite d'écrémage (ou *skimming*) pour avoir une vue globale, suivi d'une stratégie de balayage (ou *scanning*) pour repérer les informations voulues. Ces deux étapes naturelles pour une navigation visuelle n'existent pas dans la cas d'un accès non visuel à la page donnée. Nous présentons dans cet article le travail en cours sur le concept de tonnerre de mots (ou *tag thunder*), traduction à l'oral du concept de nuage de mots. Après avoir segmenté une page web en blocs, les mots-clés significatifs sont extraits de chaque bloc et pondérés afin d'obtenir un nuage de mots. Parmi ces mots, les mots-clés navigants permettent à l'utilisateur de naviguer en profondeur dans les divers blocs de la page. L'ensemble est transformé en tonnerre de mots par vocalisation. La présentation de ce travail en cours permet aussi de présenter les diverses pistes de recherche envisagées.

Abstract.

Tag thunder : providing skimming and scanning techniques in non visual situations

Skimming and scanning are two different strategies for *speed reading* and are used when accessing web pages. Skimming is used when we discover a page and it allows a reader to get a first glance of its content. Scanning usually follows, in order to get a more precise idea of the content or to find a specific content. However, these techniques are not available in non visual situations. This article introduces the concept of tag thunder, which aims to provide speed reading techniques similar to skimming and scanning to visually impaired people. A tag thunder is the oral transposition of a tag cloud, where keywords are presented with typographic effects depending on the word's importance. Within a tag thunder, each keyword is pronounced differently to enable the user to perceive its importance. Navigation keywords enable the user to select a corresponding block. This work in progress is based on three steps. First the web page is segmented to extract its main blocks. Then, for each block, keywords are extracted to create tag clouds. Finally, tag clouds are converted to tag thunders and presented to the user. Our solution is implemented within a Firefox extension and opens several research tracks.

Mots-clés : accessibilité, oral, tag thunder, nuage de mots, extraction, segmentation.

Keywords: accessibility, tag thunder, tag cloud, extraction, segmentation.

1 Introduction

La plupart des internautes découvrant une page web réagissent selon le même schéma : obtention d'une vue globale de la page suivie d'une recherche active et rapide d'informations spécifiques. La première étape, dite stratégie d'écrémage ou *skimming* permet d'appréhender les différentes zones de la page, leur disposition, etc. La seconde étape dite stratégie de balayage ou *scanning* permet à l'internaute de déterminer si la page consultée correspond effectivement à ses attentes en terme de contenu. Dans les deux cas, la réussite et la rapidité de ces stratégies dépendent fortement de la structuration de la page en zones bien identifiées (Dias & Conde, 2007), de la typographie et de la mise en forme des contenus (titres, éléments graphiques, zones distinctes, couleurs, etc). Pour les internautes en situation non visuelle, ces stratégies ne sont pas possibles (Ahmed *et al.*, 2012b) : par exemple la figure 1 montre ce qu'un voyant et un non-voyant perçoivent au premier abord après le chargement d'une page web. Des stratégies palliatives non visuelles existent comme par exemple l'accélération du débit pour les *screen readers*¹ (Borodin *et al.*, 2010), les commandes permettant de sauter d'un titre ou d'un lien à un autre, la lecture du début et fin de chaque paragraphe, etc. Malgré cela, ces stratégies sont encore loin d'égaler celles d'un utilisateur voyant (Bigham *et al.*, 2007).

1. Lecteur d'écran : système de synthèse de la parole couplé à une plage Braille (par exemple logiciel JAWS)

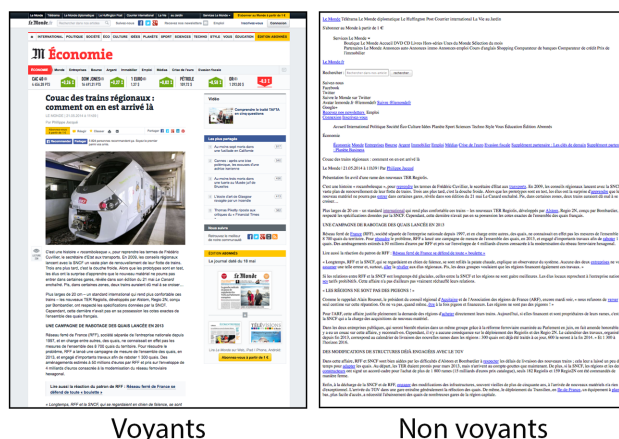


FIGURE 1 – Comparaison entre le visuel d'une page web et sa perception par un non-voyant avec les lecteurs d'écran

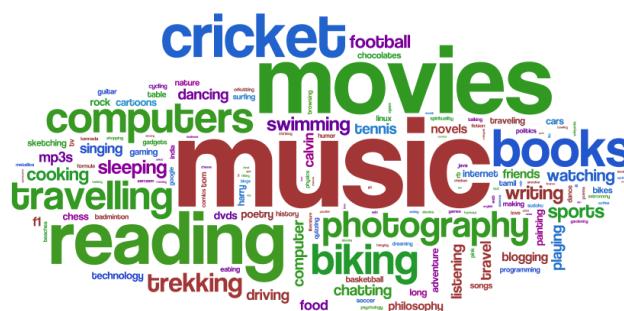


FIGURE 2 – Exemple de nuage de mots ou *tag cloud*

Cet article se focalise sur les possibilités d'accès non visuel, global et rapide à l'information. Nous proposons ici la transposition à l'oral du concept de nuage de mots (*tag cloud*) qui représente un ensemble de mots-clés plus ou moins mis en avant visuellement, comme le montre la figure 2². À l'instar des saillances dispositionnelles et typographiques du *tag cloud*, cette stratégie, appelée tonnerre de mots (ou *tag thunder*), veut organiser les mots en produisant des saillances spatiales et « typophoniques ».

La solution proposée dans cet article utilise des techniques de recherche d'information et de traitement automatique de la langue pour segmenter la page web en blocs, déterminer les mots-clés de chaque bloc et leur importance et enfin utilise un générateur vocal pour le rendu oral. Cette proposition est articulée comme suit. La partie suivante présente des travaux existants pour faciliter l'accès aux contenus dans des situations non visuelles. Nous présentons ensuite les concepts de tonnerre de mots et de mot-clé navigant, ainsi que les divers paramètres à prendre en compte. La partie suivante décrit les stratégies utilisées pour la segmentation et l'extraction des mots-clés qui constituent les tonnerres de mots. Elle est suivie d'une présentation de l'architecture logicielle qui implémente notre solution. Enfin, la dernière partie présente le travail restant à réaliser et les perspectives envisagées.

2 État de l'art

Des stratégies de *skimming* pour les non-voyants utilisent des techniques de résumé de texte (Ahmed *et al.*, 2012a). La page web est tout d'abord transformée en une séquence linéaire de texte puis des techniques classiques de sacs de mots permettent d'extraire des unités sémantiques. Cependant, cette approche montre ses limites pour la perception rapide des contenus de pages Web : la phase de linéarisation ne conserve pas le rôle important de la typo-disposition pour une bonne appréhension des contenus et la mise en place d'une boucle perception/action efficace (Dias & Conde, 2007). D'autres

2. Image de Anand S, <https://flic.kr/p/5BFE3V>, licence CC BY 2.0

propositions ont tenté d'intégrer cette observation et proposent des solutions avec différentes modalités d'accès au texte et à sa mise en forme (synthèse de la parole, retours tactiles ou vibrotactiles). La suite de cet état de l'art présente les principaux travaux qui ont nourri notre réflexion.

Pour améliorer la sensibilité des systèmes de synthèse de la parole à partir de textes (TTS), (Maurel *et al.*, 2003) proposent un modèle d'oralisation par reformulation en l'appliquant à divers phénomènes syntaxiques, typographiques ou dispositionnels, appelés Architecture Textuelle par les auteurs (Virbel, 1989). Les textes jugés les plus importants sont reformulés sous différentes formes discursives pour obtenir un premier jeu de stratégies orales. Ils sont ensuite réduits avec des marques prosodiques adéquates pour produire un nouveau jeu de stratégies qui sont alors plus compactes). Grâce à cette approche, une amélioration sensible de la mémorisation et de la compréhension est obtenue lors de l'oralisation par un TTS de documents fortement structurés. Cependant, (Maurel *et al.*, 2006) montre que la charge cognitive est plus difficilement gérée que lors d'une lecture visuelle de ces mêmes documents.

Le projet *Acessibility through Simplification & Summarization* (Parmanto *et al.*, 2005) (AcceSS) utilise des stratégies de simplification et de résumé. La simplification retire de la page les contenus jugés moins importants, donnant ainsi un accès plus rapide au contenu jugé important. Le résumé, quant à lui, donne un aperçu du contenu similaire à celui perçu visuellement. Le principe de base est de créer une nouvelle page pour chacune des sections (tête-de-page, bas-de-page, menu gauche, contenu textuel...). Pour cela, une méthode de marquage, basée sur des « modèles de genres » (presse, e-commerce, blog...) ou d'heuristiques, génère un ensemble de repères placés sur les éléments de la page. Un point central de navigation (page *Guide Dog*) est créé pour servir de pivot et de sommaire entre toutes les pages créées. Une première expérimentation avec le logiciel JAWS montre des résultats positifs. L'accès à une information particulière est plus rapide et les utilisateurs ont apprécié la navigation dans la page *Guide Dog* et plus généralement la version du site transformée par AcceSS. Une des limites de cette solution est l'utilisation d'une méthode de pattern matching pour identifier les sections de la page selon le genre qui est encore faillible. D'autre part, il n'y a pas de simplification au niveau du texte, donc une difficulté toujours grande dans le cas de contenus textuels importants.

SeEBrowser (Semantically Enhanced Browser) est un navigateur web audio spécialisé pour non-voyants (Michail & Christos, 2007). Des annotations sémantiques permettent une navigation par raccourcis. Des ontologies sur les pages web sont tout d'abord définies par annotation manuelle. Elles définissent ainsi des relations entre les différents éléments de ces pages et correspondent à la hiérarchie de la page web. L'ontologie produite est alors utilisée comme un arbre de navigation à plusieurs niveaux de granularité. Les nœuds sont appelés *Browsing Shortcuts* (BSs). Lorsque l'utilisateur charge une page web, il peut demander la liste de ses BSs puis les parcourir pour les écouter via des raccourcis claviers. Si le BS courant contient d'autres BSs, un son spécifique est émis et accessible. L'expérimentation montre que les BSs offrent un gain de temps statistiquement significatif pour des tâches de recherche d'information (scanning) et que l'utilisation de la hiérarchie produite évite la surcharge d'information. Le temps pour trouver une information reste cependant très long car les utilisateurs écoutent tous les BSs avant d'en sélectionner un. Une piste de solution serait de réordonner les BSs par ordre d'importance et de ne garder que les plus importants, mais l'inconvénient étant alors une perte d'information.

HearSay (Borodin *et al.*, 2010) est un navigateur web non visuel multi-modal développé à l'Université de Stony Brook (New York, USA) depuis 2004. Il accepte en entrée la reconnaissance vocale, le clavier, les interfaces tactiles et permet des sorties audio, écran et braille. Ses fonctionnalités sont nombreuses :

1. un module de segmentation analyse la structure et la mise en forme géométrique des pages web afin de les segmenter. Les utilisateurs peuvent alors naviguer entre les objets textuels reliés sémantiquement, par exemple le titre d'un article et son résumé. Le système utilise une variation de l'algorithme VIPS de Microsoft (Cai *et al.*, 2004) ;
2. un système d'annotation pour que les utilisateurs puissent ajouter du texte alternatif aux images et autres contenus ;
3. des algorithmes de différence comparent les pages visitées par l'utilisateur et repèrent les changements. Cela permet de passer rapidement sur les informations répétées (comme les menus) mais aussi d'obtenir les mises à jour dynamiques d'une page web ;
4. un analyseur de contexte qui permet de détecter le contenu principal d'une page web (par exemple le début d'un article) et d'identifier les informations pertinentes pour l'utilisateur en utilisant le contexte présent autour du lien cliqué.

L'expérimentation de HearSay montre une amélioration significative du temps nécessaire à des non-voyants pour trouver le contenu principal d'une page grâce au module d'analyse de contexte. Le module *Dynamo* aide fortement les utilisateurs à localiser les changements dynamiques de contenu et leur évite de lire les informations répétées (menus par ex.), permettant ainsi de se concentrer sur le contenu principal. Globalement, les autres fonctionnalités proposées par le logiciel ont aussi apporté un confort de navigation aux utilisateurs. Malgré ces points positifs, deux aspects restent en deçà de la

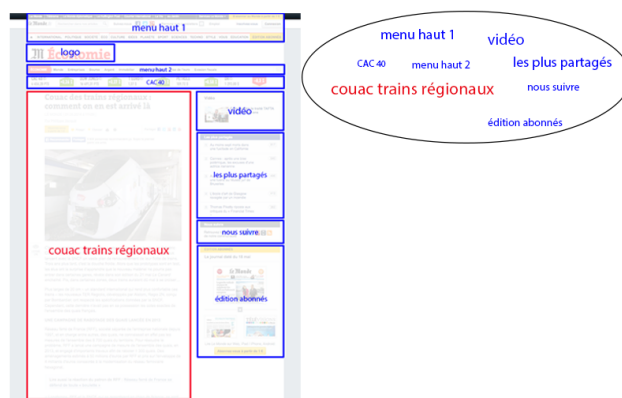


FIGURE 3 – Exemple de segmentation d'une page en blocs et création d'un tag cloud

lecture visuelle. Tout d'abord l'accès à la structure de la page n'est pas complet puisqu'il ne permet d'atteindre que le contenu principal ou jugé pertinent selon le contexte. Les autres éléments de la page restent accessibles séquentiellement, ce qui peut être long et laborieux. Ensuite, le *skimming* n'est pas permis sur la page, et le manque d'accès rapide et efficace à l'information constitue à nos yeux le deuxième aspect manquant à HearSay.

De plus en plus de travaux utilisent désormais des stratégies tactiles (Ziat *et al.*, 2007) (Giudice *et al.*, 2012) (Ahmed *et al.*, 2011). En intégrant des patterns dans les pages web, (Issa *et al.*, 2010) rend perceptible au toucher certains objets de la page et les relations qu'ils entretiennent. Les documents transformés sont alors proposés aux non-voyants sur un papier thermogonflable qui les met en relief. En positionnant le papier sur un écran tactile, on peut alors demander l'oralisation des objets présentés. L'inconvénient de cette méthode étant la nécessité d'avoir à disposition un papier spécial, (Safi *et al.*, 2014a) propose un concept proche utilisant la perception vibrotactile avec un dispositif constitué d'un gant et d'une tablette. Les contrastes lumineux produits par la tablette sont captés au survol des zones et génèrent des vibrations dans le gant. Ces approches sont intéressantes pour les stratégies de *scanning* mais sont insuffisantes pour véritablement réaliser du *skimming*.

Malgré les nombreux travaux, les stratégies de *skimming* doivent encore être améliorées pour permettre avoir un aperçu rapide du contenu d'une page web dans les situations non visuelles. La partie suivante présente les concepts de « tonnerre de mots » et de mot-clé navigant pour accéder au contenu d'une page web.

3 Tonnerre de mots ou tag thunder

Le *tag thunder* est un concept original reposant sur la capacité de l'être humain de pouvoir concentrer son oreille sur une source de son parmi d'autres émettant du son simultanément (Cherry, 1953). Le *Cocktail Party Effect* (Guerreiro, 2013) en est un exemple. On peut concentrer son attention sur une personne alors même que plusieurs personnes parlent simultanément autour de nous. La prononciation d'un mot important par une autre personne, par exemple notre nom, attirera immédiatement notre attention. L'idée est donc la suivante : dans un tonnerre de mots, l'utilisateur sera capable de détecter les mots-clés qui l'intéressent.

Nous considérons une page web comme un ensemble de blocs. La figure 3 illustre cette division sur une page d'article de journal. Chacun de ces blocs peut alors être identifié par un mot ou une phrase représentant son type, sa structure (métadonnée) ou son contenu (mot-clé). L'ensemble de ces mots ou phrases forme ainsi une sorte de nuage de mots (*tag cloud*), c'est-à-dire une représentation visuelle simplifiée du contenu et de la structure de la page web.

Ces blocs peuvent eux-mêmes contenir plusieurs blocs. Par conséquent, ils peuvent être définis de façon récursive comme des ensembles de blocs. De nouveaux mots ou phrases peuvent alors être extraits pour former de nouveaux nuages de mots. La combinaison du premier nuage et de ses « sous-nuages » constitue ainsi un arbre de nuages de mots, comme le représente la figure 4.

Les nuages de mots restent cependant des éléments visuels. Nous introduisons l'idée de « tonnerre de mots » ou *tag thunder* qui est une sorte de nuage de mots oralisé. Comme les mots ou phrases d'un nuage de mots sont distribués

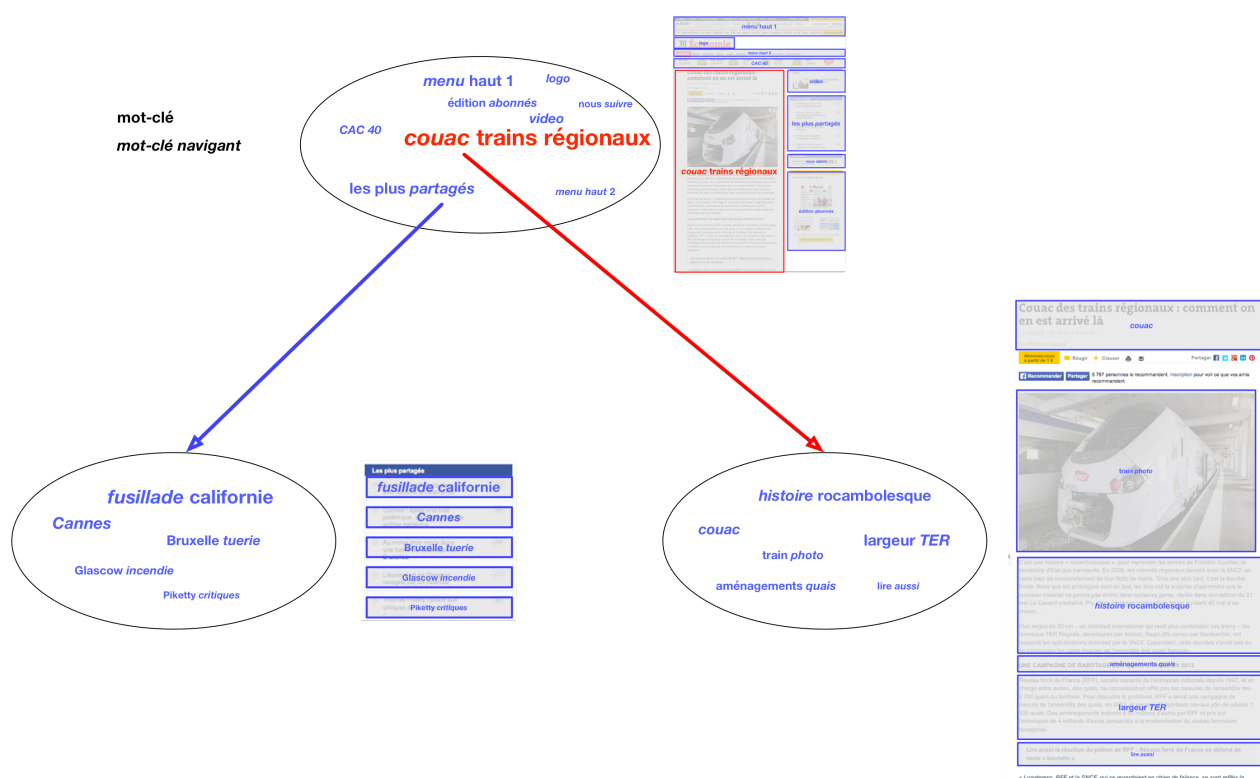


FIGURE 4 – Arbre de nuages et blocs correspondants

dans un espace avec différents effets visuels (e.g. taille de police, couleur, etc.), les mots constituant un tonnerre de mots sont distribués dans le temps avec différents effets audio (e.g., volume, débit, hauteur, prosodie, voix, répétition, multi-canaux). En utilisant une version orale de l'arbre présenté figure 4, on obtient alors un arbre de tonnerres de mots qui est une représentation de la structure de la page et de son contenu.

Nous voulons proposer une navigation dans cet arbre, permettant ainsi une navigation intra-page, c'est à dire une navigation en profondeur dans les blocs de la page. Le tonnerre de mots doit donc :

1. donner un aperçu du contenu du bloc textuel présenté (en fonction de la position dans l'arbre)
2. permettre la navigation dans l'arbre des blocs de la page

L'objectif final est que l'utilisateur puisse interagir avec le système en prononçant un mot-clé particulier entendu dans le tonnerre.

Nous définissons ainsi la notion de mot-clé navigant (MN ci-après). Un mot-clé navigant est un mot particulier du tonnerre de mots qui peut être « activé » vocalement. L'utilisateur peut ainsi prononcer le MN pour accéder au contenu correspondant au bloc identifié par le MN prononcé. Cela permet donc d'accéder au tonnerre de mots représentant un ensemble de sous-blocs s'il en contient, sinon directement à son contenu textuel (cf. figure 4). Les MN doivent être uniques, aisément identifiables et suffisamment pertinents pour représenter le contenu ou la nature de leur bloc. Il est donc nécessaire d'avoir non seulement des stratégies d'extraction de mots-clés, mais aussi des stratégies pour choisir les mots-clés navigants. Cependant, la notion d'unicité ne se réfère pas nécessairement uniquement au texte mais elle peut être liée au couple composé du texte et de sa prosodie.

Enfin, le tonnerre de mots peut être composé de mots-clés importants pour la compréhension du texte et de MN. Ainsi, l'utilisateur peut rapidement avoir un aperçu d'un bloc de texte et avoir les MN qui permettent à la fois d'appréhender le contenu et de naviguer. La figure 4 montre des exemples de noeuds composé de MN (en italique) et de mots-clés non navigants mais importants pour la compréhension du contenu. La partie suivante présente l'architecture de notre solution qui, pour l'instant, permet de créer des tonnerres de mots composés uniquement de MN.

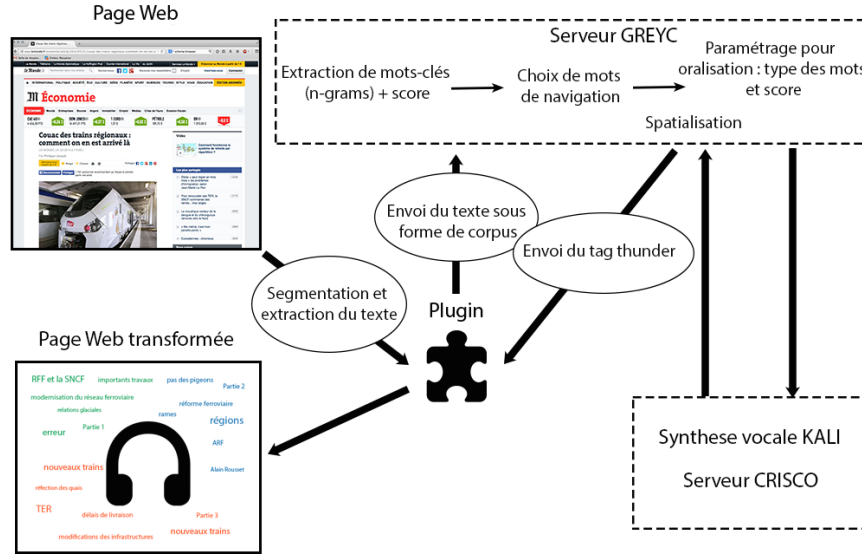


FIGURE 5 – Architecture logicielle

4 Architecture

Pour tenter de reproduire le processus d’aperçu rapide d’une page web qu’un utilisateur voyant réalise, nous devons identifier les diverses zones de la page. Puis pour chaque zone, il faut déterminer les mots importants ou représentatifs de cette zone. Cette partie présente les stratégies mises en oeuvre pour segmenter la page consultée et les techniques utilisées pour extraire un nuage de mots de chaque partie. Enfin, nous présentons comment le tout est oralisé pour être présenté à un utilisateur non-voyant.

L’architecture de notre solution est basée sur trois briques logicielles qui gèrent la segmentation de la page, l’extraction de mots-clés pour créer les nuages de mots et enfin l’oralisation des tonnerres de mots. Une extension pour le navigateur Firefox (FF ci-après) a été développée et gère les différentes briques. La figure 5 illustre l’architecture mise en place et les briques utilisées, détaillées ci-après.

4.1 Segmentation de la page en blocs

Différentes approches existent pour la segmentation de la page en blocs de texte. (Sanoja & Gançarski, 2014) propose une méthode basée sur l’analyse de l’arbre DOM (Document Object Model) représentant la page web. (Cai *et al.*, 2003) utilise l’apparence de la page dans le navigateur (approche basée vision) pour en extraire les zones. D’autres techniques existent : techniques de traitement d’image (Cao *et al.*, 2010), basées sur la structure sémantique (Adda, 2013) ou encore une résolution par graphe (Liu *et al.*, 2011). La segmentation est actuellement réalisée avec une méthode hybride que nous avons développée (Safi *et al.*, 2014b). Lorsque la page est chargée, l’extension FF applique l’algorithme de segmentation et envoie le résultat à un service web chargé de l’extraction de mots clés.

4.2 Extraction des mots-clés et nuages de mots

L’extraction de mots-clés est calculée en utilisant un algorithme basé sur TF-IDF (Term Frequency-Inverse Document Frequency). Cette mesure utilisée en Recherche d’Information, permet de faire ressortir les mots spécifiques à un contexte donné. En effet, la fréquence des termes dans le texte est pondérée par l’inverse de leur fréquence dans un corpus, ce qui fait remonter les plus spécifiques à ce texte, donc les plus pertinents.

Inspiré du travail de (Witten *et al.*, 1999), l’algorithme d’extraction de mots-clés développé ici s’appuie sur la mesure TF-IDF ainsi que sur la position des termes dans le texte pour calculer un score (e.g., dans le texte « bonjour tout le monde

», le mot « tout » est en 3^eme position sur 4, soit 0,75). Une version améliorée est proposée par (Jiang *et al.*, 2009) en utilisant le score TF-IDF et la position avec *Ranking SVM* plutôt que la position simple.

Le calcul de l’IDF (Inverse Document Frequency) a été réalisé sur un corpus de textes constitué de 953 551 articles sur 20 années du journal Le Monde (de 1987 à 2006). Le fait que le corpus s’arrête en 2006 introduit des biais et des silences dans l’extraction de mots-clés mais a permis de mettre en place une version fonctionnelle d’un prototype. Nous y reviendrons dans la dernière partie de cet article.

Un des points essentiels du travail de création du tonnerre de mots est la détermination des MN. Comme mentionné plus haut, ceux-ci doivent être uniques. Cette contrainte introduit une étape complémentaire à l’extraction des mots-clés. En effet, une fois les mots-clés extraits, il faut s’assurer de leur unicité et choisir d’autres MN en cas de conflit. Cette étape peut alors introduire des biais sur le choix des MN en « enlevant » des mots importants dans d’autres parties de la page. En particulier dans le cas où les blocs à décrire traitent du même sujet, la difficulté réside alors à obtenir des mots-clés suffisamment explicites mais distincts les uns des autres.

Une fois les nuages de mots-clés déterminés pour chaque bloc de la page, l’extension FF les reçoit alors et doit créer le tonnerre de mots. Cette étape est aussi réalisée via l’utilisation d’un web service.

4.3 Génération du tonnerre de mots

Des différences de voix, de volume, de prosodie, de hauteur ou de débit peuvent être utilisées dans un même tag thunder en simultanée étant donné que notre système auditif peut les discerner (Bregman, 1994). Leur synchronisation (Turgeon *et al.*, 2005) et leurs fréquences (Darwin *et al.*, 2003) peuvent influencer positivement ce discernement. Le travail de recherche consiste donc à définir et évaluer des modèles de *tag thunder*. L’audibilité et la rapidité de lecture du tonnerre de mots, ou encore la facilité d’identification des mots-clés d’intérêt par et pour l’utilisateur sont des critères d’évaluation possibles.

Notre solution intègre la synthèse vocale Kali (Morel & Lacheret-Dujour, 2001), conçue par le laboratoire du CRISCO (Université de Caen Basse-Normandie). Cinq étapes, prétraitement, analyse syntaxique, génération de la prosodie, phonémisation et traitement acoustico-phonétique permettent de prononcer le texte donné en entrée. Son intelligibilité et sa vitesse d’élocution élevée en font une synthèse vocale très appréciée des personnes non-voyantes. En effet, elles sont habituées à un débit élevé de parole, ce qui leur donne un gain de temps considérable lors de la lecture de documents. Kali est rendue disponible par une API prenant en paramètre le texte à synthétiser, la langue (français ou anglais), la voix, la prosodie, le volume, le débit et la hauteur. Le traitement est presque immédiat et elle retourne ensuite l’URL du fichier WAVE créé. L’extension FF envoie les textes à oraliser à l’API Kali, reçoit les URLs des sons et peut ainsi les jouer à l’utilisateur.

5 Perspectives et conclusion

Nous proposons ici une solution pour faciliter le *skimming* et le *scanning* de pages web dans des situations d’utilisation non visuelles. Après avoir segmenté la page en divers blocs, chaque bloc est analysé afin d’en extraire des mots clés navigants que l’utilisateur pourra énoncer pour sélectionner la sous-partie correspondante. Enfin, notre solution intègre un générateur vocal pour oraliser les MN de la page.

Nous travaillons actuellement à l’amélioration de notre solution, en particulier sur le corpus pour calculer. En effet, les articles les plus récents datant de 2006, un nombre non négligeable de mots sont inconnus du corpus. De plus, le corpus étant uniquement constitué d’articles de presse, le langage utilisé est propre à ce contexte. Une amélioration en cours est de baser le calcul de l’IDF sur un ensemble de pages issues du site consulté, éventuellement combiné à un corpus généraliste.

D’autres champs d’expérimentation sont ouverts par le fait qu’il s’agit non seulement d’extraire de simples mots-clés, mais de déterminer les mots-clés navigants. En effet, lorsque les divers blocs de la page traitent du même sujet, il faudra réussir à déterminer des MN distincts mais suffisamment précis pour ne pas avoir de biais dans les nuages extraits.

Une autre amélioration concerne l’utilisation des formats audio 5.1 et 7.1 afin d’utiliser plus de canaux pour la représentation spatiale d’une page web. L’utilisation de techniques binaurales est aussi envisagée³. La synthèse Kali utilisée n’est

3. Technique de spatialisation sonore. Voir http://fr.wikipedia.org/wiki/Enregistrement_binaural

Effet visuel	Effet sonore
disposition du texte	spatialisation du son
saillance typographique	volume et fréquence de répétition
navigation en profondeur dans les blocs	mot-clé navigant et type de voix

TABLE 1 – Équivalents entre effets visuels et effets sonores

pas compatible avec l'utilisation de langages normalisés d'oralisation. L'utilisation d'un TTS compatible permettrait alors de travailler directement à l'aide de formats VoiceXML+SSML.

De même que pour la génération des listes de mots-clés navigants, les paramètres possibles pour la restitution du *tag thunder* sont nombreuses. Comme mentionné plus haut, le débit, la prosodie, etc. sont divers facteurs sur lesquels nous devons mener de multiples expérimentation pour déterminer la meilleure solution pour qu'un utilisateur puisse faire son choix. En particulier, nous voulons tester le fait que la spatialisation du son reflète la spatialisation des textes, c'est à dire que des mots issus d'un même bloc de texte auront des caractéristiques spatiales similaires. De même, les effets typographiques constituent un champ d'expérimentation, pour lequel nous pensons jouer sur les paramètres de volume et de fréquence de répétition des mots. Le tableau 1 résume le parallèle entre les appréhensions visuelles et non-visuelles que nous voulons expérimenter.

En ce qui concerne les structures de données, la modélisation de ce que nous avons appelé ici l'arbres de nuages reste à formaliser afin de pouvoir utiliser des modèles et/ou algorithmes existants.

Enfin, la solution proposée doit être évaluée avec un protocole expérimental pour vérifier si les stratégies de *skimming* et de *scanning* non visuel sont améliorées. La combinaison de cette solution avec d'autres systèmes, en particulier vibrotactile, permettrait alors de concevoir des systèmes d'aides multimodaux.

Références

- ADDA N. F. S. R. G. (2013). Pré-segmentation de pages web et sélection de documents pertinents en questions-réponses. *TALN-RÉCITAL 2013*, p. 479.
- AHMED A. A., YASIN M. A. & BABIKER S. F. (2011). Tactile web navigator device for blind and visually impaired people. In *Applied Electrical Engineering and Computing Technologies (AEECT), 2011 IEEE Jordan Conference on*, p. 1–5 : IEEE.
- AHMED F., BORODIN Y., PUZIS Y. & RAMAKRISHNAN I. (2012a). Why read if you can skim : towards enabling faster screen reading. In *International Cross-Disciplinary Conference on Web Accessibility - W4A2012*, Article No. 39.
- AHMED F., BORODIN Y., SOVIK A., ISLAM M., RAMAKRISHNAN I. & HEDGPETH T. (2012b). Accessible skimming : Faster screen reading of web pages. In *25th Annual ACM Symposium on User Interface Software and Technology (UIST)*, p. 367–378.
- BIGHAM J. P., CAVENDER A. C., BRUDVIK J. T., WOBROCK J. O. & LANDER R. E. (2007). Webinsitu : A comparative analysis of blind and sighted browsing behavior. In *9th International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS)*, p. 51–58.
- BORODIN Y., BIGHAM J. P., DAUSCH G. & RAMAKRISHNAN I. (2010). More than meets the eye : A survey of screen-reader browsing strategies. In *International Cross Disciplinary Conference on Web Accessibility (W4A)*, p. 1–10.
- BREGMAN A. S. (1994). Auditory scene analysis : The perceptual organization of sound.
- CAI D., HE X., MA W.-Y., WEN J.-R. & ZHANG H. (2004). Organizing WWW Images Based on the Analysis Of Page Layout And Web Link Structure. *Microsoft Research Asia, Beijing, China*.
- CAI D., YU S., WEN J.-R. & MA W.-Y. (2003). *VIPS : A vision-based page segmentation algorithm*. Rapport interne, Microsoft technical report, MSR-TR-2003-79.
- CAO J., MAO B. & LUO J. (2010). A segmentation method for web page analysis using shrinking and dividing. *International Journal of Parallel, Emergent and Distributed Systems*, **25**(2), 93–104.
- CHERRY E. C. (1953). Some experiments on the recognition of speech, with one and with two ears. *Journal of the acoustical society of America*, **25**(5), p. 975–979.

- DARWIN C. J., BRUNGART D. S. & SIMPSON B. D. (2003). Effects of fundamental frequency and vocal-tract length changes on attention to one of two simultaneous talkers. *The Journal of the Acoustical Society of America*, 114, p. 2913.
- DIAS G. & CONDE B. (2007). Accessing the web on handheld devices for visually impaired people. In K. WEGRZYN-WOLSKA & P. SZCZEPANIAK, Eds., *Advances in Intelligent Web Mastering*, volume 43 of *Advances in Soft Computing*, p. 80–86.
- GIUDICE N. A., PALANI H. P., BRENNER E. & KRAMER K. M. (2012). Learning non-visual graphical information using a touch-based vibro-audio interface. In *Proceedings of the 14th international ACM SIGACCESS conference on Computers and accessibility*, p. 103–110 : ACM.
- GUERREIRO J. (2013). Using simultaneous audio sources to speed-up blind people’s web scanning. In *10th International Cross-Disciplinary Conference on Web Accessibility (W4A)*, p. 1–2.
- ISSA Y. B., MOJAHID M., ORIOLA B. & VIGOUROUX N. (2010). Analysis and evaluation of the accessibility to visual information in web pages. In *Computers Helping People with Special Needs*, p. 437–443. Springer.
- JIANG X., HU Y. & HANG L. (2009). A ranking approach to keyphrase extraction. *SIGIR’09*.
- LIU X., LIN H. & TIAN Y. (2011). Segmenting webpage with gomory-hu tree based clustering. *Journal of Software*, 6(12), 2421–2425.
- MAUREL F., MOJAHID M., VIGOUROUX N. & VIRBEL J. (2006). Documents numériques et transmodalité. *Document numérique*, 9(1), 25–42.
- MAUREL F., VIGOUROUX N., RAYNAL M. & ORIOLA B. (2003). Contribution of the transmodality concept to improve web accessibility. *Assistive Technology Research Series*, 12, 186–193.
- MICHAIL S. & CHRISTOS K. (2007). Adaptive browsing shortcuts : Personalising the user interface of a specialised voice web browser for blind people. In *Data Engineering Workshop, 2007 IEEE 23rd International Conference on*, p. 818–825 : IEEE.
- MOREL M. & LACHERET-DUJOUR A. (2001). Kali, synthèse vocale à partir du texte : de la conception à la mise en oeuvre. *Traitement Automatique des Langues* 42, p. 193–221.
- PARMANTO B., FERRYDIANSYAH R., SAPTONO A., SONG L., SUGIANTARA I. W. & HACKETT S. (2005). Access : accessibility through simplification & summarization. In *Proceedings of the 2005 international cross-disciplinary workshop on web accessibility (W4A)*, p. 18–25 : ACM.
- SAFI W., MAUREL F., ROUTOURE J.-M., BEUST P. & DIAS G. (2014a). Blind browsing on hand-held devices : Touching the web... to understand it better. In *Data Visualization Workshop (DataWiz 2014) associated to 25th ACM Conference on Hypertext and Social Media (HYPERTEXT 2014)*.
- SAFI W., MAUREL F., ROUTOURE J.-M., BEUST P. & DIAS G. (2014b). A hybrid segmentation of web pages for vibro-tactile access on touch-screen devices. In *3rd Workshop on Vision and Language (VL 2014) associated to 25th International Conference on Computational Linguistics (COLING 2014)*, p. 95–102.
- SANOJA A. & GANÇARSKI S. (2014). Block-o-matic : A web page segmentation framework. In *Multimedia Computing and Systems (ICMCS), 2014 International Conference on*, p. 595–600 : IEEE.
- TURGEON M., BREGMAN A. S. & ROBERTS B. (2005). Rhythmic masking release : effects of asynchrony, temporal overlap, harmonic relations, and source separation on cross-spectral grouping. *Journal of Experimental Psychology : Human Perception and Performance*, 31(5), p. 939.
- VIRBEL J. (1989). The contribution of linguistic knowledge to the interpretation of text structures. In *Structured documents*, p. 161–180 : Cambridge University Press.
- WITTEN I. H., PAYNTER G. W., FRANK E., GUTWIN C. & NEVILL-MANNING C. G. (1999). KEA : Practical Automatic Keyphrase Extraction. *NRC/ERB-1057*.
- ZIAT M., GAPENNE O., STEWART J. & LENAY C. (2007). Haptic recognition of shapes at different scales : A comparison of two methods of interaction. *Interacting with Computers*, 19(1), 121–132.

Interfaces textuelles, une difficulté pour l'utilisateur sourd ? Réflexion et éléments de réponse autour d'une étude de cas

Juline le Grand
eROCCA, 196A route du rocher de Lorzier, 38430 MOIRANS
juline.le-grand@erocca.com

Résumé. Avec l'essor des tablettes tactiles et smartphones, les usages des terminaux mobiles ont profondément évolués, révolutionnant de fait la conception des interfaces homme-machine. A travers cet article nous allons détailler les problématiques soulevées en conception d'interfaces majoritairement textuelle lorsque l'utilisateur est sourd. La personne sourde peut présenter des spécificités linguistiques, socio-culturelles ou encore psychologiques et sensorielles qui requièrent une attention particulière quant à la façon de penser l'interface. Pour le projet Rapsodie, dont l'objectif est de fournir un outil de communication améliorée et alternative dédié aux sourds, nous avons été amenés à réfléchir à la manière de concevoir notre interface centrée sur cet utilisateur atypique. Par notre approche, nous avons cherché à déterminer lesquelles de ses « spécificités » requièrent une prise en compte en conception, en particulier la notion de l'écrit qui, bien que prépondérante de manière générale dans les interfaces, pose un problème important pour certains sourds. Nous proposerons des solutions possibles et exposerons leurs limites, souvent liées à des verrous technologiques.

Abstract.

Textual interface, a difficulty for deaf user? Reflection and elements of response around a case study

With the rise of tablets and smartphones, usage of mobile devices has profoundly evolved, revolutionizing at once the design of human-machine interfaces. Through this article we will detail the issues raised in the design of mainly textual interfaces when the user is deaf. Indeed, the deaf person may have linguistic, socio-cultural, or psychological and sensorial specificities requiring special attention as to how to think the man-machine interface. As part of the Rapsodie project, which aims to provide an improved and alternative communication tool dedicated to the deaf, we have been led to think about how to design our interface centered on an atypical user. With our approach, we sought to determine which of its "specificities" require consideration, in particular the notion of writing which, although generally predominant in interfaces, poses a problem for some deaf. We will propose possible solutions and describe their limitations, mostly related to technological barriers.

Mots-clés : Surdit , reconnaissance automatique de la parole, synth se vocale, Communication Am lior e et Alternative, Interface Homme Machine.

Keywords: Deafness, Automatic Speech Recognition, Text To Speech, Augmentative and Alternative Communication, User Interface.

1 Introduction

En quelques d cennies, suite   l'automatisation de masse d'abord dans l'industrie puis dans de nombreux domaines, les pratiques dans l'interaction de l'homme avec la machine ont totalement chang es. Jusque dans les ann es 30-40, c' tait   l'homme de s'adapter   la machine. Machines qu'il rencontrait en tr s grande majorit  sur son lieu de travail. Depuis les ann es 80, les terminaux ont envahi les foyers, on compte par exemple 6.5  crans en moyenne par foyer¹. Cet essor a  t  favoris  par les avanc es technologiques ainsi que la croissance des capacit s de traitement et de stockage ont repouss  les limites de l'informatique. En parall le, cette d mocratisation a  t  favoris e par le fait que les ergonomes ont travaill    des interfaces adapt es au plus grand nombre, visant la simplicit  d'utilisation et l'intuitivit . Par leur caract re personnalisable (choix de langue, contrastes de couleur, choix d'applications/logiciels, inter-connectivit , etc.). Les terminaux sont d sormais abordable par tous, que l'on soit novice en informatique, peu importe le niveau scolaire, ou que l'on ait des d ficiences quelconques, il nous est possible d'utiliser un tr s grand nombre de machines.

La notion d'accessibilit  est devenue un enjeu primordial et on a de plus en plus cherch    adapter les technologies   diff rents handicaps voire   en cr er certaines pour pallier au handicap. Nous pouvons citer par exemple les logiciels de lecture d' cran pour les malvoyants, la commande vocale pour le contr le de l'environnement par les personnes   mobilit  r duite ou encore le sous-titrage pour les malentendants. Malheureusement, m me   l'heure actuelle, tout le

¹ <http://www.lamonde.fr/technologies/article/2014/02/20/6-5-ecrans-en-moyenne-par-foyer-en-2013-4370567-651865.html>

monde n'est pas sur le même pied d'égalité face à l'accès aux technologies (mal voyant pour surfer sur internet, sourds pour l'accès à l'information indisponible en Langue des Signes Française (LSF), etc.). Bien que les Nouvelles Technologies de l'Information et de la Communication (NTIC) contribuent à réduire les discriminations, par un effet pervers, elles renforcent le handicap du fait que certains services ne sont pas accessibles à tous. (Mottez, 1977 ; Kandsi et Belaidi, 2004)

Pour pouvoir adapter les machines/technologies aux utilisateurs et répondre à leurs besoins, il est primordial d'identifier leurs besoins et définir précisément ces utilisateurs sous toutes leurs dimensions (sociales, cognitives, linguistiques, physiologiques, etc.). Seulement à partir de ces données il sera possible de proposer un service, un outil ou une interface plus adaptée à leurs besoins. Dans cet article nous aborderons la question de l'utilisation des interfaces homme-machine par des personnes sourdes. Il sera question également des dimensions selon lesquelles les interfaces textuelles devraient ou non être adaptées à ce type d'utilisateur.

2 L'utilisateur sourd

2.1 La surdité, plusieurs profils de sourds

En 2014 en France selon l'étude DRESS 2014 (Haeusler et al., 2014), 11.2% de la population sont touchés par une surdité. Cela représente plus de 6 millions de personnes. Cette proportion est similaire dans les autres pays du monde et s'accroît plus les revenus du pays sont faibles (OMS, 2014). Toujours selon cette référence, l'Organisation Mondiale de la Santé définit la surdité comme étant une perte d'audition de plus de 25dB par rapport au seuil d'audition normale qui est de 20dB. La surdité est dans 98% des cas acquise au cours de la vie (maladies infectieuses, médicaments ototoxiques, infections chroniques de l'oreille, presbycusie, choc acoustique) et 2% sont due à des causes congénitales (maladie infectieuse de la mère pendant la grossesse, asphyxie à la naissance, médicaments ototoxiques). Le taux de surdité à la naissance est de 1 pour mille.

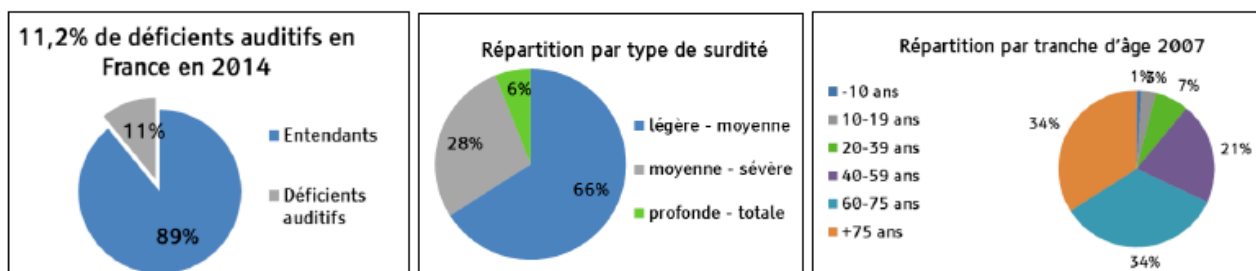


FIGURE 1 : Statistiques sur la surdité, Dress 2014

Etre sourd ne se limite pas à la « déficience ». Face à la perte auditive, on constate en réalité non pas une surdité mais des surdités en fonction de plusieurs dimensions physiologiques, éducatives ou socio-culturelles. Nous pouvons citer principalement :

- le degré de la perte auditive : elle peut être légère avec une perte de 21 à 40 dB H.L.², moyenne de 41 à 70 dB H.L., sévère de 71 à 90 dB H.L., profonde de 91 à 119 dB H.L. et enfin totale pour une perte de plus de 120 dB H.L.. Pour donner quelques exemples, avec une surdité légère on ne perçoit plus des chuchotements ou le bruissement des feuilles avec le vent ; avec une surdité de plus de 55dB on ne peut plus entendre la conversation normale ou le froissement d'une feuille de papier ; un sourd sévère n'entend plus le bruit d'un sèche-cheveux ou d'une voiture ; un sourd profond n'entendra pas une alarme voire une tronçonneuse ; en cas de surdité totale, on n'entend pas un coup de feu à quelques mètres ;
- la période d'apparition de la déficience : l'âge de survenue de la surdité influe fortement sur la 'gestion' de la surdité par la personne. Les sourds pré/péri-linguaux, c'est-à-dire ceux dont la surdité est arrivée in utero ou avant les 3 ans de cause souvent génétique ou infectieuse, auront souvent plus de difficultés pour entrer dans le langage et se tourneront souvent vers la LSF et la culture sourde. Par opposition, les devenus sourds ont une audition qui s'est dégradée au cours de la vie, souvent vers 50 ans, à cause d'une exposition à des bruit trop intense ou de la presbycusie. Cette surdité à l'âge adulte sera plus souvent vécue par comme un handicap et malheureusement facteur d'isolement pour les personnes âgées (Haeusler et al., 2014) ;

² dB H.L. : perte en décibels par rapport à l'oreille normale (20 dB H.L.) en référence aux normes ISO (www.biap.org)

Interfaces textuelles, une difficulté pour l'utilisateur sourd ? Réflexion et éléments de réponse autour d'une étude de cas

- l'appartenance à la culture sourde ou entendante : selon la période d'apparition et la présence ou non de personnes sourdes dans l'entourage et/ou les recommandations médicales, l'enfant sourd sera élevé dans la culture sourde ou entendante tel que le montre Delaporte (Delaporte, 2000). Le sourd adulte pourra choisir lui-même sa culture selon ses affinités et son vécu vis-à-vis de sa surdité. Delaporte définit la culture sourde de la manière suivante « sur ce sentiment partagé par tous [qu'il existe une identité sourde] se construit une culture, c'est-à-dire un ensemble collectif de représentations, de savoirs, de pratiques, de règles sociales, de comportements, de rituels de politesse, de valeurs, de manières de se catégoriser et de se nommer qui, avec les signes, se transmettent de génération en génération. ». Cependant la place du sourd dans la communauté entendante est vue comme une place d'handicapé et non de minorité culturelle tel que se revendiquent les Sourds (sourds se revendiquant de la culture sourde). Selon Goasmat (Goasmat, 2008) être Sourd est une « autre norme de vie possible », les sourds signants (sourds locuteurs de la LSF) rejettent la notion de handicap car pour eux il n'y a pas de handicap auditif mais une norme entendant et une norme sourd où le français est le lien entre les deux cultures/normes ;
- la présence ou non d'un appareillage : le sourd (ou ses parents pour les surdités précoces) peut également choisir d'être appareillé ou non. L'offre et la qualité des appareils auditifs s'est considérablement développée ces dernières années, le sourd ou malentendant peut ainsi espérer un gain jusqu'à 60dB. Concrètement, une personne ayant une perte d'environ 75 dB HL ne pouvant plus percevoir une sonnerie de téléphone voire un aspirateur, pourra, grâce à des appareils, réentendre la parole et même des chuchotements. Les implants cochléaires, adressés aux surdités importantes, sont directement liés à la vision médicale de l'« oreille cassée » qu'il faut réparer. Il s'agit d'un dispositif composé d'un récepteur installé dans la boîte crânienne derrière le pavillon de l'oreille relié à un faisceau d'électrodes dans l'oreille interne vers les terminaisons du nerf auditif situées au niveau de la cochlée et d'un processeur qui vient transmettre les sons captés en signaux électrique vers le récepteur. L'implant est une autre alternative pour laquelle, en 2010 d'après le Centre d'Information sur la Surdité et l'Implant Cochléaire (CISIC), plus de 10 000 personnes avaient déjà eu recours ;
- les pratiques éducatives et langagières choisies par la personne sourde ou son entourage : on relève deux courants éducatifs que sont, d'une part le bilinguisme avec la pratique de la LSF (entre 50 et 80 000 locuteurs sourds) et l'apprentissage du français écrit, et d'autre part l'oralisme qui prône l'apprentissage du français (écrit et parlé) grâce à la lecture labiale et la Langue Française Parlée Complétée (LFPC). 95% des enfants sourds naissent dans une famille entendante (Delaporte, 2000), ce qui va beaucoup influencer sur les choix éducationnels des enfants sourds. Si la famille n'est pas sensibilisée à la surdité, elle choisira souvent le courant oraliste et l'appareillage pour les enfants. De la même manière que l'éducation détermine le comportement sociétal de l'individu dans n'importe quelle culture, Delaporte explique que la relation qu'aura le sourd avec la société entendante et le monde qui l'entoure est déterminé par l'éducation fournie par ses parents et les rencontres qu'il fera (scolarité, milieu médical, adulte sourd).
- l'illettrisme : le diagnostic de la surdité chez l'enfant se fait souvent tardivement et il en découle une entrée dans le langage souvent ralentie, un taux important d'illettrisme et un rapport à l'écrit particulier chez certains sourds mentionné par (Bertin, 2010 ; Renard, 2001). Bien que discutable, on avance ainsi le chiffre de 80% d'illettrisme chez les sourds sévères et profonds, menant à une scolarisation et à une évolution professionnelle moins évidente que pour les personnes entendantes.

	Surdité précoce (2% des sourds)	Devenu sourds (98% des sourds)
Surdité	Moyenne à totale (sourds)	Souvent légère à moyenne (malentendants)
Appareillage/implant	Non	Oui
Langue	Bilinguisme LSF + lecture labiale	Oralisme + lecture labiale
Culture	Sourde	Entendante
Écrit	Français plus ou moins maîtrisé	Identique que pour les entendants
Vision	Anthropologique	Médicale
Intégration sociale (illustration avec le taux de chômage (src. Haeusler et al., 2014))	Minorité peu intégrée (de 30 à plus de 50% de chômage)	Intégré, handicap plus ou moins problématique (chômage environ 20%)

TABLE 1 : récapitulatif des 2 profils de sourd types

Ces différents aspects vont déterminer le profil de l'individu sourd et sa relation au monde et à la société qui l'entoure (Dalle, 2003). Bien souvent, le profil du sourd se construit, fruit du hasard et des rencontres (Delaporte 2000) qui vont

le guider plus ou moins malgré lui vers telle ou telle pratique. Ces caractéristiques des profils de sourd que nous venons de voir peuvent mener à ce que, face aux NTIC et en particulier les interfaces, l'utilisateur sourd ait des besoins particuliers qu'il faudra prendre en compte dans la conception d'interfaces homme-machine.

2.2 Les besoins de l'utilisateur sourd

En tant qu'utilisateur de nouvelles technologies, les spécificités du sourd que nous avons décrites dans la partie précédente le différencie des autres utilisateurs.

Accéder aux informations sonores. L'altération de l'audition empêche le sourd d'accéder, selon le degré de la déficience, à un très grand nombre d'informations naturellement transmises par le son dans la culture entendante (parole, alerte sonore, bruits de la nature, etc.). Pour accéder à certaines informations auditives pour l'entendant, le sourd va se baser sur des repères visuels ou tactiles (mouvements de la nature, variations de couleurs, vibrations ressenties) comme le décrit (Delaporte, 1998). Un des besoins de la personne sourde est de pouvoir accéder à ces informations. Les repères sonores, tels que le bruit indiquant la présence d'une personne ou la parole, n'existent pas pour le sourd : si un élément sort du champ visuel, il disparaît (Goasmat, 2008). Bien qu'ils aient pu mettre en place certaines stratégies dans la vie quotidienne, il subsiste une perte d'information auxquelles les sourds souhaitent pouvoir accéder. Impossible par exemple de passer un appel téléphonique pour un appel d'urgence, entendre une alarme, difficile de suivre lors d'une réunion professionnelle ou de passer un entretien, etc. Un certain nombre de situations restent très problématiques au quotidien malgré de plus en plus d'aménagements et de services proposés. Bien qu'ils trouvent des stratégies pour de nombreux aspects de la vie quotidienne, l'autonomie des sourds est mise à mal par la transmission sonore des informations dans la société entendante dans laquelle ils vivent.

D'avantage de LSF. On constate souvent chez les personnes présentant une déficience auditive, un bilinguisme et une multi-modalité du langage entre français oral, écrit, la lecture labiale et la LSF (Millet et Estève, 2008). La plupart des personnes sourdes combinent plusieurs modes de communication et souvent le français (écrit ou oral), la langue des signes et/ou le LFCP et la lecture labiale. Toutefois, cela reste souvent compliqué pour un sourd de communiquer avec un entendant. Le second besoin des sourds serait donc de pouvoir atténuer cette « barrière » linguistique entre les sourds et les entendants pour favoriser l'intégration sociale des sourds dans la communauté entendante et l'accès aux différents services auxquels ils ont droit.

Prendre en compte la culture Sourde. La dimension culturelle est également importante chez les personnes sourdes et, Goasmat l'explique, il existe une véritable problématique d'intégration sociale des sourds. Le sentiment d'appartenance à la culture sourde s'est vu renforcé par le rejet exprimé depuis des siècles par les entendants (Bertin, 2010 ; Kerbrouc'h, 2009 ; Goasmat, 2008). Perçue par le milieu médical comme un handicap, les Sourds vivent la surdité comme une culture, la culture sourde en opposition à la culture entendante. C'est une culture visuelle dans laquelle « il y a une manière spécifiquement sourde de s'imprégner en permanence de tout ce que le monde peut apporter d'informations visuelles. Or, comme le décrit Goasmat, l'intégration se traduit par l'abandon de toute spécificité culturelle (langue, coutumes, etc.). On voit bien ici la problématique que pose l'identité sourde. Malgré ce sentiment, on observe un réel désir de la part des sourds d'être intégrés dans la société entendante. La reconnaissance des sourds signants comme minorité culturelle et non comme handicapés constitue un autre besoin.

Favoriser le regard. Le regard joue un rôle important dans la culture sourde car c'est par lui que la communication est établie. Delaporte (Delaporte, 1998) montre que « Le regard n'est jamais passif ni au repos, il est sans cesse sollicité par tout ce qui est en mouvement. [...] Les sourds organisent donc en permanence leur rapport à l'espace sur deux critères : pouvoir utiliser la totalité de leur champ visuel, et ne pas risquer que des choses importantes se passent derrière leur dos. [...] Dans toute conversation, le regard tient une place centrale. Celui qui est en situation d'écoute ne quitte pas le visage de son vis-à-vis ». Ainsi, quitter des yeux le sourd qui s'exprime est impoli. Dès qu'une personne ou un objet sort du cadre visuel du sourd, le contact est totalement rompu. Il existe donc un besoin de privilégier le contact visuel, dans le cas de la communication en particulier et ne pas intégrer d'éléments animés autre que nécessaire car cela perturbe l'attention du sourd.

Des difficultés relationnelles. Comme nous l'avons vu dans le paragraphe sur la culture sourde, il existe une réelle problématique de l'intégration du sourd. Sur le plan professionnel, (Metzger et al, 2004) démontrent la difficulté de l'intégration du sourd dans un environnement professionnel entendant. Ainsi, en plus d'assurer leur travail ils devront trouver des stratégies pour pallier à leur déficience. Naturellement, ces stratégies les mènent vers l'usage d'outil et technologies (e-mail, chat, ...) ou des collègues qui serviront d'interprètes. Bien souvent, faute d'efficacité de communication, l'employé sourd se replie sur lui-même. Ces situations mènent parfois à des situations relationnelles complexes comme le décrit Dalle-Nazébi dans (Dalle-Nazébi, 2009). Des divers événements de l'Histoire de la surdité, en découle une faible confiance en les entendants. Il existe un réel besoin d'amélioration du « vivre-ensemble ».

Interfaces textuelles, une difficulté pour l'utilisateur sourd ? Réflexion et éléments de réponse autour d'une étude de cas

Pour conclure, nous pouvons dire que ce que recherchent les personnes sourdes c'est être égaux aux entendants face à l'éducation, au travail, aux loisirs, et ne plus être vu comme handicapés. En ce qui concerne les interfaces homme-machine, ces besoins se traduisent par des interfaces composées essentiellement d'éléments visuels et qui intègrent des fonctionnalités pour représenter graphiquement les informations sonores auxquelles ils n'ont pas accès naturellement. Par ailleurs, l'écrit devra être réduit au minimum.

3 Le sourd face aux technologies

3.1 Les apports de la technologie pour les sourds

L'avènement de la technologie apporte un bouleversement dans la vie des sourds. Le minitel apparu dans les années 80 a été le début de la métamorphose de la communauté sourde. Par la suite, le téléphone mobile a permis les SMS, puis le sous-titrage pour la télévision numérique ou au cinéma ainsi que l'informatique et toutes les portes que cela ouvre. Les technologies permettent en particulier la communication privée (par écrit) entre sourds et entendants, sans avoir à passer obligatoirement par une tierce personne, ce qui est un véritable gain d'autonomie. Les sourds peuvent également accéder plus facilement aux connaissances, vivre leur vie privée, professionnelle ou communautaire. Face à ces technologies, Goasmat dans (Goasmat 2008) cite Benoit Virole en parlant de « confort-interactionnel ».

L'utilisation des technologies par les sourds se fait en parallèle des services d'interprétation et centres relais (ex. Tadéo, Interprétis, Asourd). Indicateur de la barrière linguistique et de l'impact sur le plan de l'autonomie, les sourds signants sont obligés de passer par des services d'interprétariat pour un certain nombre de tâches. Dans le monde professionnel, les entreprises font appel à ces services pour la traduction français/LSF lors de réunions ou d'entretiens ; dans le milieu scolaire on fait appel à des médiateurs spécialisés qui vont assister le professeur vis-à-vis d'élèves sourds signants ; pour des rendez-vous médicaux ou administratifs, les sourds signants font également souvent appel à ce type de service.

L'usage de la téléphonie est possible dans le cas de certaines surdités avec des amplificateurs, des boucles magnétiques ou le Bluetooth qui émettent directement dans les appareils auditifs. L'utilisation des SMS est très largement utilisée par les Sourds comme l'indique M. Power dans (Power et Power, 2004). Grâce à la visiophonie disponible sur certains téléphones, les locuteurs de la LSF peuvent converser dans leur langue au même titre que les entendants ont des conversations téléphoniques. Preuve des évolutions, en 2011, un numéro d'urgence dédié aux sourds a également été lancé, le 114. Les SMS ne nécessitent pas une grande maîtrise de l'écrit. Or l'écrit est un véritable plaisir pour les sourds d'après (Hamm, 2012). On retrouvera cet engouement pour la communication écrite des sourds sur les sites/logiciels de chat ou les nombreux forums que l'on peut trouver sur la toile (La main des sourds, Le village des Sourds).

Il y a par ailleurs de nombreux sites dédiés à la transmission de l'information pour les sourds signants (websourd-media ou Pissourd) qui permettent aux internautes de consulter des contenus concernant l'actualité, des démarches administratives, la vie quotidienne et bien d'autres thèmes en langue des signes. Dans leurs pratiques, en dehors de la barrière langagière et des problèmes de déchiffrement et d'écriture du français, les sourds sont donc des internautes et des utilisateurs comme les autres puisqu'ils utilisent les mêmes outils que les entendants. Pour la saisie dans les moteurs de recherche, les systèmes d'autosuggestion facilitent les recherches. Parmi les entendants, même en cas de difficultés avec le français (orthographe problématique voire illettrisme), il est possible d'utiliser les outils informatiques (Lenoir, 2003). Il en est de même pour les sourds. Des logiciels tels que skype ou ooVoo sont aussi très utilisés pour la communication des sourds puisqu'il permet d'établir le contact visuel. Cependant, il existe des contraintes techniques comme celle du débit requis pour une fluidité de la transmission qui peut rendre la communication impossible. Comme cela est expliqué dans l'étude (John et al., 2009), « l'usage d'internet est non seulement une grande richesse mais également un outil 'thérapeutique' ». Internet est très utilisé par les sourds comme moyen de communication encore plus que les entendants utilisent le téléphone. En effet, la présence de webcams permet la communication gestuelle propre aux sourds. Selon l'étude, les expressions telles que les émotions ou les mimiques sont perçues au même titre que les intonations pour la parole. Selon John et al., internet aurait favorisé le lien social entre sourds, la communauté sourde étant très soudée, de nombreux échanges se sont vus rendus possibles grâce à internet, en particulier à l'adolescence.

La technologie contribue donc à rétablir le lien-social entre les sourds et entendants. Bertin explique que grâce à ces outils, les sourds peuvent être citoyens comme les entendants et accéder aux services auxquels ils ont droit et pour lesquels il serait difficile d'accéder sans la technologie. Néanmoins, une petite partie des personnes sourdes ne peuvent pas accéder pleinement à ces services car ils ne maîtrisent pas le français écrit et l'écrit est prédominant dans les IHM des outils et services. Pour eux, l'accessibilité des interfaces passe obligatoirement par la langue des signes.

3.2 TALN et surdité

Face au manque d'autonomie des personnes sourdes et aux difficultés d'intégration sociale et professionnelle, le traitement automatique des langues naturelles (TALN) est une alternative qui se développe. Actuellement, les performances des systèmes de synthèse et de reconnaissance vocale sont relativement satisfaisants et permettent d'envisager leur utilisation dans le cadre de la communication des personnes sourdes. Des systèmes existent, on peut citer par exemple les solutions commercialisées par Nuance en reconnaissance vocale ou Acapella et Protéor pour la synthèse vocale. Le problème des dispositifs existants est qu'ils sont basés exclusivement sur des données écrites, or ce mode de représentation graphique, nous l'avons déjà abordé dans la section 2.1, n'est pas adapté à tous les profils de sourds. Ces dix dernières années, plusieurs équipes de recherche ont travaillé sur le TAL et les langues gestuelles.

La LSF étant une langue à part entière, avec son vocabulaire, sa syntaxe, ses indicateurs prosodiques comme le montrent (Balvet et Sallandre, 2005 ; Boutet et Garcia, 2010 ; Cuxac, 2010), elle peut donc être soumise aux techniques de traitement automatique. (Rezzoug et al., 2006 ; Brafford et Dalle, 2005) ont travaillé sur des systèmes de synthèse de la LSF. Un signeur virtuel est par exemple utilisé avec succès dans plusieurs gares SNCF³. En revanche, la reconnaissance automatique de la LSF se heurte au manque de descriptions linguistiques, à la méconnaissance de ses mécanismes et au manque de recul sur ses mécanismes d'articulation (Amblard et Voisin, 2007 ; Bossard, 2002 ; Garcia et al., 2011).

En ce qui concerne le Langage Français Parlé Complété, plusieurs projets ont également été menés. Ce code étant basé sur la prononciation des mots du français (Cornet, 1967), il est plus aisé de développer des outils car les combinaisons sont limitées et connues. Piloté par le RIAM, le projet LABIAO (Le Pouliquen et al, 2007) avait pour but de fournir une synthèse 3D des gestes du LFPC. Pour aider à la lecture labiale, le LFPC a été développé afin de désambigüiser les prononciations identiques du point de vue de la lecture labiale seule. L'usage premier du LFPC est l'apprentissage de la langue orale pour les enfants sourds et malentendants. Il est assez peu utilisé par les adultes. Ce projet croise l'usage de l'imagerie numérique et l'usage de la synthèse et de la reconnaissance vocale pour une application dans l'enseignement en fournissant un assistant codeur LFPC et dans le cadre de la téléphonie mobile en fournissant un assistant qui code la communication de l'entendant pour le sourd. L'avantage principal du projet LABIAO est son interopérabilité. L'inconvénient est que ce dispositif ne convient qu'aux sourds oralistes utilisant le LFPC ce qui ne couvre qu'une infime partie des personnes sourdes. D'autres travaux (Beautemps et al., 2007 ; Daassi-Gnaba et al., 2010) traitent également du LFPC.

Pour assister la lecture labiale, le projet LIPCOM (Coursant-Moreau et Destombes, 1999) a été initié par IBM et visait à « faciliter la réception de la parole par les personnes sourdes face à un interlocuteur entendant ». LIPCOM fournit un sous-titrage phonétique de la parole du locuteur entendant. LIPCOM a été testé 5 ans dans un contexte scolaire. Cet outil ne s'auto-suffit pas mais vient se placer en complément à la lecture labiale et aux restes auditifs de personnes appareillées. Le locuteur parle dans un micro et grâce à une reconnaissance vocale, on affiche la transcription à destination du sourd en phonétique en temps réel. Lors des premiers essais, il a été démontré que l'outil favorisait la compréhension des élèves. Grâce aux avancées de la reconnaissance vocale, des applications telles que RogerVoice ou Pedius sont apparues, proposant un sous-titrage en temps réel de la parole du locuteur entendant à destination de la personne sourde en situation d'appel téléphonique. Mais dans le cas d'aménagement de poste, en dehors de quelques mesures d'aménagement de l'espace (ex. installation d'avertisseurs visuels, sensibilisation du personnel entendant) aucun logiciel ne permet la communication en face à face ou par téléphone sans l'intervention d'une tierce personne.

Ainsi, le TALN est une source d'accessibilité pour les personnes isolées par la surdité. Les technologies ne sont pas encore matures pour être utilisées à grande échelle mais les applications possibles sont nombreuses (reconnaissance de signes, sous titrage par la reconnaissance vocale, synthèse de gestes (LSF ou LFPC), traduction français/LSF, ...) pour des usages au quotidien dans le privé ou sur le lieu de travail, ainsi que les loisirs ou encore l'administration et le médical. Reste à surmonter les verrous technologiques qui subsistent et trouver une alternative ergonomique pour des interfaces adaptées aux différents profils de sourds.

4 Adapter des interfaces basées sur l'écrit lorsque l'utilisateur signe

4.1 Contexte et objectifs du projet Rapsodie

Pour répondre au coût social de la surdité en Europe et plus particulièrement à la question du chômage observé en France (jusqu'à 50% pour les sourds profonds (Kerbrouc'h, 2009 ; Haeusler, 2014), le projet Rapsodie a été initié par eROCCA. Le but de ce projet est de proposer aux personnes sourdes un support à la lecture labiale pour assister la

³ <http://www.accessibilite.sncf.com/la-demarche-d-accessibilite/24/equipements/article/jade>

Interfaces textuelles, une difficulté pour l'utilisateur sourd ? Réflexion et éléments de réponse autour d'une étude de cas communication entre sourds et entendants en face à face ou par téléphone. Dans le projet, les contextes d'usages ont été limités à des hôtes(ses) de caisse et des agents en logistique dans un magasin de bricolage. Des systèmes et services dédiés existent bien mais les conditions d'utilisation, les prix et la disponibilité ne permettent pas une autonomie satisfaisante pour les personnes concernées.

Un tel outil d'aménagement de poste permettrait de favoriser l'autonomie des personnes sourdes ainsi que leur évolution professionnelle en rétablissant la communication entre les sourds et les entendants. Pour cela, l'outil de CAA intègre un dispositif de synthèse vocale et de reconnaissance vocale comme le décrit la FIGURE 2.

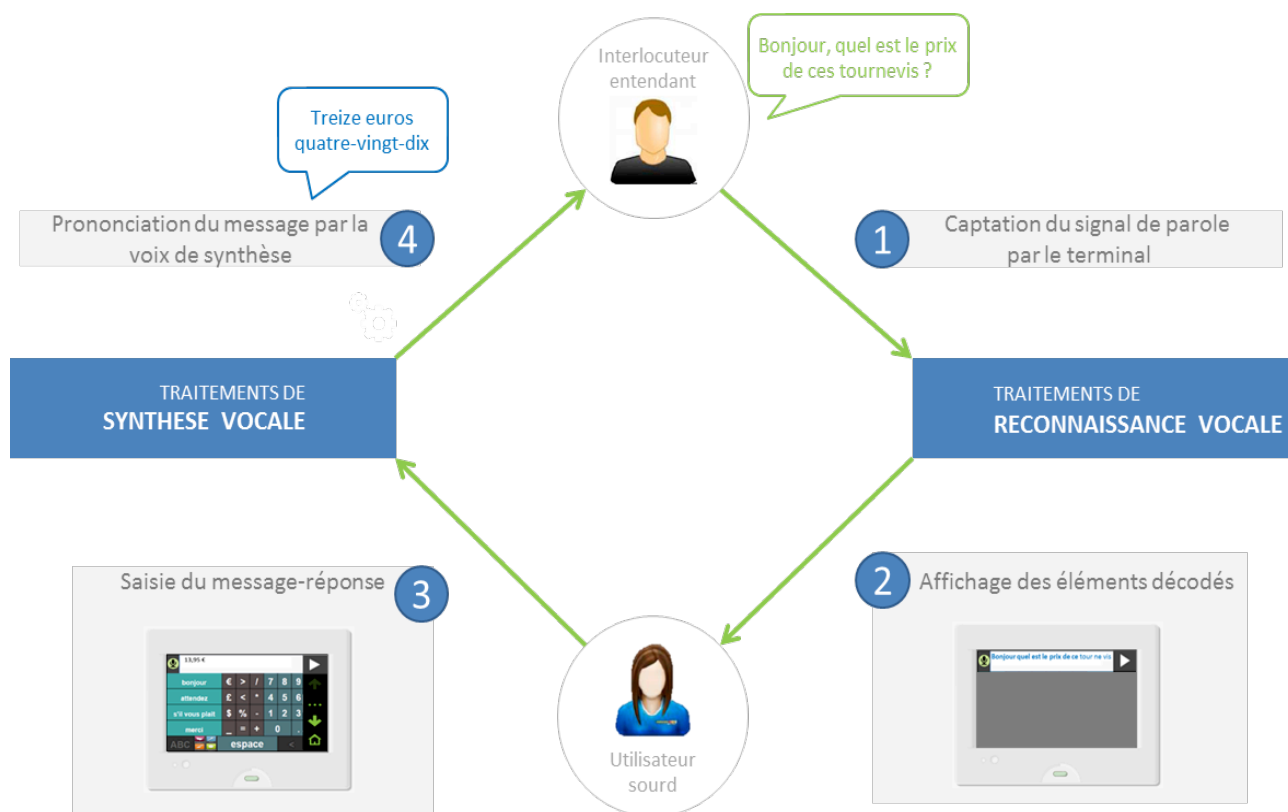


FIGURE 2 : schéma du fonctionnement du terminal de CAA issu du projet Rapsodie

Le problème réside dans le fait que ces systèmes sont basés sur la langue française écrite que ce soit pour la saisie de message par l'utilisateur sourd ou pour l'affichage à destination des éléments décodés par la reconnaissance vocale de son interlocuteur entendant. Dans quelle mesure l'écrit pour les entrées et sorties des fonctionnalités de synthèse et de reconnaissance vocale va-t-il poser problème aux utilisateurs sourds ? Quelle alternative est-il possible de proposer ?

4.2 Méthodologie

4.2.1 Objectifs et protocole

Dans le cadre du projet, nous avons mené une étude auprès de personnes sourdes pour évaluer leur besoins sur plusieurs points vis-à-vis du prototype que nous souhaitons développer :

- le mode d'affichage des résultats de la reconnaissance vocale ;
- le vocabulaire métier (phrases types très couramment utilisées (Falzon, 1991)) ;
- les pratiques linguistiques sur le terrain ;
- les interfaces / outils déjà utilisés (indicateur d'éventuelles habitudes) ;
- comment concevoir l'outil pour une communication la plus naturelle possible (interface, positionnement au poste de travail, fonctionnalités, ...).

Deux études ont eu lieu en parallèle. Pour obtenir des informations sur ces différents points nous avons mené des observations en situation réelle puis des entretiens, d'une part auprès des 6 collaborateurs sourds observés dans les magasins et d'autre part auprès des 10 personnes sourdes qui se sont portées volontaires pour des entretiens en institut

menés par l'Inria dans le cadre du projet. L'étude concernant la méthodologie, les choix de l'affichage ainsi que les entretiens en institut a été réalisée par l'Inria, chargée de cet aspect du projet.

Grâce à cette étude de l'utilisateur, nous espérons pouvoir proposer une solution qui soit le mieux adaptée possible aux besoins de l'utilisateur sourd.

Entretiens en instituts. L'objectif est ici de valider le mode d'affichage des résultats de la reconnaissance vocale qui soit convenable pour les sourds. Trois types d'unités lexicales de décodage automatique ont été étudiés dans le cadre du projet à savoir mots, syllabes et mélange de mots et syllabes en fonction des mesures de confiance (Orosanu et Jouvett, 2013). Une fois reconnus, les mots et syllabes peuvent être affichés sur le terminal de l'utilisateur. Les différentes modalités sélectionnées étaient le mode orthographique, les signes de l'alphabet phonétique international ou la phonétique simplifiée avec les signes de l'alphabet (cf. TABLE 2).

Ces différentes modalités ont donc été évaluées lors d'entretiens individuels pendant lesquels chacune des phrases reconnues par le système a été présentée au sujet sourd ou malentendant dans un contexte situationnel proche de celui

		Unités lexicales utilisées lors de la reconnaissance		
		Mots	Syllabes	Mots + Syllabes
Mode d'affichage	Mode orthographique	je voudrais être livré combien ça coûte		je voudrais être li vré combien ça kou te
	API	ʒə vu dʁɛ ɛtʁ livʁɛ kɔ̃bjɛ sa kʊtə	ʒə vu dʁɛ ɛtʁ li vʁɛ kɔ̃ bjɛ sa ku tə	ʒə vu dʁɛ ɛtʁ li vʁɛ kɔ̃bjɛ sa ku tə
	Phonétique simplifiée	je voudré ètr livré konbyin sa koute	je vou dré ètr li vré kon byin sa kou te	je voudré ètr li vré konbyin sa kou te

d'une communication réelle : il est vendeur dans un magasin de bricolage et un client entendant lui pose une question.

TABLE 2 : tableau issu des travaux menés dans le cadre de Rapsodie par l'équipe Parole du Loria

Ces modalités ont été évaluées selon deux critères que sont l'amélioration de la compréhensibilité de la phrase transcrite et l'appréciation de la modalité d'affichage. Deux phases expérimentales ont été menées d'une ou deux heures séparées d'au moins deux semaines.

Le sujet était assis à une table dans une pièce calme, face à l'examineur qui lui propose environ 20 phrases. Lors de la première session, les phrases ne comportent pas d'erreur de décodage. Il est question de tester les réactions vis-à-vis des différentes unités lexicales, les trois modes d'affichage et l'affichage ou non d'informations non lexicales comme les pauses, hésitations. Lors de la deuxième phase, les phrases comportaient des erreurs, afin de faire l'exercice avec des sorties réelles de la reconnaissance vocale. Des modes d'affichage présentant ces erreurs ont été testés :

1. les unités lexicales reconnues étaient affichées selon deux modalités (orthographique et phonétique simplifiée) et il n'y avait pas d'indication des unités lexicales erronées ou potentiellement erronées ;
2. les unités lexicales réellement erronées étaient indiquées manuellement en utilisant une des trois modalités d'affichage des erreurs (couleur, phonétique simplifiée au lieu de mot incorrect, couleur + phonétique simplifiée) ;
3. les unités lexicales que le système considérait comme erronées étaient indiquées selon les trois modalités d'affichage des erreurs.

Observations et entretiens en magasin. La première étape fut de mener une phase d'observation sociolinguistique des utilisateurs cibles en situation. 6 personnes sourdes de profils variés ont été observées pendant une durée de 2 à 6 heures chacune. Ensuite des entretiens individuels ont été réalisés concernant les pratiques langagières et culturelles ainsi que les besoins et attentes. Puis nous avons soumis au sujet les phrases à déchiffrer pour évaluer les différents modes d'affichage. Durant l'intégralité de ces observations, un micro permettait d'enregistrer les conversations et l'environnement sonore du milieu dans lequel évoluait la personne observée. Pour les personnes en logistique, un micro-cravate était discrètement placé sur la personne observée afin d'enregistrer les situations d'interactions auxquelles il était confronté au quotidien. L'observateur se plaçait à distance afin de ne pas influencer le comportement des sujets tout en pouvant observer et entendre les interactions. Pour les observations en caisse, un microphone directionnel était placé à la caisse de façon à pouvoir enregistrer la parole des clients. L'observateur était placé à côté de

Interfaces textuelles, une difficulté pour l'utilisateur sourd ? Réflexion et éléments de réponse autour d'une étude de cas

l'hôte(esse) de caisse pour faciliter les observations. Cela a pu influencer les clients mais les interactions sont restées naturelles. Après les observations, une séance de discussion d'environ 30 minutes a été réalisée avec 4 sujets. Lors de ces entretiens, il a été demandé de déchiffrer des exemples de phrases issues de la reconnaissance vocale. Durant les entretiens, l'examineur et les sujets étaient installés dans une pièce calme (bureau individuel ou salle de repos) assis en face à face à une table et le micro disposé au milieu. Bien qu'ayant à faire à des personnes sourdes, plusieurs oralisaient. Nous avons donc choisi d'enregistrer même lorsque le sujet ne parlait pas. Selon le protocole élaboré pour les entretiens en institut, des phrases issues de leur protocole, imprimées sur des petites fiches, ont été présentées au sujet dans un ordre aléatoire pour ne pas influencer le déchiffrement, mais avec la forme orthographique présentée à la fin. Ensuite les différents modes d'affichage pour une même phrase ont été présentés et il a été demandé au sujet de les classer par ordre de facilité de déchiffrement.

4.2.2 Public étudié

10 personnes dans des instituts (Lorraine et Ile de France), 8 hommes et 2 femmes âgés de 35 à 65 ans, les sujets avaient tous un certain niveau de maîtrise de la langue française, la plupart étant des personnes devenues sourdes ou malentendantes (surdité légère à sévères (OMS, 2014)), une personne était sourde de naissance. Les pratiques langagières étaient la langue des signes française, le français oral, le français signé, le français oral combiné au LFPC et 6 personnes en magasin (Rhône-Alpes et Ile de France), 4 hommes et 2 femmes, âgés de 26 à 48 ans. Tous les sujets ne maîtrisaient pas le français écrit, 4 avaient une surdité sévère ou profonde et 2 étaient malentendants. Tous pratiquent au moins les bases de la LSF et la lecture labiale, les deux personnes malentendantes étaient appareillées et oralisaient et 2 sourds sévères (dont une appareillée) étaient capables d'oraliser.

4.2.3 Résultats

L'objectif principal des entretiens était d'évaluer les préférences des utilisateurs sourds pour le mode d'affichage par écrit des éléments issus de la reconnaissance vocale. Lors des entretiens tous les sujets ont préféré la sortie en mots en premier choix puis en syllabes avec la phonétique française en second choix. Il a été démontré par la seconde phase des entretiens en institut, que le fait d'afficher en mots combiné avec des syllabes pour les éléments moins bien reconnus permettait de mieux déchiffrer les phrases comportant des erreurs de décodage. En ce qui concerne la présentation des erreurs de décodage, l'affichage (1) en rouge des éléments peu fiables faisait porter l'attention de l'utilisateur sur les mots en question et il ne parvenait pas à déchiffrer l'ensemble de la phrase. En revanche, lors de nouveaux essais, il a été démontré que l'affichage (2) en gras des éléments fiables et en normal pour les éléments moins fiables permettait un déchiffrement plus facile. Exemple ici avec la phrase « combien de litres de peinture doit-on acheter pour peindre une chambre d'enfant » :

(1) k o n b y i n d e u l i t r d e u p a t u r d o i v e n t a c h e t e r p o u r p e i n d r e u n e c h a m b r e d a n f a n

(2) k o n b y i n d e u l i t r d e u p a t u r d o i v e n t a c h e t e r p o u r p e i n d r e u n e c h a m b r e d a n f a n

chambre d'enfant » :

En observant les situations d'interactions auprès de personnes sourdes avec des personnes entendant, nous avons pu extraire un vocabulaire et une phraséologie métier. Cela nous a permis de proposer une série d'icônes correspondantes aux phrases les plus courantes en caisse et faciliter ainsi la saisie de messages pour une interaction fluide et n'obligeant pas l'utilisateur à passer par l'écrit. Certaines de ces phrases ont été associées à des icônes (cf. section 4.3).

En ce qui concerne les pratiques linguistiques, dans 5 cas sur 16 l'outil n'est pas pertinent car les personnes n'ont pas de problèmes pour communiquer avec les entendants que ce soit en direct ou par téléphone. Pour seulement un cas sur 16, l'utilisation du français écrit est très problématique. Dans 6 cas sur 16, la langue première de la personne est la LSF. Pour le cas d'interfaces ou d'outils déjà utilisés par les sourds, les entretiens ont démontré que l'usage de l'informatique (ordinateur des caisses) et la compréhension des interfaces de terminaux mobiles de manière générale se fait sans problèmes. Pour les types d'interactions nous avons constaté que les échanges sont à grande majorité professionnels (avec les clients ou collègues sur des sujets relatifs au travail). Les sujets malentendants ont davantage d'échanges avec les clients et collègues tandis que les sujets de surdité plus importante sont plus renfermés face aux clients et collègues entendants et évitent parfois l'interaction. Lors de réunions, des interprètes sont appelés pour traduire aux personnes signantes. Mais au quotidien, ce sont les collègues malentendants ou entendants qui connaissent la LSF qui traduisent en cas de besoin. Sur le terrain, nous constatons donc que très peu de solutions techniques sont utilisées. Lorsque l'appel téléphonique est inévitable (besoin d'une référence ou d'un prix) les personnes ne pouvant utiliser leur ligne interne se voient obligées de faire venir la personne de la caisse centrale et de faire expliquer le problème par le client. Les personnes interrogées ne semblaient pas avoir de requêtes spécifiques concernant l'interface en dehors du fait qu'elle rende accessible la parole des entendants.

D'après nos observations, la configuration des caisses et les habitudes professionnelles rendent possible l'utilisation d'un tel terminal de communication. Toutefois, pour une bonne captation de la parole du client, il faudra un micro le plus près possible du client.

4.2.4 Prochaines évaluations prévues

Des essais de l'interface seront effectués prochainement auprès des personnes observées afin de valider l'interface et les adaptations réalisées. Par cette étape de validation nous allons évaluer les préférences entre la saisie par icônes et la saisie orthographique par clavier et confirmer l'approche des modes d'affichages des éléments issus de la reconnaissance vocale.

4.3 Entre choix et possibilités

Suite aux observations et aux entretiens, nous avons défini les spécifications fonctionnelles et l'interface utilisateur du prototype. Les conclusions des observations ont mené vers plusieurs réflexions :

- L'usage de l'écrit est problématique pour certains utilisateurs voire impossible pour quelques cas exceptionnels, nous avons dû choisir quels utilisateurs favoriser. De manière générale, l'outil est destiné aux hôtes(ses) de caisse sourds de magasins de bricolage. Après une étude des profils de sourds et les observations sur le terrain nous avons conclu que les personnes souffrant d'une surdité légère à moyenne n'ont pas besoin de dispositif dédié et sont plutôt intégrés à la culture entendante. En revanche, les personnes sourdes sévères et profondes depuis l'enfance pratiquent souvent la LSF et la lecture labiale, ils représentent moins de 1% des sourds mais ce sont eux qui sont visés en tant qu'utilisateurs futurs. Parmi eux, une majorité aura quelques difficultés avec le français écrit, mais, suite à l'étude menée dans le cadre du projet, les personnes sourdes interrogées arrivent à déchiffrer les éléments proposés. Une adaptation de l'interface devrait permettre d'atteindre une partie de ces utilisateurs. En cas d'illettrisme réel, la personne ne pourra pas utiliser le terminal.
- Adaptation de l'interface. Certains employés des magasins avaient déjà utilisé des outils de CAA commercialisés par eROCCA (écho 200). Dans un souci de continuité, nous avons choisi de conserver une partie du design (couleurs, organisation du menu, paramètres). Nous avons cherché au maximum à reprendre des habitudes dans l'interface, des repères visuels, une continuité avec l'existant en conservant des icônes tels que la maison, la flèche de lecture ou le micro qui sont courants dans les interfaces de manière générale.

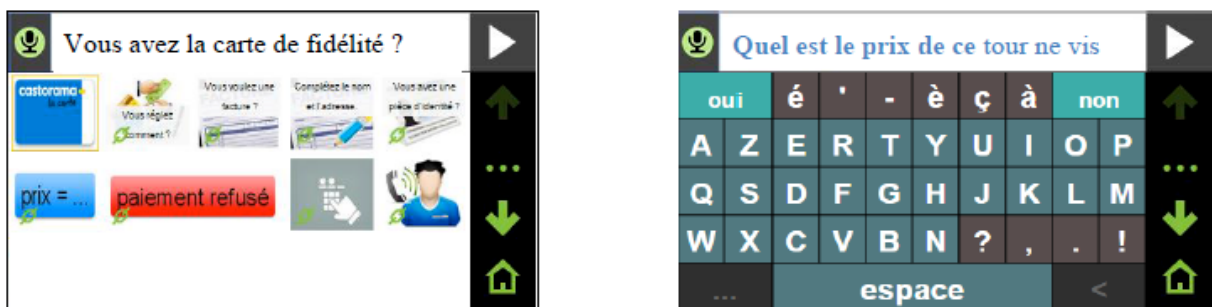


FIGURE 3 : aperçus de l'interface conçue

Nous l'avons vu dans la section 2.2, les sourds sont très attachés au caractère visuel des choses. Dans le cadre de la conception de notre interface, nous avons veillé à apporter certains repères visuels (ex. couleurs différente des textes pour la synthèse vocale et la reconnaissance vocale, disposition fixe des différents boutons, état du micro allumé-éteint-trop de bruit pour le décodage).

Pour favoriser la saisie et aider les utilisateurs plutôt visuo-gestuels qu'oralistes, plusieurs « mesures » sont proposées. Un système de prédiction de mots est activable dans les options pour aider à la saisie. Nous avons étudié la possibilité d'intégrer un dispositif de saisie intégrale en icônes ainsi qu'une traduction des éléments décodés par la reconnaissance vocale en icônes. Mais créer cette iconographie (à partir du vocabulaire et la phraséologie métier) sortait du cadre du projet. Nous avons malgré tout proposé un système de saisie par icônes basés sur les mots/phrases métiers que l'utilisateur côtoie chaque jour, permettant au sourd de s'approprier un langage visuel en lien direct avec les concepts manipulés au quotidien (cf. FIGURE 3). Par ailleurs, nous l'avons décrit dans la section 3.2, l'utilisation de la reconnaissance et de la synthèse de signes n'est pas assez mature pour être proposée ici. Néanmoins, la pertinence de ces technologies pour les interfaces dédiées aux sourds semble ne faire aucun doute. En combinant des éléments visuels et en facilitant l'usage de

Interfaces textuelles, une difficulté pour l'utilisateur sourd ? Réflexion et éléments de réponse autour d'une étude de cas
l'écrit, nous espérons pouvoir permettre à une partie des utilisateurs sourds signant d'accéder plus facilement au terminal.

- Permettre une interaction fluide et la plus naturelle possible. La disposition et les fonctionnalités de l'outil de CAA tiennent compte des règles d'interaction de la culture entendante ainsi que celle de la culture sourde. L'utilisateur sourd doit pouvoir regarder son interlocuteur lorsqu'il parle puis consulter sur le terminal sa réponse si la lecture labiale ne lui a pas permis de comprendre ce qui a été dit. La saisie est optimisée par la présence d'icônes et la prédiction pour que le temps de réponse de l'utilisateur soit le plus rapide possible. Les traitements et capacités de l'outil ont également été étudiés afin de ne pas impacter l'interaction.
- L'aménagement de poste. La disposition des caisses impose un micro « déporté » pour capter au mieux la parole du client. Il faudra également une « station d'accueil » du terminal pour s'adapter à la position de l'utilisateur (assis ou debout) ainsi qu'aux éventuels problèmes de reflets liés à l'éclairage du magasin. Par souci de stigmatisation, le terminal devra être discret et perçu comme un outil de travail.
- Rendre robuste et fiable la communication. Nous l'avons vu dans la section 2.2, les difficultés de communications entre sourds et entendants sont parfois source de mauvaise interprétation voire de tensions. Pour que les informations transmises soient sûres, il était nécessaire de travailler à un système de reconnaissance vocale aux performances optimisées. Pour cela des travaux de débruitage et un affichage modéré selon les mesures de confiances permettant de contrer les problèmes de mots hors-vocabulaire, ainsi qu'une étude de l'adaptation du modèle de langage ont été menés. L'étude de terrain a permis de mettre en évidence que, pour des raisons techniques (bruit et mouvements continus) il serait impossible dans le cadre du projet Rapsodie de proposer un dispositif adapté également pour les métiers de logistique.

5 Conclusion et perspectives

Pour répondre aux problématiques posées par les utilisateurs sourds, nous avons effectué ces choix en fonction des spécificités des utilisateurs et des possibilités qu'offre la technologie. La solution proposée ne permettra pas de répondre à tous les objectifs fixés par le projet. Mais tant que le Traitement Automatique de la Langue des Signes (TALS) ne sera pas mature, il est difficile de proposer un terminal dédié aux sourds tel qu'ils l'attendent.

L'interface textuelle s'adresse-t-elle à tous les utilisateurs ? Les interfaces sont conçues pour des utilisateurs bien identifiés selon leur âge, langue, handicap, domaine d'activité, etc. Face à des personnes sourdes, il semble, d'après les références ainsi que nos observations, qu'il n'y aura pas à prendre de décision propre à la surdité dans la plupart des situations. Grâce à des appareils et/ou à une maîtrise du français, la déficience auditive n'est plus un obstacle pour une majorité des sourds. En revanche, face à des sourds signants, une interface textuelle peut devenir problématique puisque la langue maternelle de l'utilisateur est la Langue des Signes Française et que cette langue n'a pas d'écriture. Qu'advient-il de l'interface textuelle lorsque traduire signifie ne plus passer par l'écrit mais par des signes, des avatars, des vidéos ou encore des icônes ?

Une étude approfondie lors des tests utilisateurs nous permettra de valider notre approche et la pertinence des solutions proposées.

Remerciements

eROCCA remercie tout particulièrement l'INRIA pour les travaux menés par Agnès Piquard-Kipffer, Jérémy Miranda, Odile Mella dans le cadre du projet. Nous tenons à remercier également la BPI France, le Conseil Général de la Haute-Savoie, le Feder Lorraine et le Feder Rhône-Alpes qui ont contribué au financement du projet Rapsodie (www.erocca.com/rapsodie) ainsi que les membres du consortium pour le travail réalisé pour le projet.

Références

AMBLARD M., VOISIN E. (2007). Vers une analyse automatique de la LSF. *Rapport de recherche INRIA*.

BALVET A., SALLANDRE M.-A. (2005). Problèmes et méthodes pour l'analyse d'énoncés en LSF. Actes de la *Journée TALS, conférence TALN 2005*, vol. 2, 361-365.

BEAUTEMPS D., GIRIN L., ABOUTABIT N., BAILLY G., BESACIER L., BRETON G., BURGER T., CAPLIER A., CATHIARD M.-A., CHÊNE D., CLARKE J., ELISEI F., GOVOKHINA O., JUTTEN C., LE V.-B., MARTHOURET M., MANCINI S., MATHIEU Y., PERRET P., RIVET B., SACHER P., SAVARIAUX C., SCHMERBER S., SÉRIGNAT J.-F., TRIBOUT M., VIDAL S. (2007). Telma: Telephony for the hearing-impaired people. From models to user tests. Actes de *ASSISTH'2007*, 204-208.

- BERTIN F. (2010). *Les Sourds : une minorité invisible*. Paris : Autrement.
- BOSSARD B. (2002). Problèmes posés par la reconnaissance de gestes en Langue des Signes. Actes des *4e Rencontres des Etudiants Chercheurs en Informatique pour le Traitement Automatique des Langues*, 445-454.
- BOUTET D., GARCIA B. (2006). Finalités et enjeux linguistiques d'une formalisation graphique de la Langue des Signes Française (LSF). *Les langues des signes (LS) : recherches sociolinguistiques et linguistiques*, R. Sabria éd., Glottopol n°7, 32-52.
- BRAFFORD A., DALLE P. (2005). Traitement informatique de la LS : modèles, représentations, outils d'analyse vidéo, d'animation d'avatar et de communication. Actes du *2e Congrès de l'International Society for Gesture Studies (ISGS) : Interacting Bodies*. École Normale Supérieure Lettres et Sciences Humaines, support électronique.
- BRANGIER, E., GRONIER G. (2000). Conception d'un langage iconique pour grands handicapés moteurs. Actes de *Handicap 2000, Nouvelles technologies : assistance technique aux handicaps moteur et sensoriel*, 93-100.
- CORNETT R.-O. (1967). Cued Speech. *American Annals of the Deaf*, 112, 3-13.
- COURSANT-MOREAU A., DESTOMBES F. (1999). LIPCOM : prototype d'aide automatique à la réception de la parole par les personnes sourdes. *Glossa* 68, 36-40.
- CUXAC C., ANTINORO PIZZUTO E. (2010). Émergence, norme et variation dans les langues des signes : vers une redéfinition notionnelle. *Langue et société*, 2010/1, 131, 37-53.
- DALLE P. (2003). La place de la langue des signes dans le milieu institutionnel de l'éducation : enjeux, blocages et évolution. *Langue française* 137, 32-59.
- DAASSI-GNABA H., ZBAKH M., LOPEZ KRAHE J. (2010). Combinaison de reconnaissance de la parole, reconnaissance des émotions et tête parlante codeuse en LPC pour les personnes sourdes et malentendantes. *Sciences et Technologies pour le Handicap*, 239-253.
- DALLE-NAZEBI S. (2009). Quand les salariés sourds prennent la parole. Présenté lors du *3ème congrès de l'Association Française de Sociologie*.
- DELAPORTE Y. (1998). Le regard sourd. "Comme un fil tendu entre deux visages...". *Terrain, Ministère de la culture et de la communication*, 30, 49-66.
- DELAPORTE Y. (2000). Être sourd dans un monde entendant : destin, rencontre, transmission et rupture. *Ethnologie Française, Société d'ethnologie française (SEF)*, distribution PUF, 389-400.
- FALZON P. (1991). Les activités verbales dans le travail. In *Modèles en analyse du travail*, R. Amalberti, M. De Montmollin, & J. Theureau (Eds.), 229-252.
- GOASMAT, G. (2008). *L'intégration sociale du déficient auditif – Enjeux éducatifs et balises cliniques*. Paris : L'Harmattan.
- GARCIA B., SALLANDRE M-A., SCHODER C., L'HUILLIER M-T. (2011). Typologie des pointages en Langue des Signes Française (LSF) et problématiques de leur annotation. Actes du *premier défi geste langue des signes, DEGELS2011*, 107-119.
- HAEUSLER L., DE LAVAL T., MILLOT C. (2014). Etude quantitative sur le handicap auditif à partir de l'enquête « Handicap-Santé ». *DRESS, Série Etudes et Recherches*, 131.
- HAMM, M. (2012). Écrire sans entendre. Une exploration de la pratique de l'écriture chez quelques sujets sourds, devenus sourds et malentendants. *Education et Formation*, 297, 138-153.
- JOHN C., MAUTRET-LABBE C., PALACIOS P. (2009). Les sourds, internet et le lien social. *Empan* 76, 100-106.
- KANDSI M., BELAIDI A. (2004). Le handicap et les NTIC. Actes du *Colloque TICE 2004*.
- KERBROUC'H S. (2009). L'emploi des sourds et malentendants : quels enjeux ?. *Connaissance de l'emploi* 66, 1-4.

Interfaces textuelles, une difficulté pour l'utilisateur sourd ? Réflexion et éléments de réponse autour d'une étude de cas
LE POULIQUEN F., PREDA M., PRETEUX F. (2007). LABIAO : première plate-forme MPEG-4 pour le Langage Parlé Complété assisté par ordinateur. Actes du 5èmes Ateliers de Travail sur l'Analyse d'Images, Méthodes et Applications, 459-464.

METZGER J-L., BARRIL C. (2004). L'insertion professionnelle des travailleurs aveugles et sourds : les paradoxes du changement technico-organisationnel. *Revue française des affaires sociales* 3, 63-86.

MILLET, A., ESTEVE, I. (2008). Pratiques langagières bilingues et multimodales de jeunes adultes sourds. Acte des XXVIIèmes Journées d'Etude sur la Parole – JEP 2008.

MOTTEZ B. (1977). A s'obstiner contre les déficiences, on augmente souvent le handicap : l'exemple des sourds. *Sociologie et sociétés*, vol. 9-1, 20-32.

POWER M.-R., POWER D. (2004). Everyone Here Speaks TXT: Deaf People Using SMS in Australia and the Rest of the World. *Journal of Deaf Studies and Deaf Education* 9, 333-343.

RAZIK J., MELLA O., FOHR D., HATON J.-P. (2008). Transcription automatique pour malentendants : amélioration à l'aide de mesures de confiance locales. Actes des XXVIIèmes Journées d'Etude sur la Parole - JEP 2008.

RENARD M. (2001). Surdités, accessibilité et illettrisme. *Les cahiers de l'Actif*, 298-301.

REZZOUG N., GORCE P., HELOIR A., GIBET S., COURTY N., KAMP J. F., MULTON F., PELACHAUD C. (2006). Agent virtuel signeur, Aide à la communication pour personnes sourdes. Actes de la 4ème conférence pour l'essor des technologies d'assistance, Handicap2006.

OMS. (2014). Surdité et déficience auditive. *Aide-mémoire 300*. (<http://www.who.int/mediacentre/factsheets/fs300/fr/>)

Orosanu L., Juvet D. (2013). Comparison and Analysis of Several Phonetic Decoding Approaches. Actes de TSD - 16th International Conference on Text, Speech and Dialogue. Springer Verlag, 8082, 161-168.

Navigation Aveugle des Pages Web sur Dispositifs Mobiles

Waseem Safi Fabrice Maurel Jean-Marc Routoure Pierre Beust Gaël Dias
University of Caen Basse-Normandie
GREYC CNRS UMR 6072
Bd Maréchal Juin
14032 Caen Cedex - France
firstName.lastName@unicaen.fr

Résumé. De nombreuses techniques spécialisées aux personnes déficientes visuelles ont réussi à extraire les informations affichées sur des écrans numériques et ont réussi à transformer ces informations d'une manière linéaire, soit dans un format écrit sur des dispositifs spéciaux en Braille ou en une sortie vocale. Toutefois, les lecteurs d'écran ne transforment pas la structure 2-dimensionnel de la page web naviguée. Dans cet article, nous proposons une nouvelle technique vise à renforcer la capacité des personnes déficientes visuelles à naviguer le Web en se concentrant sur l'amélioration de l'accès vibrotactile non-visuel des pages web sur dispositifs mobiles, basée sur l'extraction et la réorganisation de la structure de textes et les éléments graphiques des pages web, et de conversion automatique de ces structures visuelles et des informations textuelles dans des pages vibrantes utilisant un langage vibro-tactile graphique.

Abstract.

Blind Navigation of Web Pages on Touch-Screen Devices

Many techniques specialized for Visually Impaired People succeeded to extract the information displayed on digital screens and succeeded to transform this information in a linear way either into a written format on special Braille devices, or into a vocal output using text to speech synthesizers. However, screen readers failed to transform the 2-dimensional structure of the navigated web page. In this paper, we propose a new technique aims to enhance the Visually Impaired People ability to navigate the Web by focusing on improving non-visual vibrotactile access to web pages on touch-screen devices, based on extraction and re-organization the structure of texts and graphical elements for web pages, reformatting and converting automatically these visual structures and textual information into vibrating pages using a graphical vibro-tactile language.

Mots-clés : Dispositifs mobiles, personnes déficientes visuelles, langage vibro-tactile graphique.

Keywords: Touch-screen devices, visually impaired people, vibro-tactile.

1 Introduction

In October 2013, the world health organization estimated that the number of (Visually Impaired People) VIP in the world is 285 million, 39 million of them are blind, and 246 million of them have low vision ¹. VIP depend on screen readers in order to deal with computer operating systems and computational programs. One of most important and desired targets by VIP is navigating the Web, considering the increased importance and expansion of web-based computational programs. Screen readers present some solutions to navigate the Web, either by transforming a web page into a written Braille, or into a vocal output. Some screen readers installed on touch devices transform a web page into a Vocal-Tactile output. But there are some drawbacks for these proposed solutions: on the one hand, the Braille techniques are costly, and only few number of VIP have learned Braille (in France, there are about 77 000 visually impaired people and only 15 000 of them have learned Braille) ². On the other hand, transforming the information of a web page into a vocal format might not be suitable in public and noisy environments. Finally most of Braille solutions are not suitable for mobile devices (Maurel et al., 2012). In addition to these drawbacks, the most important one is the failure to transform the 2-D web page structure, because as reported by many authors, perception the 2D structure greatly improves navigation efficiency and memorizing the information because it allows high level reading strategies (rapid or cursory reading, finding or locating information,...) (Maurel et al., 2003). Our work focuses on developing and evaluating a sensory substitution system based on vibro-tactile solution which may solve the mentioned drawbacks.

¹ <http://www.who.int>

² <http://www.opc.asso.fr/>

Where we study how to increase the VIP perception of a 2-D web page structure, and how to enhance their techniques to navigate the Web on touch-screen devices. This suggested solution is very cheap comparing with prices of Braille devices, and also it could be more efficient in noisy and public environments comparing with vocal-tactile solutions. Our contribution is three-fold:

- Designing a Tactile Vision Sensory System (TVSS) represented by an electronic circuit and an android program in order to transform light contrasts of touch-screen devices into low frequencies tactile vibrations, and running a series of experiments with blind persons in order to validate our hypotheses,
- Designing an algorithm for segmenting web pages automatically in order to support the VIP by a way which may enhance their ability to navigate the textual and graphical contents of web pages, and
- Running a series of experiments with sighted persons in order to compare the differences between automatic and manual web page segmentation.

The paper is organized as following: first, in section 2 we view the state of the art for VIP targeted technologies, and we propose our own framework. The first pre-tests achieved with blind persons will then be described in the third section, and an analysis of results will be presented in the fourth section. In section 5, the state of the art for web pages segmentation methods is reviewed. In section 6, our supervised hybrid segmentation method is presented. In section 7, we present the results obtained in an experiment, where the differences between automatic and manual web page segmentation are compared. Finally, perspectives and conclusions are presented.

2 VIP Targeted Technologies and Proposed Framework

Current products for VIP such as screen readers depend mainly on speech synthesis or Braille solutions, such as ChromeVox³, Windows-Eyes⁴, and Jaws (Job Access With Speech)⁵. Braille displays are complex and expensive electromechanical devices that connect to a computer and display Braille characters. Speech synthesis engines convert texts into artificial speech. Some screen readers can support a tactile feedback when working on touch devices, such as Mobile Accessibility⁶, Talkback⁷ for Android, and VoiceOver⁸ for iPad. Many of these products propose shortcuts for the blind user to display a menu of HTML elements existed in the web page, for example headers, links, and images. But, the main drawback of all these products is that they transfer the web page information into a linear way, and without any indication for the global web page structure (2D layouts). Many researches tried to enhance the way by which VIP interact with web pages, such as (Alaeldin et al., 2012), that proposed a tactile web navigator to enable blind people to access the Internet. This navigator extracts texts from web pages, and sends these texts to a microcontroller responsible of displaying the text in Braille language using an array of solenoids.

A tactile web browser for hypertext documents has been proposed by (Rotard et al., 2005). This browser renders texts and graphics for VIP on a tactile graphics display, and it supports a voice output to read textual paragraphs and to provide a vocal feedback. The authors implemented two exploration modes, one for bitmap graphics, and another one for Scalable Vector Graphics. Main drawback of this proposed system is that it needs a pin matrix device, which is expensive and cannot be integrated with handled devices. Another interesting model called MAP-RDF ("Model of Architecture of web Pages") (Boulssa et al., 2009) proposed a method to improve the accessibility to visual information for blind persons. This model allows representing the structure of a web page, and provides the blind users with an overview of the web page layout and the document structure semantics. The main drawback of this model is that it could be applied only on well structured web pages which contain meta-data, so it could not be applied to most web pages which rarely contain meta-data. Tactos is a suggested perceptual interaction system (Lenay et al., 2003). It consists of three elements: 1- tactile simulators (two Braille cells with 8 pins) represent a tactile feedback system, 2- a graphics tablet with a stylus (represents an input device), 3- computer (Tixier et al., 2013). The graphics tablet and the stylus allow the user to explore graphical contents on the screen such as circles, rectangles, and characters. While the user explores the contents, the system transforms pixels under the stylus into tactile stimulation on the Braille cells. 30 prototypes of Tactos have been released, to be used by a lot of users in many domains. Tactos has been successfully used to recognize simple and complex shapes.

First glance could be defined as the ability -in a blink of an eye- to understand the document layout and its structural semantics (Maurel et al., 2012). We aim of our work to increase the ability of visually impaired persons to understand the web page 2-dimentional layout in order to enhance their tactics to navigate the Web. A commercial tablet connected to a vibro tactile set-up is used for that. The first phase in our model is to extract visual structures in the navigated web page, and to convert these visual blocks to zones (segments) for facilitating the navigation in later phases. We achieve this phase depending on a hybrid segmentation method. Then the system will represent on the tablet screen the extracted visual elements as symbols using a graphical language (this language is under-development). The third phase is to

³ <http://www.chromevox.com/> [Access 01/03/2015]

⁴ <http://www.synapseadaptive.com/gw/wineyes.htm> [Access 1/03/2015]

⁵ <http://www.freedomscientific.com/> [Access 01/03/2015]

⁶ <https://play.google.com/store/apps/details?id=es.codefactory.android.app.ma.vocalizerfrdemo&hl=fr> [Access 01/03/2014]

⁷ <https://play.google.com/store/apps/details?id=com.google.android.marvin.talkback&hl=fr> [Access 01/03/2014]

⁸ <http://www.apple.com/fr/accessibility/> [Access 01/03/2015]

browse these graphical symbols depending on size of the used touched screen device, and then in the fourth phase, our system provides a vibro-tactile feedback when the blind user touches the tablet. The intensity and the frequency of the vibration depends mainly on gray level under the finger. In this paper, we focus only on the first and fourth phases which specialize in extracting visual structures in the navigated web page, and in giving the user a vibro-tactile feedback by transforming light contrasts of touch-screen devices into low-frequencies tactile vibrations. To achieve the desired system, we have designed an electronic circuit which controls two micro-vibrators placed on the hands. A Bluetooth connection with an android tablet allows controlling the vibration intensity (Amplitude) of vibrators. An Android dedicated program on the tablet displays an image on the screen and detects where the user touches the tablet screen. The gray level of touched points is transmitted to the embedded device in order to control the vibration intensity. Only one micro-vibrator was used for pre-tests described in this paper. Figure 1 illustrates the designed electronic circuit, and the used vibrator.

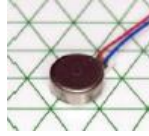


Figure 1. (a) The used micro-vibrator



Figure 1. (b) The designed electronic circuit

Figure 1. (a) The used micro-vibrator: the range of vibration frequency goes from 20 Hz up to 260 Hz. (b). The embedded system designed with the Bluetooth module.

3 Pre-tests Protocol

3.1 Objectives of Pre-tests

Our objective of the designed protocol is enhancing the ability of VIP to recognize the 2-D structure of a web page. In order to test the prototype mentioned in section 2, we designed some images contain different structures (detailed in section 3.2), and we tested the prototype firstly on 15 sighted persons (their eyes were closed) (Maurel et al., 2012), and later on 5 blind persons. Testing the protocol on sighted and blind persons gave us a more understanding of tactics and strategies achieved by sighted and blind persons to navigate the designed structures. This will be useful in designing the desired graphical vibro-tactile language (all results are detailed in next sections).

3.2 Designed Protocol for Vibro-Tactile Access

Each experiment (either for sighted or blind persons) consists of 4 ordered phases of training (learning task), and four ordered phases of evaluation (evaluation task). All the experiments were filmed, and the designed program stocked many parameters in log files (coordinates X, Y, pressure on the screen, and the time at each touch). Figure 2 presents the 4 images of training phases, and figure 3 presents the 4 images of evaluation phases.



Figure 2 (a). Image a of training task



Figure 2 (b). Images b (NT2), c (NT3), d (NTG)

Figure 2. Images of training task



Figure 3 (a). Image a (Evaluation task)

Figure 3 (b). Images b (IDP1), c (IDP2), d (IDP3)

Figure 3. Images of evaluation tasks.

In the training task, each user discovered firstly the graphical elements in each image presented in figure 2 (images a, b (NT2), c (NT3), and d (NTG)), and users were informed about names of graphical elements. The name of each image NT2, NT3, NTG, indicates how many transitions are necessary to access the square center, for example NT2 proposes 2 transitions to access the center of the square.

The evaluation task consists also of 4 phases, the first one allows to discover the image 3.a and to name each square inside it, then next phases are about images 3.b, 3.c, and 3.d, where we asked users to discover contents of each image, then to describe these contents, and to redraw discovered elements inside each image. We chose these images depending on following considerations:

- Image 3.a contains all squares on which users have trained in the training task, so it could test the ability to memorize and to distinguish the shapes.
- Image 3.b contains 3 rectangles with matched sizes and with vertical order, and the image 3.c contains 3 rectangles with different sizes and many relations of directions, so testing images 3.b, and 3.c could test the ability of distinguishing sizes, and distinguishing relations of directions.
- Image 3.d contains different shapes (a rectangle and a polygon), so it could test the ability to distinguish different shapes in the same image.
- The tested images contain examples of expected results of the segmentation process, so success of distinguishing these shapes by blind users could be an indicator of their ability to distinguish results of segmenting web pages.

The results of pretests with sighted persons were already published in (Maurel et al., 2012). Table 1 presents some results of the experiment for images NT2, NT3, NTG (the time required to distinguish graphical elements and the number of errors for the 15 sighted persons. The users have been asked to name the shapes in figure 3.a, and for each shape, we evaluated the number of correct and incorrect answers). In table 1, number of errors represents the number incorrect answers.

Shape Name	Average Time in Seconds	Number of errors
NT2	28	2
NT3	36	6
NTG	22	3

Table 1. Results of experiments of sighted persons for images NT2, NT3, NTG. (Table extracted from (Maurel et al., 2012))

We notice from table 1 that the lowest number of errors is assigned to image NT2, and the largest time and max number of errors is assigned to image NT3.

3.3 Experiment Steps with Blind Persons

The tests performed with each one of the 5 blind persons consisted of following: personal and technical questions, explanations of the test objective, a training task, and finally an evaluation task. The approximated average time for the test for each person is about 1 hour.

3.3.1 Personal and technical questions

Before starting the tests with the 5 blind persons, we asked them to support us with some information about their age and date of their blindness. Table 2 summarizes answers of personal questions.

User-ID	0	1	2	3	4
Age (Years)	63	67	59	56	36
Sex	Male	Female	Male	Female	Female
Date of the blindness	since birth	since 32 years	since 25 years	since 10 years	since 15 years

Table 2. Personal information of the blind persons

We also asked users to provide us with some technical information about their experience in dealing with operating systems, screen readers, and what are the main problems when they navigate the Web. Table 3 shows a summary of answers for these technical questions. The first two columns indicate the number of operating systems (either Windows or Linux) used either on fixed or portable computers. The third, fourth, and fifth columns indicate the number of users who use JAWS (Job Access With Speech), NVDA (NonVisual Desktop Access), and ORCA, either on fixed or portable computers.

	Windows	Linux	JAWS	NVDA	ORCA
Number of users with fixed computer	4	0	4	2	0
Number of users with portable computer	2	1	2	1	1

Table 3. Used operating systems and screen readers

No one of the five blind persons uses a tablet, and the screen readers used with cellular phones are Talks and MobileSpeak with Nokia, and Voiceover with iPhone. Only one of the 5 persons uses a telephone to access the Web (access via iPhone). The main problems of accessing the Web via fixed or portable computers, or via iPhone telephone were: problems of access to Flash files, problems of AJAX technologies, and no ability to know the global structure of web sites. (These problems have been reported to us by the 5 users).

3.3.2 Explaining the objective of the test

To give the blind persons a good idea about the test, we explained in details what are the objectives and the phases of each tasks, and described contents of the embedded system; we also explained the final objectives of the project, and why we concentrate on vibro-tactile technique regardless of other techniques. This phase was important to initiate users for accepting kindly the test and for doing their best to interact with next steps as correctly as possible.

3.3.3 Training and learning task

In this training task, the user discovers the graphical elements in each image presented in figure 4 (images a, b (NT2), c (NT3), and d (NTG)), and the users were informed of each shape name. This task was very important for users to test the system before the evaluation task, and to know exactly how the system transforms different the grey level under the touched points on the tablet screen to a vibration mode. It is also very useful for users to control their speed of mapping the screen either to discover either the borders or the contents. During this task, the program recorded the touching information in log files (X, Y coordinates, Pressure, and Time). Table 4 indicates training times in minutes for each user, and for each image in figure2.

User ID / Image	ID0	ID1	ID2	ID3	ID4	Total	Average
A	4,99	2,60	3,55	9,97	3,96	25,08	5,02
b (NT2)	4,40	3,11	0,99	3,30	1,00	12,80	2,56
c (NT3)	2,81	6,29	2,54	2,85	1,15	15,65	3,13
d (NTG)	3,11	3,98	2,02	3,22	1,32	13,65	2,73
Total	15,31	15,99	9,10	19,34	7,43		

Table 4. Times of training task for each user (in minutes)

We notice from table 4 that discovering the first image takes more time, and it is normal because it is the first experiment for blind users on this prototype. We notice also that there is a significant decrease in time between discovering the first and the last image in the training task. This could be an indicator that training users could decrease the time for discovering graphical elements. We can also notice the significant difference between different tested persons, for example user with ID4 needed 7.43 minutes to scan the images (A, NT2, NT3, and NTG), but the user with ID3 needed 19.34 minutes to scan the same images.

3.3.4 Evaluation task

In this task, firstly we asked each user to discover the image 3.a and to find how many squares inside it and to name each founded square, then we asked them to discover images 3.b (IDP1), 3.c (IDP2), 3.d (IDP3), and to describe them to us, and to redraw discovered shapes. Table 5 illustrates an evaluation of answers for the first question to name squares in image 3.a (The blind users have been asked to name the shapes in image 3.a, and for each shape, we evaluated the number of correct and incorrect answers). (In tables 5 and 6, the symbol ✓ represents a correct answer for the touched shape, and the symbol X represents an incorrect answer or inability to select the name of the touched shape).

User-ID/ Square Name	ID0	ID1	ID2	ID3	ID4	Number of errors
NT2	✓	✓	✓	X	✓	1
NT3	✓	✓	X	X	✓	2
NTG	✓	X	✓	X	✓	2

Table 5. Results of questions for squares in image 3.a

We notice from table 5 that the lowest number of errors is assigned to image NT2, and it is the same result which we obtained during tests with sighted persons. Results of answers for other questions related to images IDP1, IDP2, and IDP3 are summarized in table 6.

User-ID	ID0			ID1			ID2			ID3			ID4		
IDP	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3
Answers about number of rectangles	✓	✓	X	✓	X	X	✓	X	✓	X	✓	✓	✓	✓	✓
Answers about sizes of rectangles	X	X	X	X	X	X	✓	X	X	X	✓	✓	✓	✓	✓

Table 6. Answers of questions for images IDP1, IDP2, IDP3

We notice from data in tables 4, 5, and 6, that the best performance is for the user with ID4, and this may be because that this female user is the youngest between others, and it could be because that she was the only one that has already used touched devices (an iPhone in her case working with VoiceOver). After answering questions about each image of images (IDP1, IDP2, IDP3), we asked each user to redraw the graphical elements founded in each touched image. Figure 4 views the redrawing results of the user ID4 (ID4 is the female user who gave best answers).

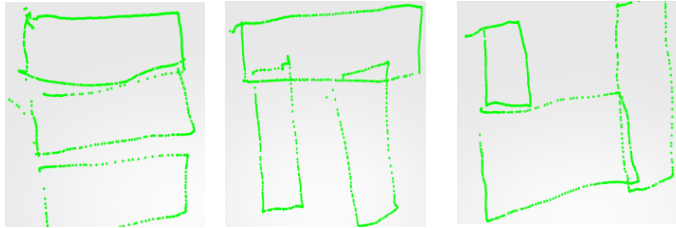


Figure 4. Results of redrawing images IDP1, IDP2, IDP3 for the user ID4

When comparing results of redrawing (Figure 4) with images IDP1, IDP2, and IDP3, we find that the results are interesting, and we can conclude the following:

1. An ability of distinguishing sizes of shapes, because the degree of scaling between redrawn shapes is nearly equal to the degree of scaling between real shapes (IDP1, IDP2, IDP3).
2. An ability of distinguishing relations of directions, because relations of directions (vertical order, left to, right to,...) between redrawn shapes is nearly equal to relations of directions between real shapes.

The average of times in minutes consumed for each evaluation question is summarized in table 7.

User ID / Image	ID0	ID1	ID2	ID3	ID4	Total	Average
A	1,23	9,87	5,39	7,84	1,85	26,17	5,23
IDP1	4,39	14,99	1,41	3,75	1,41	25,96	5,19
IDP2	7,70	9,22	0,79	1,99	13,85	33,55	6,71
IDP3	2,71	12,94	2,81	4,03	12,58	35,06	7,01
Total	16,03	47,02	10,39	17,61	29,68		

Table 7. Times of the evaluation task for each user (in minutes)

4 Analysis of Results

To get an idea about the most touched and the least touched areas on the screen during learning and evaluation tasks, we divided the touched-screen into 16 areas (as in figure 5, r00...r03, r10...r13, r20...r23, r30...r33), and calculated the average of touches in each area for all users.

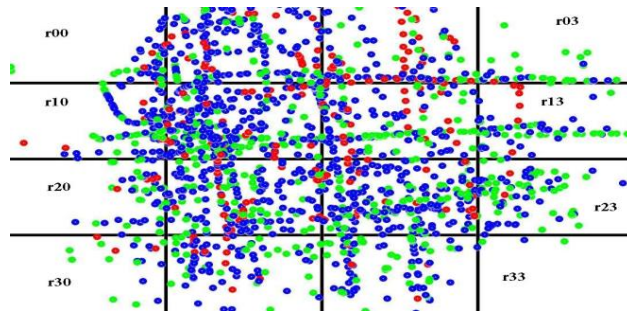


Figure 5. The 16 areas of the touched-screen of user with ID4 (Red points represent touched points with max pressure values, blue points represent touched points with pressure values less than the max and greater than the average, green points represent touched points with pressure values less than the average)

We have founded that the most touched areas are r12, r11, r22, r21, and the least touched areas are r30, r33, r32, r00 as described in figure 6. This information could be useful in next phases of our research in completing the graphical vibro-tactile language by putting the important information in the most touched areas.

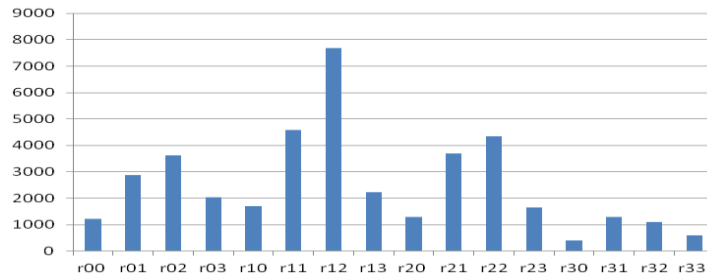


Figure 6. Most and least touched areas on the touched-screen

During analysis the results, we have noticed that there are a lot of differences between the pressure values for all users (Pressure value depends on the used tablet; in these experiments we have used Asus Model TF101 with Android operating system). To analyze pressure values, we calculated the max pressure value between all users, it was 3.19, and the average was 1.73, then we redrew the touched points for each user with considering that points with pressure values equal to the max value have been drawn in red color, points with pressure values less than the max and greater than the average have been drawn in blue color, and points with pressure values less than the average have been drawn in green color. Figure 5 represents an example of these points in different colors (for IDP3 evaluation task of the user who has id 4). After analyzing all the images drawn for all users, we have noticed that majority of red points (max pressure) are in images for which users gave right answers. This notice may be useful in designing our graphical vibro-tactile language, since we can observe when the user decreases or increases his touch pressure. The increasing of pressure may indicate that the user touches graphical elements interesting for him, and the decreasing may indicate that the user touches graphical elements non-interesting for him. During the tests we observed also that users try sometimes to scan the screen very quickly, it might be because they try to get a lot of information in a short time.

5 State of The Art for Segmenting Web Pages

Segmenting a web page is a fundamental phase for understanding its global structure. Extracting the global structure of web pages is useful in many domains such as information retrieval, data extraction, and similarity of web pages. Many approaches have been suggested for segmenting web pages, such as:

- DOM-based segmentation: it depends on analyzing the DOM tree (Document Object Model), and extracting the main structure of web pages depending on HTML tags as described in (Sanoja et al., 2013). In their paper, the authors determine first the layout template of a web page, and then it divides the page into minimum blocks, and finally collects these minimum blocks into content blocks.
- Vision-based segmentation: this method divides the web page depending on the visual view of web page contents on a web browser. The well known tool VIPS (VIsion based Page Segmentation) (Deng et al., 2003) is based on this approach.
- Image processing based segmentation: this approach captures an image for the visual view of a web page, and then depends on image processing techniques to divide the captured image into sub-blocks (Cai et al., 2004) (Cao et al., 2010).
- Text-based Segmentation: this approach focuses on extracting only information about texts in a web page. After dividing the web page into blocks of texts, it could be possible to find the semantic relations between these textual blocks. This method is useful in many information retrieval domains such as question answering applications (Foucault et al., 2013).
- Fixed-length segmentation: this approach divides the web pages into fixed length blocks (passages), after removing all HTML tags, where each passage contains a fixed number of words (Callan, 1994).
- Densitometric analysis based segmentation: this approach depends on methods applied in quantitative linguistics, where text-density refers to a measure for identifying important textual segments of a web page (Kohlschütter et al., 2008).
- Graph-based segmentation: This approach depends on transforming the visual segments of a web page into graph nodes, then applying many common graph methods on these nodes for combining them into blocks, or for making a clustering for these nodes. Some common works which depend on this approach are (Chakrabarti et al., 2008) (Liu et al., 2011).

6 Suggested Hybrid Segmentation Algorithm

Most of segmentation algorithms render first the web page using a web browser, and then segments the HTML elements into many blocks depending on the visual layout. Our constructed hybrid segmentation algorithm has been tested on 154 pages collected manually from many newspapers and e-commerce sites (www.leparisien.fr, www.lefigaro.fr, www.liberation.fr, www.amazon.fr, www.materiel.net, www.photobox.fr). The results have been integrated with our under-development Android program. The obtained results are promising because the segmentation algorithm can efficiently extract the web page blocks depending on the visual structure, and the algorithm can also convert correctly these blocks into zones (clustering the blocks). Our algorithm mixes three segmentation approaches, vision-based segmentation, DOM-based segmentation, and graph-based segmentation.

6.1 Vision-Based Approach

In this phase, we render the web page using Selenium⁹ web driver and Mozilla FireFox browser, and we get its visual structure by Java-script code injection inside the HTML source code of the rendered web page. The obtained visual structure indicates the global hierarchy of the rendered web page. This phase assigns additional information for each DOM HTML element such as XPath and bounding box (location [X0, Y0], and size [height and width]).

The input of this phase is a web page HTML source code, and its output is an augmented HTML web page with injected information about bounding boxes and DOM XPath for each HTML element. In next sections, we refer to

⁹ <http://www.seleniumhq.org/> [Access 01/03/2015]

bounding boxes by blocks (i.e. each bounding box represents an HTML element, and may contain other bounding boxes).

6.2 DOM-Based approach

After obtaining the visual structure of a web page, we analyze its DOM structure by applying filters and re-organization rules for enhancing results of next phases. We divide the DOM elements depending on the specification of HTML5 content models proposed by the World Wide Web Consortium (W3C)¹⁰. This specification divides the HTML tags into 7 categories (Metadata content, Flow content, Sectioning content, Heading Content, Phrasing content, Embedded content, and Interactive content). The first applied filter is Metadata-Content-Filter, which deletes all the elements considered as metadata content elements except "title" tag. We delete the other tags because they do not contain useful visual information in next steps. The deleted tags are "base", "command", "link", "meta", "noscript", "script", and "style". We add also to this filtered group the following tags "comment", "br", and "doctype".

The second applied filter is Dead-Nodes-Filter, where it deletes all HTML nodes that do not affect on the appearance, for example nodes with height or width equals to "0px" (zero pixel); or nodes with style properties ("display:none" or "visibility:hidden" or "hidden:true"). After applying the previous filters, we apply some re-organization rules in order to enhance visualizing the information in next phases. One example of these rules is Paragraph-Reorganization-Rule, where this rule re-constructs all paragraph child-nodes in one node contains the extracted sub-texts. We made this rule after analyzing many DOM structures, and observing that the text in some paragraph nodes is distributed between many child-nodes such as <i> (italic), , (bold), , (emphasized), <small>, <mark> (marked), (deleted), <ins> (inserted), <cite> (defining a title of work), <u> (underline), and <sub> (subscript). So extracting these sub-texts and collecting them in one text is useful and more efficient for visualizing them as one block in next phases rather than visualizing them as many separated blocks. This rule was applied on all tags of type <P> (paragraph), and on all tags existed in the heading content category (specification of HTML5 content models), this group contains the following nodes (h1, h2, h3, h4, h5, h6, and hgroup). We also apply this rule on many other HTML tags which might contain textual child-nodes such as <a> (hyperlink tag), <abbr> (abbreviation), <acronym> (this tag is not supported in HTML5), <address> (contact information for the author/owner), <bdi> (Bi-directional Isolation), <button>, <label>, (list element), and <q> (quotation). We used Jsoup tool¹¹ in order to access to a web page DOM structure, and getting its HTML hierarchy. The result of this phase is a filtered DOM-tree; each of its nodes is visible and contains XPath and bounding box information. The designed filters and re-organization rules were integrated with our framework, and then we applied these rules and filters on the vision-based segmented web pages (154 pages mentioned previously). After getting the filtered DOM-tree for each page, we represented the obtained bounding boxes on the used tablet Samsung GALAXY Tab 2 (10.1inch, dimensions HeightXWidthXDepth 175.3X256.7X9.7 mm, Android version 4.1.2), after making a scaling of sizes of bounding boxes to be appropriate with the new size of used tablet.

6.3 Graph-Based approach

After segmenting the web page depending on its visual structures and analyzing its DOM-structure, we apply a new graph-based segmentation algorithm called "Blocks2Zones Clustering" in order to group many similar blocks together in one zone. Clustering many blocks together is necessary in order to decrease the number of viewed blocks in some interfaces (instead of viewing many blocks, we view one zone which represents these blocks and then the user can navigate intra-elements inside the zone by double clicking on the graphical element of the chosen zone), and to group closed blocks in one zone. The pseudo-code of the proposed algorithm is:

```

Blocks2Zones Clustering Algorithm
Input (Blocks, N° of desired Zones)
Output: Graph of N nodes (N Zones)
1- Transform the blocks into a graph (Un-Directed graph)
    1.1. Blocks → Nodes,
    1.2. Make relations between the nodes, and assign weights for these relations.
2- If number of zones <= number of blocks
    end the algorithm,
Else
3- Find the node with the smallest size (node A)
    (size of node == size of the rectangle bounds the node )
4- For node A, find the connection which has the largest weight (node B).
5- Group the nodes A, and B (A+B).
6- Repeat steps 3-4-5 till number of blocks == number of zones

```

¹⁰ <http://www.w3.org/TR/2011/WD-html5-20110525/content-models.html> [Access 01/03/2015]

¹¹ <http://jsoup.org/> [Access 01/03/2015]

The output graph is described as following: $G = (V, E)$, where G is undirected graph, V is a set of vertices (nodes or zones), and E is a set of edges (connections between zones), and $|V|=n$ (number of desired zones). We define the set of vertices as $V=\{v_i: 1 \leq i \leq n \text{ where } n \text{ is the number of desired zones, and } v_i \text{ is set of sub-zones}\}$. We define the set of connections as following $E=\{e_j: e_j(v_{j1}, v_{j2}) : v_{j1} \in V, \text{ and } v_{j2} \in V\}$. To calculate weights between nodes, we used the Euclidian distance between the centers of two nodes (center of a node is the center of the rectangle which bounds the node). In following, an example of applying this algorithm on the main page of web site w3schools.com is given. In this example, we want that the algorithm segments the main page into 9 zones. The first iteration of the algorithm will divide the page into 13 zones (because there are 13 main nodes in the DOM structure – <Div> nodes) as illustrated in figure 7.a, and will construct a graph of 13 nodes as illustrated in figure 7.b.

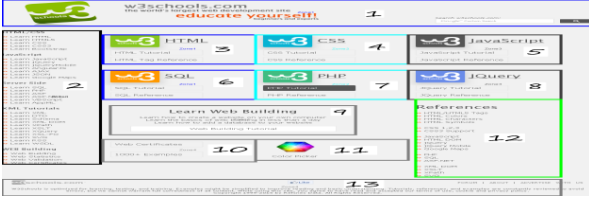


Figure 7.a Segmentation for 13 zones

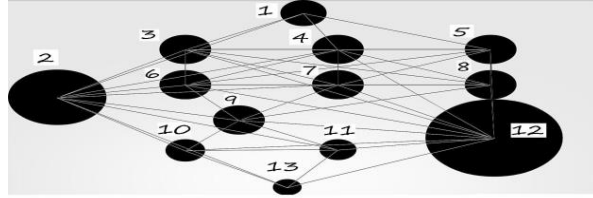


Figure 7.b Constructed graph

Figure 7. First segmentation of www.w3schools.com

Figures 8, and 9, illustrate next iterations of the algorithm and how the 13 zones are converted to 9 zones.

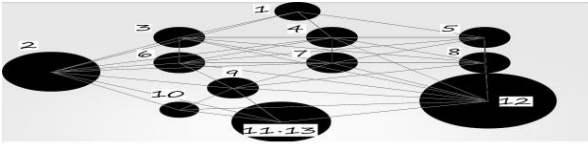


Figure 8.a 12 zones segmentation

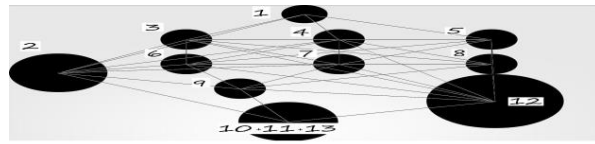


Figure 8.b 11 zones segmentation

Figure 8. 12, 11 zones segmentation of w3schools.com



Figure 9.a 10 zones segmentation

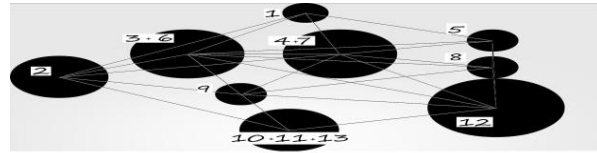


Figure 9.b 9 zones segmentation

Figure 9. 10, 9 zones segmentation of w3schools.com

Applying this hybrid segmentation algorithm on a filtered DOM-tree (obtained from applying Vision-based approach and DOM-based approach) converts a web page to a set of zones, each zone contains many other zones or blocks, since each block represents a visual structure of HTML element and may contain many other blocks. The purpose of the proposed vibro-tactile access protocol is then to transform semantics of symbols in these zones, or blocks, or HTML elements into vibrations with different frequencies and amplitudes.

7 Manual and Automatic Web Page Segmentation Differences

In order to evaluate our algorithm results, and to know how users understand web layout structures based on their visual perception, we made an experiment as following: we asked to 15 volunteers a manual segmentation of different kinds of web pages. The volunteers were of different ages (between 25 and 50 years old), and most of them were informatics specialists. We presented, for each volunteer, 4 printed copies (A4 size papers) of 8 web pages (2 pages from www.cddiscount.com, 2 pages from www.photobox.com, 2 pages from www.rueducommerce.fr, 1 page from www.w3schools.com, and 1 page from www.leparisien.fr). We asked each volunteer to segment the 4 copies of each web page into 3, 4, 5 and 6 zones with following considerations:

- all the pages are printed in gray scale in order to avoid affecting the colors on the segmentation process. We chose this option because the current version of our algorithm does not depend on color differences between blocks,

- users can segment the page using polygons with minimum of 4 points (triangles are not allowed),
- users can start segmenting from any part or direction of the page (left, right, top, down, or center),
- we asked users to write the ordering number of zones (inside or beside the zone) while they make the segmentation process; this is very useful for us to know how the users start segmenting the pages, and how they end it.

After collecting all manual segmented copies (480 papers : 15 users X 8 web pages X 4 copies), we noticed the following:

- 70% of users do not start the segmentation process for certain number of zones by the same way they start segmenting for other numbers of zones,
- 16% of papers have been segmented starting from the center of the page, 3.5% have been segmented starting from the bottom of the page, and the majority of papers 80.5% have been segmented starting from top of the pages,
- 20% of papers have been segmented vertically, and 80 have been segmented horizontally,
- 92.6% of segments are rectangles, and 7.4% of segments are polygons with more than 4 points,
- finally, we noticed that it is very difficult to detect a segmentation method common between all users; since each user segments the pages depending on his understanding of the web page layout structure, on his visual perception of the visible elements, and on his interests and visual experience.

We have run our algorithm with the 8 mentioned web pages. The algorithm segmented each web page into 3, 4, 5, and 6 zones. The comparison between manual and automatic segmentation is illustrated in tables 8 and 9.

	Page1	Page2	Page3	Page4	Page5	Page6	Page7	Page8	Total
3 Zones	1	0	0	6	3	0	9	1	20
4 Zones	1	5	4	3	3	0	3	2	21
5 Zones	4	4	1	4	3	0	0	2	18
6 Zones	4	1	2	2	4	1	0	1	15
Total	10	10	7	15	13	1	12	6	74

Table 8. Strong-criterion based matching results of manual and automatic segmentation

Table 8 illustrates matching results based on what we call “Strong criterion”. Depending on this criterion, we can consider that 2 segmentation results are matched if the results are 100% identical without any difference. We can conclude from table 1 that pages segmented automatically and manually into 4 zones are more matched than pages segmented into other numbers of zones. The percentage of identical matching depending on the strong criterion is 15.42% (74 identical matched results of 480 segmented copies).

Table 9 illustrates matching results based on what we call “Weak criterion”. Depending on this criterion, we can consider that 2 segmentation results are matched if at least 50% of the results are identical. We can conclude from table 2 that pages segmented automatically and manually into 3 zones are more matched than pages segmented into other numbers of zones. The matching percentage depending on the weak criterion is 47.5% (228 weak identical matched results of 480 segmented copies).

	Page1	Page2	Page3	Page4	Page5	Page6	Page7	Page8	Total
3 Zones	8	10	8	9	9	2	12	6	64
4 Zones	7	5	7	8	10	4	13	7	61
5 Zones	8	6	6	5	7	6	8	9	55
6 Zones	6	4	7	5	8	5	10	3	48
Total	29	25	28	27	34	17	43	25	228

Table 9. Weak-criteria based matching results of manual and automatic segmentation

8 Conclusion and Perspectives

In this paper, we summarized our current work which aims to design an approach for non-visual access to web pages on touch-screen devices. The designed vibro-tactile protocol transforms the information viewed on the screen and touched by users to vibration by transforming light contrasts of touched pixels into low-frequencies tactile vibrations. The obtained results are interesting, since we used in these experiments only one vibration motor of low level quality (phone vibrator), and the learning period was very short, so there are many enhancements to be achieved in next versions either on the hardware/software level or on the level of learning phase (increasing the number and quality of micro-vibrators, making more control on frequencies and amplitudes sent to micro-vibrators, adding vocal abilities to the current

approach, ...). Next steps in this research will be 1) adding some image processing techniques in order to enhance the proposed segmentation algorithm 2) adding advanced techniques in text summarization to facilitate navigating textual information, 3) making real experiments with blind VIP to study effects of suggested segmentation algorithm on their web navigation models.

References

- Maurel F., Dias G., Routoure J-M., Vautier M., Beust P., Molina M., Sann C. (2012). Haptic Perception of Document Structure for Visually Impaired People on Handled Devices. *Procedia Computer Science*, Volume 14, Pages 319-329.
- Maurel F., Vigouroux N., Raynal M., Oriola B. (2003). Contribution of the Transmodality Concept to Improve Web Accessibility. In *Assistive Technology Research Series, International conference; 1st, Smart homes and health telematics; Independent living for persons with disabilities and elderly people*. Volume 12, Pages 186-193.
- Alaeldin A., Mustafa Y., Sharief B. (2012). Tactile Web Navigator Device for Blind and Visually Impaired People. In *Proceedings of the 2011 Jordan Conference on Applied Electrical Engineering and Computing Technologies, Jordan*, Pages 1-5.
- Rotard M., Knödler S., Ertl T. (2005). A Tactile Web Browser for the Visually Disabled. In *Proceedings of the sixteenth ACM Conference on Hypertext and Hypermedia. ACM, New York, NY, USA*. pages 15-22.
- BouIssa Y., Mojahid M., Oriola B., Vigouroux N. (2009). Accessibility for the Blind, an Automated Audio/Tactile Description of Pictures in Digital Documents. *IEEE International Conference on Advances in Computational Tools for Engineering Applications*, Pages: 591 – 594.
- Lenay C., Gapenne O., Hanneton S., Marque C., Genouëlle C. (2003). Sensory Substitution, Limits and Perspectives. In *Touch for Knowing Cognitive psychology of haptic manual perception*. Pages: 275-292.
- Tixier M., Lenay C., Le-Bihan G., Gapenne O., Aubert D. (2013). Designing Interactive Content with Blind Users for a Perceptual Supplementation System. In *Proceedings of the 7th International Conference on Tangible, Embedded and Embodied Interaction, Barcelona, Spain*, Pages 229-236.
- Sanoja A., Gañçarski S. (2013). Block-o-Matic: a Web Page Segmentation Tool. *La 29 édition des journées "Bases de Données Avancées"*.
- Deng C., Shipeng Y., Ji-Rong W., Wei-Ying M. (2003). VIPS: a Vision-based Page Segmentation Algorithm. *Technical Report MSR-TR-2003-79, Microsoft Research*.
- Cai D., He X., Ma W-Y., Wen J-R., Zhang H. (2004). Organizing WWW Images Based on the Analysis Of Page Layout And Web Link Structure. *IEEE International Conference on Multimedia and Expo*. Volume 1, Pages 113-116.
- Cao J., Mao B., Luo J. (2010). A segmentation method for web page analysis using shrinking and dividing, *International Journal of Parallel, Emergent and Distributed Systems - Network and parallel computing*, Volume 25 Issue 2, Pages: 93-104.
- Foucault N., Rosset S., Adda G. (2013). Pré-segmentation de pages web et sélection de documents pertinent en Questions-Réponses. *TALN-RECITAL-2013*. Pages 479-492.
- Callan J.P (1994). Passage-level Evidence in Document Retrieval. *The Seventeenth Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, Dublin*. Pages: 302-310.
- Kohlschütter C., Nejdl W. (2008). A Densitometric Approach to Web Page Segmentation. *Proceedings of the 17th ACM conference on Information and knowledge management*, Pages: 1173-1182.
- Chakrabarti D., Kumar R., Punera K. (2008). A graph-theoretic approach to webpage segmentation. *Proceedings of the 17th international conference on World Wide Web, USA*, Pages: 377-386.
- Liu X., Lin H., Tian Y. (2011). Segmenting Webpage with Gomory-Hu Tree Based Clustering, *International Journal of Software Engineering and Its Applications*, Volume 6, Pages: 2421-2425.

TALN et IHM : une approche transdisciplinaire pour la saisie de textes de personnes en situation de handicaps

Philippe Boissière¹ Nadine Vigouroux¹ Mustpha Mojahid¹ Damien Sauzin¹ Frédéric Vella¹

(1) CNRS, IRIT - Université Toulouse 3 Paul Sabatier, 118 Route de Narbonne, F-31062 TOULOUSE CEDEX 9, France
Prénom.Nom@irit.fr

Résumé. Ce papier vise d'une part, à présenter une synthèse des méthodes d'optimisation pluridisciplinaires alliant le traitement automatique de la langue naturelle et celui l'interaction homme-machine et d'autre part à soulever les nouveaux verrous scientifiques posés par l'arrivée des nouvelles technologies et des nouveaux modes d'écriture pour la saisie de textes. Cette synthèse abordera les principes d'optimisation des agencements spatiaux des claviers virtuels et ceux de la prédiction de caractères et/ou des mots à afficher en tenant compte des nouveaux supports d'interaction et des nouvelles technologies.

Abstract.

Natural Language Processing and Human Computer Interaction: interdisciplinary approach for text input by persons with disabilities.

This paper aims to present, firstly a synthesis of multidisciplinary optimization methods combining automatic processing of natural language and the human computer interaction and secondly to discuss new scientific challenges raised by new technologies and new modes of writing for text input. This summary will address the principles of optimization of virtual keyboard layout and those of characters and word prediction towards the new technologies (device and interaction technique).

Mots-clés : Agencement spatial, techniques d'interaction, système de prédiction, liste de présentation

Keywords: Layout, interaction technique, prediction system, presentation list

1 Introduction

L'homme dans son histoire a toujours voulu laisser des traces concernant ses connaissances, et communiquer à distance. Pour cela, l'homme utilisait l'écriture manuscrite jusqu'au XIX^{ème} siècle. A cette date fut inventée la machine à écrire. Elle subit plusieurs modifications au niveau de l'agencement des touches. En effet lorsque l'utilisateur saisissait un message les barres de lettres s'entrechoquaient. Cela était dû à la proximité des barres de lettres. A partir de ce constat il la disposition des touches fut changé pour éloigner ces dernières, ce qui a aboutit à l'agencement des claviers physiques QWERTY ou AZERTY. A partir du XX^{ème} siècle, le premier ordinateur utilisé comme outil de travail comportait un clavier de type QWERTY chez les anglo-saxons, et de type AZERTY pour les pays francophones¹. Cependant, aujourd'hui nos claviers sont numériques et les barres de lettres n'existent plus.

Il existe deux façons pour accélérer cette saisie : le réagencement des touches du clavier pour optimiser les déplacements du curseur et/ou la prédiction des caractères ou des mots à afficher pour diminuer le nombre de frappes sur le clavier. Ce papier vise d'une part, à présenter une synthèse des méthodes d'optimisation pluridisciplinaires alliant le traitement automatique de la langue naturelle et celui l'interaction homme-machine et d'autre part à soulever les nouveaux verrous scientifiques posés par l'arrivée des nouvelles technologies et des nouveaux modes d'écriture. Cette synthèse s'appuiera sur l'art sur les systèmes d'assistance (Garay-Vitoria et Abascal 2006), (Antoine et Maurel 2007) et (Martin et Pecci 2007).

¹ <http://fr.wikipedia.org/wiki/AZERTY>

2 Contribution de la représentation spatiale à l'amélioration de la saisie

L'une des problématique dans la saisie de texte est d'améliorer/adapter leur agencement pour accélérer la vitesse de saisie de textes et réduire la fatigue motrice des scripteurs. De cette problématique ont émergé les questions scientifiques sur : la réduction du taux d'erreur de saisie, l'affichage des claviers virtuels sur des petits écrans et la conception des claviers sur les écrans tactiles. Sur ce dernier aspect ont été posées les questions sur la taille des touches par rapport à la taille des doigts et sur la présentation visuelle des touches du clavier qui sont cachées par les doigts et les mains lors de l'interaction. Ces différentes questions seront abordées ci-après.

2.1 Intérêt de l'agencement spatial

La saisie sur le clavier physique AZERTY peut s'effectuer avec 10 doigts et pour un clavier virtuel avec un dispositif de pointage. L'étude expérimentale de (Zhai & al, 2000) a montré que la vitesse de saisie est plus rapide sur un clavier physique AZERTY (75 mots par minute contre 25 mots par minute) que sur un clavier virtuel AZERTY pour des utilisateurs valides. De ce constat de nombreuses études se sont intéressées à l'agencement de claviers virtuels afin d'accélérer la vitesse de la saisie de texte. Deux types d'agencement existent : ambigu et non ambigu. L'agencement ambigu consiste à mettre un nombre de lettres supérieur à un sur une même touche alors que l'agencement non ambigu contient une seule lettre par touche. Ces agencements ont pour but d'optimiser les distances entre les caractères à saisir dans une langue donnée.

2.1.1 Clavier ambigu

Il existe peu de travaux consacrés à l'optimisation des claviers ambigus. Nous mentionnons le clavier ambigu UKO II (Harbusch et al, 2003) (FIGURE 1) — chaque touche de celui-ci correspond à dix caractères—l'agencement général est le résultat d'un algorithme génétique.

Le clavier ambigu de type téléphone (Vigouroux et al. 2004) dispose de 10 touches (FIGURE 2). A chaque touche sont associés trois caractères. Le principe de la sélection d'un caractère est d'effectuer des clics gauches successifs (avec une souris ou un trackball) afin de faire défiler de manière circulaire la liste des caractères. Le premier clic sélectionne le premier caractère de la touche, le second clic sélectionne le deuxième caractère de la touche, et ainsi de suite. Ce principe de sélection est repris par le clavier K-Hermès (Guerrier et al, 2011) (FIGURE 4). L'agencement des touches de ce clavier suit l'ordre alphabétique afin que l'utilisateur mémorise facilement l'emplacement des lettres. La vitesse de saisie de ces deux claviers ambigus a été comparée avec celle du clavier AZERTY. Deux résultats contradictoires ont été montrés. Le résultat de (Vigouroux et al. 2004) montre que la saisie de texte est plus rapide avec le clavier AZERTY que sur le clavier téléphone. Par contre, le résultat de (Guerrier et al, 2011) montre le contraire. La raison de cette différence est double ; a) cela peut-être dû aux dispositifs de pointages utilisés (un trackball dans les travaux de (Vigouroux et al. 2004) et un joystick dans les travaux de (Guerrier et al, 2011)), b) cela peut-être dû à l'agencement des touches et c) à la diversité des troubles moteurs.

Le principe de saisie du clavier K-Thot (Baas et al, 2010) est de sélectionner d'abord en cliquant (non maintenu) n'importe où dans la touche contenant la lettre que l'utilisateur veut afficher, puis sortir le pointeur de la souris, du côté où se trouve la lettre (FIGURE 3). Dans le cas, où l'utilisateur a cliqué sur une touche ne contenant pas la lettre voulue, il peut cliquer à nouveau sur cette touche pour annuler son erreur.

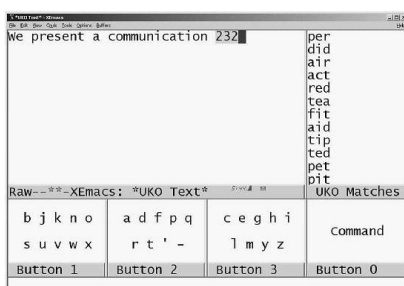


FIGURE 1: Clavier UKO II



FIGURE 2: Clavier téléphone

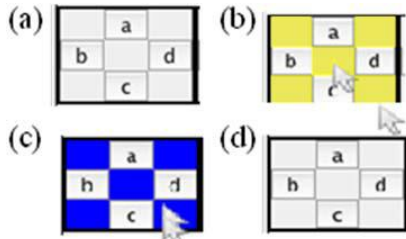


FIGURE 3: Clavier K-THOT

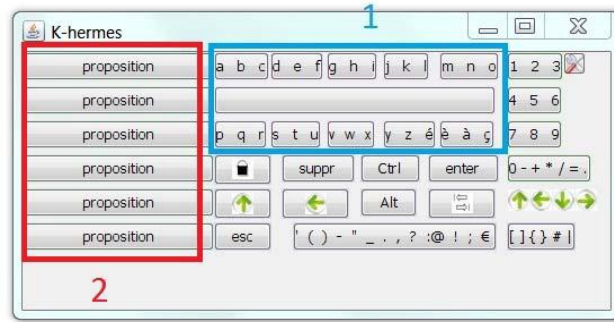


FIGURE 4: Clavier K-Hermès : 1) touches du clavier et 2) liste de prédications des mots

2.1.2 Clavier non ambigu

La majorité des travaux sur l'agencement spatial des touches ont été réalisés sur la langue anglaise. Ils visent à optimiser la position des caractères afin de rapprocher au mieux ceux qui ont le plus de chance d'être saisis consécutivement. Ces approches sont algorithmiques —algorithme glouton (Getschow et Rosen, 1986), (Lewis, 1992) appliqué à la génération du clavier Metropolis (Zhai et al, 2000) ; algorithme génétique (Raynal et Vigouroux, 2005)) appliqué à la génération du clavier GAG— ou sont basés sur les fréquences d'utilisation des caractères de la langue anglaise, comme par exemple FITALY (FIGURE 5), OPTI (FIGURE 6), (MacKenzie et Zhang, 1999).

Dans le cadre de la langue française, nous avons par exemple le clavier GAG (FIGURE 8) et le clavier Annie (Vigouroux et al, 2015) (FIGURE 9). Le clavier GAG (Raynal et Vigouroux, 2005) consiste en l'application d'un algorithme génétique à la langue française, à partir des connaissances extraites de BDLEX (Decalme et Perennou, 1998). Nous remarquons que, pour le français, le clavier GAG est meilleur de 60,27% en WPM (Word Per Minutes ou mots par minutes) que le clavier AZERTY. Le clavier Annie a été conçu par une personne handicapée atteint d'une myopathie (amyotrophie spinale de degré 2). Suite à des entretiens, elle nous a commenté son principe d'agencement : elle a disposé au centre les caractères les plus fréquemment utilisés en essayant de respecter de manière cyclique l'ordre alphabétique. L'explication de cette disposition est due à sa difficulté de déplacement du curseur du dispositif de pointage.

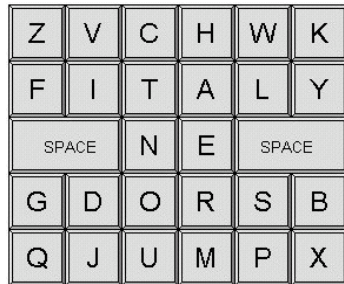


FIGURE 5: Clavier FITALY



FIGURE 6: Clavier OPTI

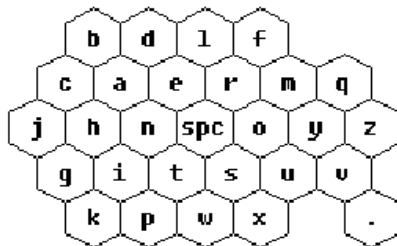


FIGURE 7: Clavier Metropolis

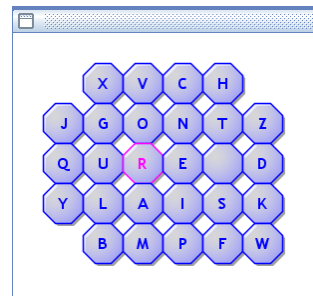


FIGURE 8: Clavier GAG

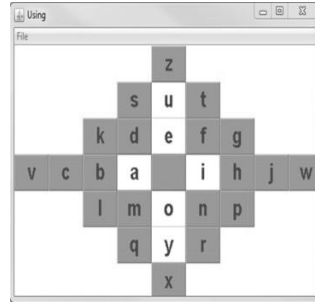


FIGURE 9: Clavier Annie

Ces claviers sont généralement utilisés avec une technique de pointage à base d'une souris, d'un trackball, etc. Cependant, la saisie avec ce type de pointage reste longue et demande une mémorisation des agencements des touches. Avec l'apparition de nouvelles technologies comme par exemple les Smartphones, de nouvelles techniques d'interaction et des modalités gestuelles sont explorées pour tenter d'améliorer la saisie de texte.

2.2 Techniques pour la saisie de texte

L'explosion du tactile questionne à nouveau l'agencement des claviers virtuels. La saisie de texte sur des supports tactiles (téléphone portable, tablette) est très inconfortable car l'utilisateur ne peut pas *sentir* les touches avec ses doigts ou les distinguer (Go et Endo, 2007). L'inconvénient du tactile peut être aussi la taille des touches à atteindre, surtout lorsque la taille des doigts est importante et le fait que ces derniers cachent les informations à l'écran (Go et Tsurumi, 2010). Pour pallier ces problèmes plusieurs études ont été conduites. (Go et Endo, 2007) ont redéfini le centre de la touche et la forme de la touche (par exemple, le clavier CATKey FIGURE 14) en fonction de la position des impacts des doigts sur ces dernières. Quant à (Go et Tsurumi, 2010), ils ont ajouté la technique d'interaction du « Pie Menu » (FIGURE 15) qui propose des caractères en fonction de la position du doigt sur le clavier.

2.3 Entrée de textes par saisie gestuelle

Une modalité de saisie concerne la saisie de texte par la reconnaissance de mouvements sur des petits écrans. Cette modalité a été très explorée en situation de saisie en mobilité. Il existe deux sous-types de saisie : analogique et non analogique. La première (FIGURE 10) est basée sur la reconnaissance d'un caractère ou d'une représentation squelettique d'un caractère (EdgeWrite (Wobbrock et Myers, 2006)). La deuxième est basée sur le pointage de cibles par un stylet. Celles-ci peuvent être les touches du clavier virtuel qui constituent les caractères à saisir (par exemple DASHER (Ward et al, 2000), (FIGURE 11)). Elles peuvent aussi constituer un ensemble de primitives qui définissent un caractère (par exemple UniGlyph, FIGURE 12 (Poirier et Belatar, 2007)) ou des cibles à sélectionner (par exemple, Seven Segment Input (7SI) FIGURE 13 (Al Faraj et al, 2008)). Néanmoins, ces deux modes de saisie gestuelle engendrent des contraintes à la fois cognitives, qui sont dus à la mémorisation des tracés des lettres, et motrices qui demandent beaucoup de déplacements du pointeur pour réaliser une lettre. Ces deux contraintes peuvent provoquer de la fatigue visuelle et motrice.



FIGURE 10: Tracé de la lettre « y » sur le système EdgeWrite

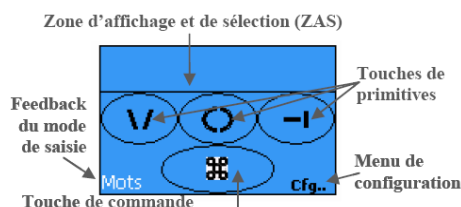


FIGURE 12: Clavier UniGlyph

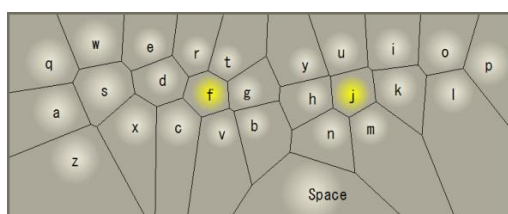


FIGURE 14: Clavier CATKey

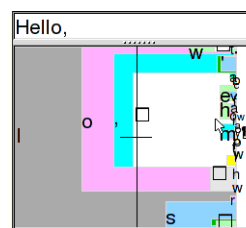


FIGURE 11: Clavier DASHER

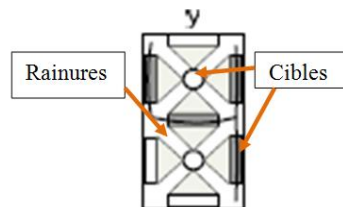


FIGURE 13: Clavier 7SI

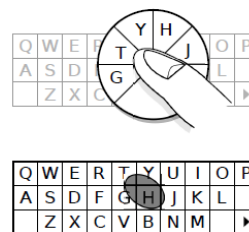


FIGURE 15: Clavier avec Pie Menu

3 Apport de la linguistique à la saisie de textes

Nous venons de voir ce que l'agencement spatial et les techniques d'interaction des claviers virtuels peuvent apporter à l'utilisateur en termes d'ergonomie, et de vitesse de saisie. Est-ce que la vitesse de saisie serait augmentée si nous pouvions également anticiper, voire prédire, ce que l'utilisateur veut dire ou écrire. Or, la pensée de l'utilisateur est par nature riche et complexe, et n'est en rien déterministe. Prédire ou anticiper sa pensée est toute une problématique (voire une gageure) à laquelle se sont heurtés, et se heurtent encore les chercheurs. La linguistique est une aide fondamentale, mais comment mettre en œuvre les analyseurs (morphologique, syntaxique, sémantique et pragmatique) et les connaissances linguistiques associées pour prédire ou anticiper l'écriture de l'auteur? Comment dans cette optique, ces différents outils, peuvent-ils s'articuler entre eux pour apporter les informations nécessaires à cet objectif? Les études statistiques peuvent être aussi utiles pour l'apprentissage des modèles de langage. Sur quels corpus d'apprentissage allons-nous faire tourner nos modèles pour que les systèmes d'aide à la saisie soient performants ?

Les systèmes d'aide à la saisie se divisent en deux classes selon qu'ils font de la prédiction ou de la complétion. Dans un système de complétion, le système ne propose pas de liste de mots, mais complète un mot ou une partie de mot dès qu'il n'y a plus d'ambiguïté (comme le système VITIPI (Boissière et al. 2012)). Un système d'assistance est composé de deux parties: un moteur linguistique qui vise à générer une liste de mots à partir de connaissances linguistiques et/ou statistiques et une interface de présentation des listes de mots. Pour plus de détails, nous renvoyons les lecteurs à deux états de l'art sur les systèmes d'assistance (Garay-Vitoria et Abascal 2006) et (Antoine et Maurel 2007).

3.1 Moteur linguistique

Un moteur linguistique est la partie du système qui prend en charge la prédiction des mots et éventuellement leur correction.

La typologie des moteurs linguistiques se divise en trois approches :

- Une approche purement syntaxique dans laquelle on décrit explicitement toutes les règles (morphologiques, syntaxiques, sémantiques, pragmatiques) ;

- Une approche purement statistique ou probabiliste qui se base sur les probabilités de quoi d'apparition d'un mot en fonction de ceux qui précèdent en utilisant les différentes méthodes en cours dans ce domaine.
- Une approche mixte qui combine approche syntaxique et approche statistique.

La première approche syntaxique consiste à prévoir toutes les règles de grammaire qui représentent la structure syntaxique de la phrase. En programmant les règles d'accords entre le déterminant et le nom, le nom et son adjectif, on peut comme le proposaient les projets Kombe ((Godbert et al. 1992), (Milhaud et al. 1992)) et Illico ((Pasero et Sabatier 1998 a), (Pasero et Sabatier 1998 b)) générer automatiquement les accords. Cette approche se prête bien à la correction automatique de fautes d'orthographe qui fait partie des fonctionnalités du moteur linguistique. Le projet VITIPI (Boissière et Dours 2001), (Boissière et al. 2012) utilise des règles de réécriture des Groupes Phonétiques Orthographiques (GPO) (Perennou et al. 1986) qui, à un phonème donné, associe les différentes graphies. Cela permet la correction de fautes d'orthographe usuelles (Boissière et Dours 1996) des personnes dyslexiques. Le passage d'une langue à une autre nécessite la reprogrammation des règles orthographiques et de syntaxe. De même, le changement de registre de langages implique que certaines règles peuvent être en usage dans certains registres et pas dans d'autres.

La deuxième approche probabiliste repose sur l'apprentissage des modèles de langages (au niveau mot et phrase) à partir de corpus avec une bonne couverture lexicale et syntaxique. Les tous premiers systèmes d'assistance à l'écriture (Hunnicut 1985) se contentaient de calculer la fréquence des mots. La plupart des systèmes actuels utilisent des n-grams (fréquence de suite de n lettres ou de n mots) comme dans le module SybiLettre du projet Sybille (Schadle 2003), dans le système de (Trnka et McCoy 2007), ou plus récemment dans le projet Eugénio (Garcia et al. 2014). Le système Dasher (Ward 2001) utilise la méthode de Monte-Carlo. La limitation de cette approche statistique est qu'elle repose entièrement sur la représentativité et la couverture lexicale des corpus : si le corpus n'est pas assez représentatif du langage de l'utilisateur, l'apprentissage du modèle sera insuffisant et par conséquent les performances du système seront moindres (Trnka et McCoy 2007). Des études ont été réalisées pour essayer de déterminer le corpus optimal pour un système de Communication Assistée et Augmentée (CAA). Les études de Trnka (Trnka et McCoy 2007) ont démontré que pour construire un modèle n-gram performant, il fallait disposer d'une *grande* quantité de textes généraux, d'une quantité *raisonnable* de textes du même style et sur le même sujet, et d'une *petite* quantité de textes similaires à ce que va écrire l'utilisateur. Nous ne partageons pas cette opinion. Nous pensons qu'il faut un corpus parfaitement adapté au registre de l'utilisateur (Boissière et al. 2012). Plus le taux de mots hors vocabulaire est faible et plus le lexique est petit, meilleur est le rendement du système. Il ne sert à rien de mettre dans la base de connaissances des mots que l'utilisateur n'utilisera jamais. L'étude de (Kristensson et Vertanen 2012) rapporte que le corpus de MacKenzie et Soukoreff (MacKenzie et Soukoreff 2003) est efficace pour l'évaluation des systèmes de prédiction.

L'approche mixte consiste à combiner les avantages des deux premières approches tout en minimisant leurs inconvénients respectifs. On conserve la problématique de l'apprentissage des modèles statistiques sur des corpus les plus "*représentatifs*" possibles, tout en rajoutant des informations linguistiques, comme des attributs syntaxiques qui ne figurent pas dans l'approche purement statistique. On pondère ensuite ces informations linguistiques pour déterminer, en fonction des attributs syntaxiques attachés aux mots précédents, le ou les attributs les plus probables du mot qui suit (système HandiAs (Le Pévédic 1997), (Maurel et al. 2001) et FASTY (Wester 2003)). Les informations linguistiques ne se réduisent pas aux seules informations syntaxiques ; les informations de niveau sémantiques commencent à entrer en ligne de compte. Ainsi, le système SybiSem (Wandmacher et al. 2007 a) a été créé pour améliorer les performances de Sybille en introduisant des informations sémantiques par la méthode de l'Analyse Sémantique Latente (ASL). Le principe est issu de la recherche documentaire : à un document, on associe un vecteur de mots clés. Pour l'ASL, le document est remplacé par un mot auquel est associé un vecteur de mots clés. Après une décomposition en valeurs singulières, l'espace se réduit à 200 ou 300 dimensions. On prend ensuite les 100 derniers mots écrits, (qui constituent le champ sémantique du discours), et on projette ce champ sémantique sur le vecteur d'un mot pour trouver les mots les plus probables. Le plus gros problème est de pouvoir extraire les informations sémantiques à partir d'un corpus.

Nous proposons de catégoriser les utilisateurs des systèmes d'aide à la saisie en trois populations :

- les personnes handicapées physiques (atteintes des membres supérieurs) qui les utilisent pour accélérer leurs saisies de textes et réduire leur fatigabilité ;
- les personnes qui ont des troubles linguistiques (dyslexiques, dysorthographiques), qui les utilisent pour trouver la bonne orthographe ;
- les personnes en situation de mobilité qui se servent de téléphones portables, smartphones et autres tablettes pour écrire des textos et des Short Message Service (SMS) avec des abréviations plus ou moins codifiées.

Il va de soi que ces trois catégories d'utilisateurs n'ont pas toutes les mêmes besoins en termes d'assistance à l'entrée de textes. Comment alors un moteur linguistique peut-il répondre à des objectifs aussi divers voire parfois antagonistes — Trouver la bonne orthographe versus écrire en SMS (Beaufort et al. 2010) ? Des logiciels conçus pour des personnes

dyslexiques existent comme WordQ² et Médialexie³, mais sont-ils vraiment utiles et efficaces pour cette population pour laquelle ils ont été conçus?

Faut-il créer un moteur générique, capable de faire une prédiction de mots relativement performante, sur lequel viendraient se greffer des modules satellites qui adapteraient les prédictions aux usages des différentes populations ou faut-il créer dynamiquement son moteur de prédiction en fonction de sa saisie comme le démontre (Sakkos al. 2015).

3.2 Compromis entre performance du système d'assistance et charge cognitive du scripteur

L'autre challenge dans les systèmes de saisie de texte réside dans la recherche d'un compromis entre la performance du système de prédiction et la minimisation des attentions visuelles et cognitives. La performance d'un système d'aide à la saisie de texte se mesure avec le KSR (Keystroke Saving Rate) (Zagler et Seisenbacher 2000). Le KSR est dépendant du nombre de mots dans la liste de prédiction: plus la liste de mots prédits est longue, plus le KSR augmente et par conséquent, plus le temps de recherche d'un mot dans la liste augmente (Pouplin et al 2015). La problématique réside donc dans la recherche d'un compromis entre la maximisation du KSR et la minimisation de la charge visuelle et cognitive de recherche de la bonne prédiction : le bénéfice apporté par de bonne performance peut être annulée par l'augmentation des charges.

Trop peu de recherches se sont intéressées à l'ergonomie des listes de prédiction. (Wester 2003) rapporte les résultats d'un questionnaire sur l'utilisabilité du système FASTY : 3 utilisateurs handicapés sur 6 étaient d'accord avec les paramètres de présentation de la liste de prédiction (position de la liste dans l'écran, taille, fonte, couleur, police des caractères de la liste de mots) alors que les trois autres étaient en total désaccord. L'étude de (Pouech et al. 2009) a montré que les personnes "valides" étaient moins attentives —non perception du mot correct— aux prédictions affichées par la liste de mots que les personnes avec un handicap moteur. Les résultats trop partiels de ces deux études nous suggèrent que des études complémentaires, intégrant des systèmes de suivi de regard pour comprendre le choix de la prédiction doivent être conduites sur des populations plus élargies.

L'enquête récente de (Poulain et al. 2015) confirme que le paramètre de configuration le plus important dans les systèmes de prédiction selon les professionnels de la rééducation (près 50 % d'entre eux) est la gestion de la liste de mots. Les autres paramètres des systèmes d'assistance à la saisie sont souvent utilisés avec les valeurs par défaut, peu adaptés à l'utilisateur. Les mots les plus probables devraient être mis en saillance.

Dans les systèmes de complétion, cette question de la surcharge cognitive n'existe plus puisqu'il n'y a plus de listes de prédiction. (Dubus 1996) a comparé le système de complétion VITIPI avec le système de prédiction HandiWORD dans un Centre Spécialisé d'Enseignement Secondaire. Les élèves qui appréciaient le plus VITIPI étaient ceux qui maîtrisaient le mieux le Français et son orthographe. Pour les plus jeunes (jusqu'à la 4ème), la présence d'une liste de prédiction les rassurait sur l'orthographe des mots. Dans une étude, (Anson et al. 05) a montré que pour des utilisateurs exempts de troubles physiques : 1) la saisie avec un clavier virtuel et sans complétion de mot est plus lente qu'avec un système de complétion (en faisant la moyenne des 10 participants, on obtient une moyenne de 9.79 mots par minute pour la complétion contre 10.03 pour la prédiction) ; 2), la saisie de textes avec un système de complétion est globalement plus lente qu'avec un système de prédiction de mots. (Boissière et al. 2006) ont observé que le KSR d'un système de complétion est inférieur à un KSR d'un système de complétion lors de comparaison du système de prédiction Sybille (Schadle 2003) et du système VITIPI (Boissière et Dours, 1996).

Conclusion

Dans leur étude avec une population de tétraplégiques (Pouplin et al. 2014) ont démontré que la vitesse de saisie par un clavier dynamique est réduite de 37 % en comparaison avec un clavier standard et que l'ajout d'un système de prédiction n'a aucun effet sur la vitesse. Cela pose la question de savoir pourquoi nous n'arrivons toujours pas à nous habituer à d'autres agencements que ceux de AZERTY ou QWERTY ?

Les résultats des études respectivement de (Raynal 2005), (Pouech et al. 2009) et (Pouplin et al. 2014) démontrent clairement la nécessité de concevoir des claviers dynamiques plus ergonomiques et de définir le nombre et la position des mots prédits. L'évaluation des systèmes de prédiction ne pourra se faire que par l'introduction de systèmes de suivi du regard pour mesurer les stratégies visuelles mises en place dans le processus de sélection du mot. Les performances

² <http://www.cleomonde.com/collections/wordq-et-speakq>

³ <http://www.medialexie.com/?langue=fr>

moindres en termes de vitesse de saisie mais également en KSR posent la question de l'utilisation des systèmes de complétion versus des systèmes de prédiction de mots ? Faut-il un système performant versus un système moins fatigant ?

Références

- GARAY-VITORIA N., ABASCAL J. (2006) Text prediction systems: a survey. *Univ Access Inf Soc* (2006) 4: 188–203 DOI 10.1007/s10209-005-0005-9.
- ANTOINE J-Y, MAUREL D. (2007) Aide à la communication pour personnes handicapées et prédiction de texte Problématique, état des lieux et retour sur trente ans de recherche en communication augmentée. *Revue TAL*. Volume 48 – n° 2, 9 - 46.
- MARTIN, B., PECCI, I. (2007). État de l'art des claviers physiques et logiciels pour la saisie de texte. *Revue d'Interaction Homme-Machine*, vol. 8, n°2, 147-205.
- ZHAI, S., HUNTER, M., SMITH, B.A. (2000). The Metropolis Keyboard - an exploration of quantitative techniques for virtual keyboard design. In *Proceedings of The 13th Annual ACM Symposium on User Interface Software and Technology (UIST)*. SAN DIEGO, CALIFORNIA: ACM., 119-218.
- HARBUSCH K., HASAN S., HOFFMANN H., KÜHN M., SCHÜER B. (2003). Topic and Author-Specific Suggestion Lists for Typing with Ambiguous Keyboards. *Proceedings of the Recent Advances in Natural Language Processing (RANLP 2003)*, Borovets, Bulgaria.
- VIGOUROUX N., VELLA F., TRUILLET PH., RAYNAL M. (2004). Evaluation of AAC for text input by two populations: normal versus handicapped motor persons. *8th ERCIM UI4All, Adjunct Proceedings*, Vienna (Österreich).
- GUERRIER Y., BAAS M., KOLSKI CH., POIRIER F. (2011). Étude comparative entre un clavier virtuel AZERTY et un clavier multitap pour des utilisateurs souffrant d'une Infirmité Motrice Cérébrale de type tétraplégique athétosique. *Conférence Internationale sur l'accessibilité et les systèmes de suppléance aux personnes en situations de handicaps (ASSISTH 2011)*. Paris, Vigouroux, Vella, Lepicard (Eds.), IRIT Press, 164-177.
- BAAS M., GUERRIER Y., KOLSKI C., POIRIER F. (2010). Système de saisie de texte visant à réduire l'effort des utilisateurs à handicap moteur. *Conférence Ergo IA 2010*.
- GETSCHOW C. O., ROSEN M. J. (1986). Goodenough-Trepagnier: A systematic approach to design a minimum distance alphabetical keyboard. *RESNA (Rehabilitation Engineering Society of North America) 9th Annual Conference*. 396-398.
- ZHAI S., HUNTER M., SMITH B.A. (2000) The Metropolis Keyboard - an exploration of quantitative techniques for virtual keyboard design. *Proceedings of The 13th Annual ACM Symposium on User Interface Software and Technology (UIST)*, San Diego, California, 119-218.
- MACKENZIE I. S., ZHANG S. Z. (1999) The design and evaluation of a high performance soft keyboard. *Proceedings of the ACM Conference on Human Factors in Computing Systems - CHI '99*, New York, 25-31.
- DE CALMES M., PERENNOU G. (1998). BDLEX : a Lexicon for Spoken and Written French. *1st International Conference on Language Resources & Evaluation*. Grenade, A. Rubio, N. Gallardo, R. Castro, A. Tejada (Eds.), ELRA, Paris, 1129-1136.
- RAYNAL M., VIGOUROUX N. (2005). Genetic Algorithm to Generate Optimized Soft Keyboard. *1st Int. Conf. for Human-Computer Interaction (CHI 2005)*, ACM Press, (support électronique), Portland, Oregon, USA.
- VIGOUROUX N., BOISSIERE PH., VELLA F., BOURAOUI J-L., MOJAHID M. (2015). Information communication technology and (psycho) linguistic sciences : a nice alliance for the design of alternative and augmentative communication for speech and motor impairment, *Routledge*.
- WOBBROCK J.O., MYERS B.A. (2006). Trackball text entry for people with motor impairments. *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI '06)*. Montréal, Québec, 479-488.

- WARD D. J., BLACKWELL A. F., MACKAY D. J. C. (2000). Dasher - a data entry interface using continuous gestures and language models. *UIST 2000*, 129-137.
- POIRIER F., ANTOINE J.-Y., BARHOUMI Z. (2007). Sibylle : améliorations de l'interface à partir de critères ergonomiques et des besoins des utilisateurs. *1ère Conférence Internationale Sur l'accessibilité et les systèmes de suppléance aux personnes en situations de handicaps (ASSISTH'2007)* "Pour une meilleure insertion dans la société", Toulouse, France, Ed Cépadués, 241-247.
- AL FARAJ K., MOJAHID M., VIGOUROUX N. (2008). GrooveWrite: A Multi-Purpose Stylus-Based Text Entry Method. *International Conference on Computers Helping People with Special Needs (ICCHP 2008)*, Linz, Austria.
- GO, K., AND ENDO, Y. (2007). CATKey: customizable and adaptable touchscreen keyboard with bubble cursor-like visual feedback. *Proc. HCII*, 493-496
- GO, K., & TSURUMI, L. (2010). Arranging touch screen software keyboard split-keys based on contact surface. In E. Mynatt, D. Schoner, G. Fitzpatrick, S. Hudson, K. Edwards, & T. Rodden (Eds.), *Proceedings of the 28th international conference on human factors in computing systems*. New York, 3805-3810.
- BOISSIÈRE PH, VIGOUROUX N, MOJAHID M, VELLA F. (2012) Adaptation of AAC to the context communication: a real improvement for the user Illustration through the VITIPI word completion, In Miesenberger, K., Karshmer A., Penaz P. Zagler, W: (Eds) *Computers Helping People with Special Needs, Proceedings of the 13th International Conference, ICCHP'2012, Linz, Austria, July 2012. Lecture Notes in Computer Science (LNCS 7383, Part II)*, 451-458.
- GODBERT E, PASERO R, SABATIER P. (1992) Composition de phrases assistée : Interaction des connaissances linguistiques et conceptuelles. *4ème journée sur l'ingénierie des interfaces homme-machine IHM'92*, 197-202.
- MILHAUD G, PASERO R, SABATIER P. (1992) Partial synthesis of sentences by coroutining Constraints on different levels of well-formedness. *Proceedings of COLING*, Nantes, 926-929.
- PASERO R, SABATIER P. (1998a) Linguistic games for language Learning and tests, an ILLICO application. *Computer-Assisted Language Learning (CALL)*.
- PASERO R, SABATIER P. (1998b) Concurrent Processing for Sentences Analysis, Synthesis and Guided Composition. *Natural Language Understanding and computational Logic, Lecture Notes in Computer Science*, Springer.
- BOISSIERE PH, DOURS D. (2001) VITIPI : Comment un système d'assistance à l'écriture pour les personnes handicapées peut offrir des propriétés intéressantes pour le TALN ? *TALN 2001 Tours*, 183-192.
- PERENNOU G, DAUBEZE P. LAHENS F. (1986) La vérification et la correction automatique de textes : le système VORTEX. *T.S.I Vol 5 n°4*, 285-305.
- BOISSIÈRE PH., DOURS D. (1996) VITIPI : Versatile Interpretation of Text Input by Persons with Impairments. In 5th *ICCHP (International Conference on Computers for Handicapped Persons)*. Linz, 165-172.
- HUNNICUT S. (1986) A lexical prediction for a text-to-speech system. *Rapport of department of speech communication Stockholm STL-QSPR 2-1/1985*.
- SCHADLE I. (2003) Sibylle : Système linguistique d'aide à la communication pour les personnes handicapées. *Thèse d'informatique, Laboratoire VALORIA*.
- TRNKA K, MCCOY K. (2007) Corpus Studies in Word Prediction. *Proceedings of ASSETS'07, Tempe, Arizona, (USA)*, 195-202.
- GARCIA L. DE OLIVEIRA L. AND DE MATOS D. (2014) Word and Sentence Prediction: Using the best of the two worlds to assist AAC users. *Technology and disability, Volume 26, Numbers 2,3*, 79-92.
- WARD D.J. (2001) Adaptive Computer Interfaces. *Thèse de Doctorat de Philosophie Université de Cambridge*.
- KRISTENSSON P.O. VERTANEN K. (2012) Performance Comparisons of Phrase Sets and Presentation Styles for Text Entry Evaluations. *Proceedings of IUI'12, Lisbon, (Portugal)*, 29-32.

MACKENZIE, I. S. SOUKOREFF, R. W. (2003) Phrase sets for evaluating text entry techniques. In *Ext. Abstracts CHI 2003*, ACM Press, 754-755.

LE PEVEDIC B. (1997) Prédiction Morphosyntaxique évolutive dans un système d'aide à la saisie de textes pour des personnes handicapées physiques. *Thèse de Doctorat IRIN* (N° ED-82-269)

MAUREL D. ROSSI N. THIBAUT R. (2001) Handias : un système multilingue pour l'aide à la communication de personnes handicapées. *Atelier Thématique TALN 2001, Tours*, 203-212.

WESTER M. (2003) User Evaluation of a Word Prediction System. *Thèse, Université Uppsala (Suède)*.

WANDMACHER T. BECHET N. BAEHOUMI Z. POIRIER F. ANTOINE J-Y. (2007) Système Sibylle d'aide à la communication pour personnes handicapées : modèle linguistique et interface utilisateur. *Atelier Thématique TALN 2007*, Volume 2, 539-548.

BEAUFORT R, ROEKHAUT S, COUGNON L-A, FAIRON C. (2010) [Une approche hybride traduction/correction pour la normalisation des SMS](#). Actes de la 17e conférence sur le traitement automatique des langues naturelles (TALN'10), actes électroniques, Montréal, Canada.

POUECH N., BOISSIERE PH., VELLA F., VIGOUROUX N. (2009) Influence des modes de présentation et de sélection des listes de prédiction : Étude sur trois systèmes d'aide à la saisie de textes. Dans *5^{ème} Journée d'Ergothérapie de Garches (ARFEHGA)*, Garches, 84-86

ZAGLER W.L. SEISENBACHER G. (2000) German language predictive typing : Results from a feasibility investigation. *Proceedings of the ICCHP'2000, Karlsruhe*, 771-779

POUPLIN S. ROCHE N. HUGERON C. VAUGIER I. BENSMAIL D. (2015) Recommendations and settings of word prediction software by health-related professionals for people with spinal cord injury: a prospective observational study. *European journal of physical and rehabilitation medicine (sous presse)*.

BOISSIÈRE PH, SCHADLE I, ANTOINE J-Y. (2006) A methodological framework for writing assistance systems: applications to Sibylle and VITIPI systems, In *Modelling, Measurement & Control, Série C, (bioengineering)*. Edited by Association for the Advancement of Modelling & Simulation Techniques in Enterprise, AMSE-journals (Barcelona Spain), Modelling C, Vol 67 (Supp Handicap 2006), 167-176.

DUBUS N (1996) Evaluation de l'interface intelligente d'aide à la saisie informatique, VITIPI au lycée "Le Parc Saint-Agne". *Journal d'Ergothérapie Mars 1996 MASSON*, 95-100

ANSON, D. K., MOIST, P., PRZYWARA, M., WELLS, H., SAYLOR, H., AND MAXINE, H. (2005). The effects of word completion and word prediction on typing rates using on-screen keyboards. In *Proceedings of RESNA '05*. RESNA, Arlington, VA, no pages, on proceedings CD.

POUPLIN S., ROBERTSON J., ANTOINE J-Y., BLANCHET A., KAHLOUN J-L., VOLLE PH., BOUTEILLE J., LOFASO F., BENSMAIL D. (2014) Effect of dynamic keyboard and word-prediction systems on text input speed in persons with functional tetraplegia, *Journal of Rehabilitation Research & Development (JRRD)*, Volume 51 Number 3, 467-480