

Une aide à la communication par pictogrammes avec prédiction sémantique

Aurélie Merlo^{1,2}

(1) Savoirs, Textes, Langage (STL), rue du Barreau, 59653 Villeneuve d'Ascq Cedex

(2) Ergonotics SAS, 165 avenue de Bretagne, 59000 Lille, France

aurelie@ergonotics.com

Résumé. Cette démonstration présente une application mobile (pour tablette et smartphone) pour des personnes souffrant de troubles du langage et/ou de la parole permettant de générer des phrases à partir de la combinaison de pictogrammes puis de verbaliser le texte généré en Text-To-Speech (TTS). La principale critique adressée par les patients utilisant les solutions existantes est le temps de composition trop long d'une phrase. Cette limite ne permet pas ou très difficilement d'utiliser les solutions actuelles en condition dialogique. Pour pallier cela, nous avons développé un moteur de génération de texte avec prédiction sémantique ne proposant à l'utilisateur que les pictogrammes pertinents au regard de la saisie en cours (e.g. après le pictogramme [manger], l'application propose les pictogrammes [pomme] ou encore [viande] correspondant à des concepts comestibles). Nous avons ainsi multiplié de 5 à 10 la vitesse de composition d'une phrase par rapport aux solutions existantes.

Abstract.

An Augmentative and Alternative Communication with semantic prediction.

This demo shows a mobile app (for smartphone and tablet) for people with language and/or speech disorders for generating sentences from the combination of icons and verbalize text generated by Text-To-Speech (TTS). The main criticism expressed by patients using existing solutions is the time too long required to compose a sentence. Hence, existing solutions are hard to use in dialogic conditions. To alleviate this problem, we have developed a text generation engine with semantic prediction. This engine only proposes to use the relevant pictograms in view of the current entry (e.g. after the pictogram [eat], the application offers pictograms [apple] or [meat] corresponding to edible concepts). We have multiplied from 5 to 10 the speed of composing a sentence compared to existing solutions.

Mots-clés : génération de texte, prédiction sémantique, frame semantics, handicap, aide à la communication

Keywords: natural language generation, semantic prediction, frame semantics, disability, AAC

1. Introduction

Selon l'American Speech-Language Hearing Association, une personne sur 100 dans le monde est concernée par des troubles du langage et/ou de la parole. Ces troubles peuvent faire suite à des causes diverses telles que des maladies neuro-dégénératives (e.g. Sclérose Latérale Amyotrophique), des accidents (e.g. accidents cardio-vasculaires) ou encore des opérations chirurgicales lourdes de la sphère ORL (e.g. laryngectomie, trachéotomie, glossectomie).

La communication parlée étant l'interface majeure des interactions humaines, les troubles du langage et/ou de la parole entraînent un isolement social et une perte de l'autonomie des personnes touchées. Des solutions techniques existent pour pallier ces troubles ; ce sont des aides à la communication ou Augmentative and Alternative Communication (AAC) selon la terminologie anglo-saxonne. Ces aides permettent aux patients, à l'aide de saisies textuelles ou pictogrammiques, de composer des phrases et de les verbaliser en sortie à l'aide d'un système de Text-To-Speech (TTS). Néanmoins, les solutions existantes de suppléance ne sont actuellement pas considérées comme satisfaisantes par les patients eux-mêmes. La principale critique émise est le temps de composition d'une phrase beaucoup trop long pour une utilisation dans un contexte dialogique naturel. En effet, les solutions existantes ne permettent de générer que 5 à 10 mots par minute contrairement à la vitesse de la parole non-altérée estimée à environ 200 mots par minute. Seuls alors la famille, les amis et le corps médical sont assez patients pour interagir avec le patient.

Cette démonstration présente une aide à la communication par pictogrammes pour support mobile (smartphone et

tablette). Nous avons augmenté sensiblement la vitesse de composition d'une phrase à l'aide de pictogrammes (de 5 à 10 fois par rapport aux solutions existantes) grâce à la mise en place de solutions innovantes telles qu'un moteur de prédiction sémantique. L'objectif principal de ce moteur est de ne prédire uniquement que les pictogrammes pertinents au regard de la saisie de l'utilisateur pour générer en sortie des phrases pertinentes du point de vue du sens et de la syntaxe (ex : après le pictogramme [manger]¹, le moteur prédit uniquement des pictogrammes comme [pomme], [viande] ou encore [chocolat] représentant des concepts comestibles).

Nous proposons ici de décrire la prédiction sémantique en place dans l'application. Nous avons développé une ontologie d'affichage des pictogrammes et un module de gestion de la surface pour rendre les phrases grammaticalement et orthographiquement correctes selon la langue. Ces éléments seront présentés en démonstration.

2. Présentation générale de l'application

L'application est une aide à la communication par pictogrammes à destination de personnes ayant des troubles du langage et/ou de la parole avec des troubles cognitifs et gestuels légers. Contrairement aux solutions existantes dont les pictogrammes représentent des mots-formes, elle repose sur le principe qu'un pictogramme représente un concept, soit une information sur le monde qui peut être définie par l'association avec d'autres informations. Ainsi, le concept AVOCAT² représenté par un fruit se définit par ses propriétés comme /comestible/, /avec un noyau/, /peut être vert/ ou encore /avec une peau/. Ces propriétés distinguent ce concept du concept AVOCAT représenté par un humain et possédant des propriétés différentes. Le concept AVOCAT fruit peut être mis en relation avec d'autres concepts selon ses propriétés comme les concepts VERT, MANGER ou encore ÉPLUCHER. Lorsque l'utilisateur sélectionne un pictogramme, il sélectionne un concept et génère une forme fléchie. Par exemple, la sélection du concept MANGER, génère la forme fléchie *je mange*³. Il peut à tout moment modifier cette forme fléchie ou cette forme peut être modifiée automatiquement selon des règles de surface (e.g. accords, conjugaison).

Les pictogrammes sont présentés sur une interface façon mind map (cf. Figure 1 ci-dessous). L'interface est composée de 6 sections : les actions, les circonstanciels (temps, lieux, instruments...), les objets concrets, les objets abstraits, les êtres vivants et les favoris. Les sections des adjectifs et des adverbes de manière apparaissent uniquement à la sélection d'un nom ou d'un verbe. Chaque section contient des catégories qui peuvent contenir elles-mêmes des catégories ou des concepts. Le concept sélectionné par l'utilisateur (e.g. MANGER sur l'interface ci-dessous) est placé au centre autour duquel s'organise la prédiction sémantique des concepts pertinents (e.g. des concepts comestibles, des lieux, des temps).



FIGURE 1 : Interface de l'application

3. Prédiction sémantique et Frame Semantics

La prédiction sémantique des concepts permet de ne proposer que les concepts pertinents au regard de la saisie en cours de l'utilisateur. Pour cela, nous avons créé une grammaire par frames dans la lignée de la Frame Semantics (Fillmore

¹ Par convention, nous mentionnons les pictogrammes entre crochets.

² Par convention, nous mentionnons les concepts en petites majuscules.

³ Le pronom personnel de la première personne du singulier est généré par défaut pour les verbes.

1968, 1982). La Frame Semantics est une théorie initiée par Fillmore à la fin des années 1960. C'est une sémantique empirique mettant en avant la continuité entre langue et expérience du monde. Une frame correspond à une situation prototypique évoquée par des lexèmes (e.g. *manger*, *dévor*er, *grignoter*). Chaque frame peut être décrite par une définition et un typage sémantique des participants de la situation ou *frame elements* (cf. Table 1 ci-dessous).

Frame	Type sémantique du sujet	Verbes du frame	Types sémantiques des objets	Exemple de phrases générées
manger	humain	manger dévorer grignoter	aliment instrument pour manger accompagnant lieu temps (durée, date et fréquence)	<i>Je mange une pomme</i> <i>Je mange avec une fourchette</i> <i>Je mange avec Paul</i> <i>Je mange dans la cuisine</i> <i>Je mange à 14h</i> <i>Je mange</i> <i>Je mange une pomme dans la cuisine</i> <i>Je mange avec Paul dans la cuisine à 14h</i>

TABLE 1 : Exemple du frame « manger »

Notre lexique contient 684 verbes correspondant à environ 250 frames. Les 5800 concepts de notre lexique sont annotés en types sémantiques appelés par les frames (e.g. les concepts POMME, AVOCAT, VIANDE, CHOCOLAT correspondent au type « aliment »). Un concept peut avoir plusieurs types sémantiques selon que la frame qui l'appelle est spécifique ou générique. Par exemple, la frame « manger » appelle le type sémantique « instrument pour manger » du concept FOURCHETTE contrairement à la frame « acheter » plus générique qui appelle le type sémantique « artefact » de FOURCHETTE.

4. Niveaux de prédiction sémantique

La prédiction sémantique à base de frames peut se faire à plusieurs niveaux dans la saisie. Nous venons de voir qu'elle peut se faire sur la relation entre le sujet et le verbe, entre le verbe et ses objets mais également sur les propriétés du concept comme la relation entre le nom et ses adjectifs (e.g. le frame « aliment » a pour type sémantique « couleur ») et entre le verbe et ses adverbes (e.g. le frame « manger » a pour frame element « vitesse »). Un système de restrictions sémantiques vient affiner la prédiction des concepts associés aux propriétés (e.g. le type sémantique « couleur » pour AVOCAT est restreint au concept VERT et non à toutes les couleurs). Nous sommes également capable d'identifier la gradabilité de certains concepts et de proposer ainsi à l'utilisateur des adverbes d'intensité quand cela est pertinent (e.g. prédictions des concepts : POMME ROUGE TRÈS, *HOMME MORT TRÈS, ROULER VITE TRÈS). Enfin, nous prenons en considération le caractère massif ou comptable des concepts pour la prédiction des numéraux (e.g. prédictions des concepts : POMME TROIS, *PEUR TROIS).

Références

FILLMORE C. (1968). "The case for case", In: E. Bach and R.T. Harms (eds) (1968) *Universals in Linguistic Theory*. London: Holt, Rinehart and Winston, 1–88.

FILLMORE C.(1982). "Frame Semantics", In H. P. CO, Ed., *Linguistics in the morning calm*, 111–137.