

# Un Système de Dialogue Vocal pour les Seniors : Études et Spécifications

Pierrick Milhorat, Stephan Schlögl, Gérard Chollet  
Télécom-ParisTech, Institut Mines-Télécom  
37-39 rue Dareau, Paris  
Email : name@telecom-paristech.fr

Jérôme Boudy  
Télécom SudParis, Institut Mines-Télécom  
9 rue Charles Fourier  
Email : jerome.boudy@it-sudparis.eu

**Résumé**—Les solutions de services dématérialisés venant en aide aux tranches d’âge supérieures de la population européenne sont particulièrement encouragées du fait de du vieillissement rapide du continent. Cependant, il existe toujours un frein dans le déploiement de ceux-ci : les interfaces sont peu ou pas adaptées aux utilisateurs cibles. Le projet vAssist propose une plate-forme de services pour les seniors accessible via un système de dialogue vocal. Cet article présente les étapes de conception et de développement du système

## I. INTRODUCTION

En 2010, la population européenne comptait 87 millions d’individus âgés de 65 ou plus, soit 17,38% de la population totale. Les estimations actuelles prévoient que le taux de personnes en âge de travailler (15-64 ans) par personne de plus de 64 ans sera de moins de deux pour un en 2060 [1]. Cet accroissement rapide a pour conséquence une réforme nécessaire de nos systèmes de retraite et encourage le développement de solutions de soutien pour les tranches d’âge supérieures de notre société. Bien que les progrès de la science nous conduisent à vivre et vieillir plus confortablement et en meilleure santé, nous rencontrerons irrémédiablement des troubles physiques et/ou cognitifs à plus ou moins long terme. Le programme européen Ambient Assisted Living (AAL) propose des solutions pour remédier à cela, notamment, et c’est l’aspect prévalent de cet article, pour parer aux insuffisances de la motricité fine et aux maladies chroniques. Les statistiques européennes montrent que aujourd’hui, 1,2 million des seniors souffrent de la maladie de Parkinson et 630 000 sont atteints de sclérose. De fait, les solutions de soutien pour cette tranche de la population du continent devraient comprendre un support croissant pour la télé-santé mais également proposer des interfaces adaptées pour compenser les restrictions dues à l’âge.

Depuis les sorties de Siri (Apple) et Voice Search (Google), le marché des interfaces vocales s’étend rapidement. Les applications supportent de plus en plus l’interaction vocale, la présentant parfois comme l’unique option pour interagir. Le contrôle par la voix est maintenant actif à bord de nos voitures, dans nos foyers, et les médias nous font envisager un futur dans lequel nous parlerons aux machines comme à des entités humaines. En revanche, les méthodes et outils pour créer des Systèmes de Dialogue Vocaux (SDV) et les adapter à leurs utilisateurs cibles sont rare. L’ensemble des services connectés auxquels la plate-forme vAssist fournira un accès seront développés selon une approche centrée sur les utilisateurs potentiels et le marché potentiel, ainsi, ils

seront conjointement développés en phase avec les demandes et attentes de ces deux pôles d’influence. vAssist ne produit pas de nouveaux services mais se concentre sur l’ajout de modules au plate-forme, standards et applications déjà disponible pour y inclure une apparence vocale intelligente. Les travaux déjà disponibles (universalAAL) sont pris en compte dans la conception technique et les stratégies d’exploitation du projet.

Cette article introduit le projet vAssist, résume les études préliminaires effectuées et leurs conclusions puis décrit le système de dialogue vocal développé pour répondre aux besoins des seniors.

## II. LE PROJET VASSIST

Le projet vAssist a pour but de concevoir et d’implémenter une plate-forme alliant des services de communication et de soins à domicile accessibles par l’intermédiaire d’un dialogue vocal à destination des personnes âgées présentant des restrictions de la motricité fine et/ou sujets à des maladies chroniques. L’interface développée doit être efficace et adapté pour une utilisation par les seniors. Bien que la modalité prédominante soit l’interaction vocale, vAssist offre également une interface graphique alternative pour les situations où cette dernière serait mieux appropriée. Enfin, pour minimiser les coûts d’installation d’un tel système et ainsi favoriser son adoption par une large catégorie de la population, les terminaux déjà existant dans la plupart des foyers européens tels que les PCs, les télévisions, les téléphones mobiles/fixes et les tablettes sont mis à profit.

### A. Les Services

Les services proposés dans vAssist visent à améliorer les applications et logiciels actuellement disponibles par l’ajout d’une interface vocale multilingue sous la forme d’un système de dialogue en langage naturel. Parmi ceux-ci, on compte des outils de communication (appel audio et vidéo, messagerie), des outils de recherche et d’organisation (gestion des contacts, calendrier/agenda, recherche internet) et des applications médicales (carnet de santé virtuel, jeux de stimulation cognitive). Bien que vAssist permette l’accès à différents types services, la plate-forme doit apparaître à l’utilisateur comme une unique entité cohérente : c’est une sorte de majordome comme l’avait imaginé Bush [2] avec “Memex” et dont nous ne réalisons le potentiel que 70 ans plus tard. Il faut néanmoins garder à l’esprit les contraintes d’adaptation imposées par la population cible. Des travaux ont démontré l’influence de

la qualité de la synthèse vocale dans le degré d'acceptation du système [3], vAssist devrait ainsi "s'exprimer" avec une voix claire, articulée et naturelle, produire des requêtes de clarification et de confirmation à intervalles réguliers et utiliser une grammaire et un vocabulaire basiques. Un clavier (virtuel ou physique) sera disponible pour répondre aux exigences d'utilisateurs et de certaines applications. Les sections suivantes décrivent plus en détails les différents services que vAssist offrira.

### *B. Appel Audio/Vidéo*

vAssist permettra la création et la gestion d'une liste de contacts permettant aux utilisateurs d'entrer en contact avec leurs praticiens, leurs proches et leurs amis en quelques mots. Dans le cas où un doublon dans la liste ou une incertitude persistera quant à l'identité du contact à appeler, le système sera en mesure d'engager un dialogue pour clarifier la demande. De plus, l'ensemble des contacts sera organisé en un réseau familial, médical et/ou professionnel de façon à répondre à des requêtes telles que "appelle ma fille" ou bien "envoie ça au docteur". D'un point de vue technique, vAssist supportera des liaisons vers les téléphones fixes et mobiles ainsi que les logiciels de voix sur IP (Skype, Google Talk).

### *C. Email, SMS, MMS*

En plus du service d'appel vocal, un utilisateur de la plate-forme sera en mesure d'envoyer des messages textuels : SMS, MMS, emails. Ceux-ci pourront être appariés avec des événements du calendrier (voir section II-E) comme la prise de médicament ou les anniversaires. Comme précisé précédemment, la priorité sera donnée à la diction vocale des messages bien qu'un clavier d'appoint sera disponible.

### *D. Gestion Des Contacts*

La gestion des contacts est basée sur les logiciels déjà intégrés dans les foyers européens (principalement sur le PC ou le téléphone). vAssist offrira des fonctionnalités d'assistance à l'édition, à la création et à la suppression d'entrées médicales, professionnelles ou privées par l'intermédiaire de dialogues vocaux. De façon à conserver une interaction efficace et cohérente, le système n'exigera que le minimum d'information (nom et prénom par exemple) pour identifier un contact. En revanche, les paramètres complémentaires sont recevables (surnom, numéro de téléphone, adresse, préférences, date d'anniversaire, etc...) quelque soit l'étape courante du dialogue. En réponse à une requête de l'utilisateur, les informations relatives à un contact seront affichées et synthétisées vocalement de façon plus ou moins détaillée selon les préférences de l'utilisateur. Ce dernier pourra ensuite manipuler cette référence selon son désir. Toute source d'ambiguïté sera automatiquement détectée et le système tentera de clarifier les tours d'interaction incertains. vAssist proposera ainsi un puissant moteur de recherche et de maintenance, masqué par une façade voulu simple, efficace et accessible, évitant ainsi à l'utilisateur la lourde tâche de parcourir une longue liste d'entrées plus ou moins complètes.

### *E. Calendrier/Rappels*

La plupart des seniors sont suivis par plusieurs médecins (généralistes ou spécialistes) à plus ou moins long terme. La diversité, le nombre et l'irrégularité des rendez-vous médicaux est source de confusion. Les traitements médicaux bien souvent associés à ce suivi sont également une contrainte constante mais nécessaire pour le bien-être du patient. vAssist propose aux seniors un calendrier éditable accessible n'importe où et n'importe quand qui, pro-activement, aide son propriétaire à organiser son planning quotidien via une combinaison de deux interfaces, l'une vocale et l'autre textuelle/tactile. La fonction de rappel est activable pour tout événement de cette base données et produit des messages sonores et visuels d'alerte sur tout les supports connectés. Les acteurs médicaux pourront également avoir accès à certains types d'information pour contrôler à distance l'évolution de l'état de santé de leurs patients (en plus du carnet de bien-être, voir section II-G). La gestion concurrente de cet agenda, par l'environnement médical, familial ou tout autre entité autorisée facilitera également l'harmonisation des événements entre les différents participants.

### *F. Recherche Internet*

S'ajoutant aux services de communication décrit précédemment, vAssist intégrera un moteur de recherche internet spécialement conçu pour les seniors. Bien que les sujets relatifs à la santé seront favorisés, un utilisateur aura "à portée de voix" des informations sur tous ses sujets de prédilection. Les résultats de la recherche seront affichés et/ou lus selon les préférences de l'utilisateur. Les aptitudes du moteur seront étendues à la couverture des services inclus dans chaque version de la plate-forme vAssist, toujours dans l'objectif de faire de cette plate-forme une entité intégrée dans un système plus large. L'ensemble de ces fonctionnalités sera accessible via des entrées vocales et textuelles.

### *G. Carnet de Bien-être*

Un suivi continu de l'état de santé du patient est un facteur important de l'efficacité des soins médicaux prodigués par l'entourage médical. Les carnets de santé sont largement utilisés pour suivre les individus et les sujets de recherche médicale, il n'est plus besoin de démontrer aujourd'hui combien ils peuvent être bénéfiques pour la prévention, les soins en continus et l'anticipation de possible effets secondaires lors de la prise de médicaments [4]. Pour les cas dans lesquels les praticiens demandent au patient de procéder lui-même un suivi et de reporter les résultats de tests médicaux réguliers (quotidiens), vAssist souhaite faciliter ces tâches en fournissant un canal d'entrée vocale de ceux-ci dans un carnet électronique. Le senior aura la possibilité de "discuter" avec la plate-forme qui le guidera dans l'accomplissement des différentes tâches à effectuer. L'envoi vers les personnes concernées par les informations enregistrées sera automatique. Tout comme pour les événements du calendrier, une fonction de rappel pourra être activée de façon à assurer la régularité de ces auto-tests.

### *H. Jeux Cognitifs*

L'évolution de nos aptitudes cognitives est inversement proportionnelle à celle de notre âge avec pour conséquence

des défaillances de la mémoire et un temps moyen d'analyse qui s'allonge [5]. La stimulation cognitive est une thérapie permettant un ralentissement relatif de cette dégradation. La communauté scientifique reste partagée quant à l'efficacité d'une telle méthode malgré la publication récente de résultats prometteurs [6]. vAssist a fait le choix d'intégrer de tels jeux au sein de l'ensemble des services médicaux et quotidiens. Les jeux seront interactifs, à travers la voix et un clavier (ou un écran tactile). Le niveau de difficulté sera adapté selon les aptitudes de l'utilisateur et les résultats (les scores) obtenus pourront être transmis à un analyste médical qui pourra alors modifier le niveau de chaque jeu et juger de l'évolution du joueur.

### III. ÉTUDE PRÉLIMINAIRE

#### A. Définition Des Utilisateurs Cibles

Pour initier la conception de services orientés vers les seniors, il est nécessaire de comprendre les attentes et les besoins de cette tranche d'âge de la population vis à vis d'un tel système. Pour ce faire, les acteurs/utilisateurs potentiels de vAssist ont été séparés en trois groupes distincts :

- Utilisateurs Primaires (seniors)
- Utilisateurs Secondaires (entourage médical)
- Utilisateurs Tertiaires (fournisseurs de services)

Les utilisateurs primaires de vAssist sont des personnes de 65 ans ou plus qui peuvent présenter des restrictions de l'appareil moteur, de l'ouïe ou de la vue ou des symptômes associés à une perte partielle de la mémoire. L'entourage médical se divise en deux catégories : les membres du corps médical qui ont été formés à la pratique de la médecine et qui sont rétribués pour leur travail. Ce sont les médecins généralistes, les neurologistes, les psychologues, les cardiologues, les gériatologues, les infirmières, les nutritionnistes ou tout autre professionnel dont le métier est de traiter les seniors souffrant de restrictions de la motricité fine et/ou de maladies chroniques. Ils garantissent des soins à domicile ou dans un centre médical. Les proches, utilisateurs secondaires, sont considérés comme des assistants au confort, ils fournissent bénévolement des services de courses, de nettoyage, d'accompagnement, etc... Enfin, les utilisateurs tertiaires, les fournisseurs de services connectés à la plate-forme, sont définis comme des entreprises qui alimentent, entretiennent et anime la celle-ci.

Avec cette définition des catégories d'utilisateurs, nous glissons vers l'étude des attentes de chacune d'elles. Dans cet article, nous nous intéresserons aux seniors en particulier.

#### B. Définitions Des Attentes

Les premières sessions d'analyse des attentes basées exclusivement sur la population des seniors ont fait appel à un mélange d'analyse qualitative et quantitative. Dans l'idée de définir les contours et une compréhension générale du domaine, la réflexion était basée sur les questions suivantes :

- Quelles sont les appareils de télécommunication et de télé-médecine principalement utilisés par les seniors ?
- Quels sont les services pour lesquels une interface vocale serait optimale, en particulier pour les personnes

à motricité fine restreinte et/ou sujets à des maladies chroniques ?

- Quelles sont les préférences et les habitudes des seniors dans l'utilisation d'interfaces vocales ?

Des groupes de travail ont été mis en place pour la collection de données qualitatives. Cette méthode a fait ses preuves dans bon nombre d'études sur l'interaction homme-machine [7]. La communauté scientifique s'accorde à dire que l'effet de groupe est un excellent animateur de la discussion et qu'il permet d'obtenir des résultats tangible et en quantité. Des questionnaires ont également été distribués pour obtenir des données quantitatives sur les relations entre les seniors et la technologie.

#### C. Méthode

Deux groupes de travail ont été organisés au cours desquels les utilisateurs primaires ont pu exprimer leur opinion sur les appareils, les services, les modalités d'interaction et leur vision du futur pour les applications de télé-santé et de télécommunication. Comme introduction, une brève présentation du projet vAssist était donnée aux participants qui avaient signé un formulaire d'accord. L'équipe scientifique était composée d'un animateur et d'un observateur, tous deux participant à la discussion, prenant des notes et reportant certaines informations sur un paperboard. Des questions prédéfinies furent utilisées pour guider la discussion. En plus d'un rapport écrit, chaque session a produit un enregistrement audio et vidéo, ces données ont été interprétées selon un protocole structuré d'analyse du contenu.

De façon à reproduire les fonctionnalités d'un prototype, la méthode du Magicien d'Oz a été employée. Celle-ci permet de tester un système sans pour autant que ce dernier soit complètement développé, autorisant ainsi l'expérimentation de nouveaux paradigmes d'interaction [8]. Un opérateur humain remplace tout ou une partie du système et simule les fonctions manquantes. Les solutions existantes d'interaction vocale ont également été présentées aux utilisateurs, la session ayant pour but de dessiner les contours du système vAssist. Enfin, une dernière partie permettait aux participants de comparer une interface vocale et une entrée au clavier à travers la rédaction d'un email ou d'un SMS via les deux modalités. Ainsi, les utilisateurs inexpérimentés ont pu s'essayer aux technologies actuelles, fournissant de précieuses informations pour la suite du projet.

Pour une analyse plus cohérente, des EmoCards [9] ont été présentées aux participants pour évaluer leur réaction face aux différentes modalités. Ces cartes permettent de définir le degré de satisfaction d'un sujet à travers 9 émotions distinctes parmi lesquelles celui-ci doit choisir (voir figure 1). Par la suite, l'équipe de recherche a demandé à chaque participants d'énumérer les points négatifs et positifs des deux types d'interactions. La section suivante présente et analyse leurs réponses.

#### D. Configuration et Premières Impressions

10 participants ont été conviés au premier groupe de travail (4 hommes, 6 femmes, âge moyen : 70,1 ans), tous avaient au moins une première expérience avec les technologies de l'information et tous présentaient des troubles de la motricité

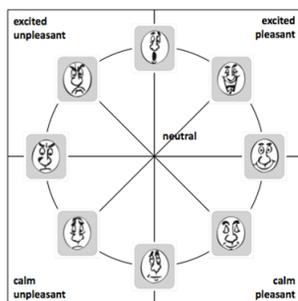


FIGURE 1. EmoCards

fine dûs à l'âge. Le premier sujet de discussion portait sur l'emplacement et le type d'appareil de télécommunication dans leurs foyers, suivi par l'utilisation que chacun avait de ces outils pour communiquer avec leurs proches. Sans surprise, le téléphone (fixe ou mobile) est utilisé pour les contacts locaux tandis que les solutions via internet sont reversées aux appels et messages longue distance. Les appareils sont généralement situés dans la cuisine, le salon ou la chambre et, dans la plupart de cas, il en existe plusieurs sous un même toit. Les participants ont souligné les avantages du rapport qualité-prix des communications textuelles (email, SMS) bien qu'amoindris, dans le cas des SMS, par la difficulté et la pénibilité de l'écriture. Une interface vocale fut globalement perçue comme une très bonne alternative.

Lors du second groupe de travail, auquel participaient 4 hommes et 4 femmes (âge moyen : 70,2 ans) présentant les mêmes conditions que les sujets de la session précédente, Siri, le logiciel d'Apple, a été montré comme exemple et base des débats à venir, suivi d'une expérimentation comparative des deux modalités (texte et parole). L'une des majeure conclusion de cette session porte sur l'importance des tours de clarification et de la personnalisation du système qui devraient être pris en compte dès l'étape de conception de l'interface vocale. D'un autre côté, l'interface graphique souffrirait d'une taille d'écran, d'un clavier et d'une police de caractère trop restreints. L'interaction vocale reste l'interaction qui produit le plus de sentiment de satisfaction pour l'utilisateur.

### E. Opinion des Seniors

Bien que ces sessions d'échange avec les utilisateurs potentiels de la plate-forme vAssist ne nous permettent que de jeter les bases de cette dernière, cela permet de définir quelques éléments essentiels de conception. La taille minimum d'un écran pour l'affichage, par exemple, devrait être celle d'une carte postale standard. Les utilisateurs n'ont pas d'habitude profondément ancrée quant à l'utilisation d'un appareil plutôt qu'un autre et s'accommoderont de ceux disponibles dès lors qu'une documentation suffisamment précise et pédagogique est livrée avec. Pour finir, la taille de la police de caractère, tout comme celle des zones tactiles devra être adaptée aux aptitudes des utilisateurs.

Plus spécifiquement sur l'interaction vocale, les participants ont préféré cette nouvelle modalité d'interaction en lieu et place des biais plus traditionnelles. Les résultats sur les EmoCards montrent que 75% des tâches effectuées à la voix ont un effet plaisant ou neutre, à comparer au taux de 37.5%

pour l'utilisation d'une interface graphique. Les seniors ont tendance à utiliser des phrases courtes et grammaticalement simples pour s'adresser au système mais expriment néanmoins le désir d'employer un langage naturel complexe. Lorsque le système montrait des difficultés pour comprendre l'intention de l'utilisateur, ce dernier souhaitait que l'appareil prenne le contrôle du dialogue pour palier à l'incompréhension ou offrir des alternatives (clavier, souris, zones tactiles). De manière générale, la synthèse vocale féminine est plus appréciée que son équivalent masculin. Nous avons pu constater que le dialogue vocal était facilement accepté et engageant pour l'utilisateur, souligné par le fait que certains participants auraient aimé que la système réponde à un concept d'identité (en ayant un prénom ou un âge par exemple). En conclusion, les seniors sont enclins à utiliser un système de dialogue vocal puisque celui-ci semble plus accessible et adapté à leurs besoins.

## IV. LE SYSTÈME DE DIALOGUE

### A. Architecture Générale

La figure 2 montre la structure générale du système de dialogue vocal que nous développons. Les entrées de l'utilisateur sont dans un premier temps converties en hypothèses sur leur contenu orthographique par le module de Reconnaissance Automatique de la Parole (RAP). Plusieurs procédés successifs sont nécessaires à l'analyse de ces hypothèses pour en extraire la signification et produire une ou plusieurs trames sémantiques. Ces dernières sont des structures de données composées d'un but (intention de l'utilisateur) et de zero ou plusieurs *slots* auxquels une valeur est assignée. A partir des trames sémantiques et suivant le contexte de dialogue dans lequel celles-ci ont été émises, le Gestionnaire du Dialogue (GD) agit sur les applications auxquelles il est connecté et guide l'interaction selon les modèles implémentés. Pour communiquer avec l'utilisateur, le GD produit également des trames sémantiques qui, à travers les deux modules que sont le Générateur de Langage Naturel (GLN) et le Synthétiseur de la Parole (SP), sont converties en signal de parole contenant le message à transmettre. Chacun de ces composants sera décrit précisément dans la suite de cette section.

### B. Reconnaissance Automatique de la Parole

Ce qui fait la spécificité d'un SDV, par comparaison avec les autres types de système de dialogue, est que l'unique modalité d'interaction est la parole. La première tâche d'un programme appartenant à cette catégorie est donc de décoder le message contenu dans un signal de parole. Malgré quelques décennies de recherche sur l'analyse de tels signaux, la conversion du signal en hypothèse textuelles est toujours hautement sujette aux erreurs. L'état de l'art dans le domaine est basé sur l'algorithme de Viterbi qui permet de trouver la séquence d'état cachés la plus probable dans un réseaux de Modèles de Markov Cachés (MMC) générant une séquence d'observations. Les Mel-Frequency Cepstrum Coefficients (MFCC) sont actuellement les descripteurs les plus utilisés pour la représentation paramétrique d'un segment de signal de parole. La distribution des séquences de tels vecteurs de coefficients correspondante à un phonème (la plus petite unité acoustique) est modélisée par un MMC [10]. Les combinaisons de phonèmes pour former les mot composant un langage sont répertoriées dans un dictionnaire de prononciation. Enfin, les contraintes linguistiques

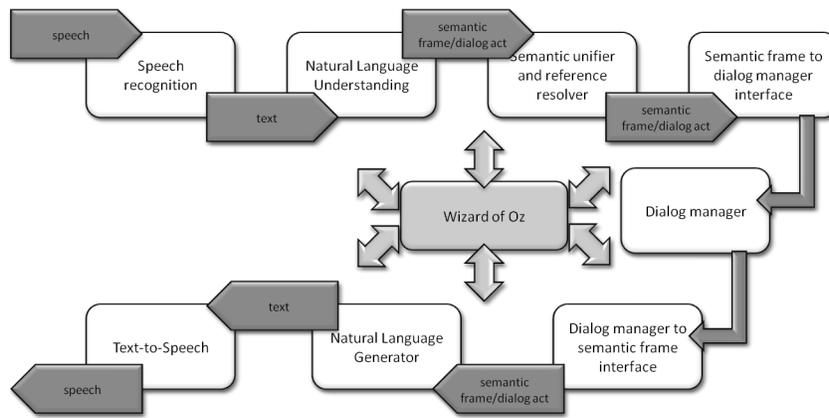


FIGURE 2. Architecture générale

sur la composition des phrases sont contenues dans le modèle de langage, la tendance actuelle pour ces derniers étant basée sur les N-grams. Ces trois sources d'information (modèles acoustiques, dictionnaire et modèle de langage) permettent de délivrer un score pour chaque combinaison de mots et de produire des hypothèses sur la transcription la plus probable du signal de parole.

Nous utilisons le moteur de reconnaissance Julius, développé au Kawahara Lab de l'université de Kyoto [11]. Il est possible, en utilisant notre configuration actuelle, de reconnaître l'anglais, le français, l'espagnol et le hollandais, le travail est en cours pour apprendre les modèles acoustiques pour l'allemand et l'italien. Julius a été notre choix depuis le début du projet mais nous souhaitons tout de même faire remarquer que n'importe quel système de RAP peut se substituer à celui-ci, dès lors que celui-ci produit une ou plusieurs hypothèses de transcription.

### C. Compréhension du Langage Naturel

Bien que la plupart des modules de Compréhension du Langage Naturel (CLN) pour les SDV uni-modaux acceptent une entrée identique, la représentation de la signification d'un message vocal diffère entre eux [12]. En d'autres termes, la structure de données produite dépend des fonctionnalités du système dans son intégralité et/ou du formalisme du GD. Les trames sémantiques ont été sélectionnées pour leur diversité. Ces trames consistent en un but, paramétré par un ensemble de paires (*nom, valeur*), ce sont les *slots*. A la présentation d'une entrée textuelle en sortie de la RAP, le module CLN a pour rôle de convertir celle-ci en trame compréhensible pour un programme informatique. Jurcicek et al. proposent un parser basé sur une succession de transformations selon des règles apprises sur un corpus d'exemples [13]. Ce corpus d'apprentissage peut être collecté, par exemple, par l'intermédiaire d'expériences de Magicien d'Oz.

### D. Unification Sémantique et Résolution des Références

Le module d'Unification Sémantique et de Résolution de Référence (USRR) tire son origine de la nécessité d'établir une correspondance entre l'espace sémantique dans lequel la CLN produit des trames et l'espace sémantique dans lequel le GD est capable de prendre des décisions quant à la suite

du dialogue la plus appropriée. Le coeur du composant est un algorithme de recherche appliqué à une forêt d'arbres de réécriture, implémenté en SWI-Prolog. La figure 3 montre un exemple d'arbre de réécriture. L'algorithme cherche à convertir les *slots* contenus dans une trame sémantique en un ensemble de *slots* à la racine des arbres. Une recherche récursive est effectuée pour obtenir la conversion. Dans le cas où une telle conversion n'est pas possible, l'algorithme produit une erreur.

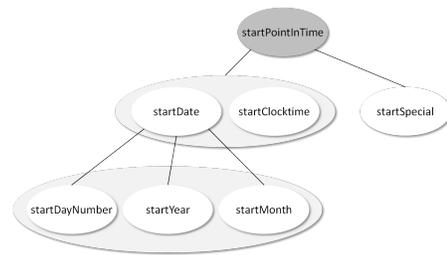


FIGURE 3. Structure d'un arbre inclus dans le procédé d'Unification Sémantique et de Résolution de Référence

### E. Interfaces

Les structures de données en entrée et sortie du GD sont différentes de celles reconnues par les modules de compréhension et de génération du langage (i.e. les trames sémantiques), pour cette raison le coeur décisionnel du système est entouré par deux interfaces qui formatent les trames sémantiques en des actes de dialogues et vice-versa. De plus, l'interface d'entrée, fournit une correspondance dynamique des trames sémantiques vers les actes de dialogue qui dépend de l'état courant de l'interaction.

### F. Gestion du Dialogue

Plusieurs algorithmes stochastiques de gestion du dialogue ont été proposés. Ce paradigme présente actuellement deux principaux défauts : il nécessite une grande quantité de données d'apprentissage et les applications sont limitées dans le nombre de *slots*, de sujets de dialogue et de possibilité d'actions pour le système. Dans notre prototype, nous avons opté pour Disco [14], un GD déterministe dont les modèles de dialogue respectent le standard ANSI/CEA-2018, qui établit une décomposition récursive des tâches en sous-tâche et en

atomes. Devant la difficulté de la conception et de la mise à jour de tels modèles, nous avons consacré une partie de nos travaux à la création de langages informatiques de description qui transfèrent le poids du design de modèles de dialogues vers celui de la description des applications à interfacier. Ces fichiers de description sont ensuite convertis en modèles de dialogue via des transformations XSLT.

### G. Génération du Langage Naturel

La qualité de la GLN est un facteur important pour l'acceptabilité d'un SDV, ce n'est cependant pas notre intérêt de recherche. De fait, la génération est basée sur des templates qui sont filtrés en fonction de l'intention de communication puis aléatoirement sélectionnées pour enfin être instanciées et transmises au module suivant du SDV.

### H. Synthèse de la Parole

OpenMary, un système de SP open-source développé par DFKI, permet, entre autres, de synthétiser le français, l'allemand et l'italien [15].

### I. Magicien d'Oz

L'intégration d'un outil de magicien d'Oz dans notre SDV est l'un des aspects importants de notre contribution. Celui-ci nous permet de collecter des données d'apprentissage et de tester en substituant n'importe quel module du système automatique par un opérateur humain. Jusqu'à présent une première expérience a servi à collecter des données d'apprentissage pour entraîner la CLN.

## V. CONCLUSION

Le projet vAssist souhaite répondre à la problématique actuelle du vieillissement de la population en Europe en proposant des solutions alternatives et en adressant les besoins réels des seniors. La méthode de conception, basée sur une approche centrée sur les utilisateurs potentiels et le marché potentiel, permet de s'assurer à chaque étape du développement de la plate-forme de l'adéquation du produit avec ses cibles. Les intérêts de recherche portent sur l'adaptation de services déjà disponibles mais difficilement accessibles pour la tranche de population concernée. En ce sens, des méthodes et algorithmes de l'état de l'art sont étoffés pour étendre la couverture des SDV.

## REMERCIEMENTS

Les recherches présentées sont effectuées dans le cadre du projet AAL vAssist.

## RÉFÉRENCES

- [1] European Union, "Active ageing and solidarity between generations," in *Eurostat, Statistical books*, 2011.
- [2] V. Bush, "As we may think," *The Atlantic*, 1945.
- [3] M. A. Walker, C. A. Kamm, and D. J. Litman, "Towards developing general models of usability with PARADISE," *Natural Language Engineering*, vol. 6, no. 3-4, pp. 363-377, 2000.
- [4] M. E. Burman, "Health diaries in nursing research and practice," *Journal of Nursing Scholarship*, vol. 27, no. 2, pp. 1547-5069, 1995.
- [5] I. J. Deary, J. Corley, A. J. Gow, S. E. Harris, L. M. Houlihan, R. E. Marioni, L. Penke, S. B. Rafnsson, and J. M. John, "Age-associated cognitive decline," *British Medical Bulletin*, vol. 92, no. 1, pp. 135-152, 2009.
- [6] B. Woods, A. E. Spector, L. Prendergast, and M. Orrell, "Cognitive stimulation to improve cognitive functioning in people with dementia," *Cochrane Database of Systematic Reviews*, vol. 4, 2005.
- [7] D. L. Morgan, "Focus groups," *Annual Review Sociology*, vol. 22, pp. 129-152, 1996.
- [8] J. F. Kelley, "An empirical methodology for writing user-friendly natural language computer applications," in *Proceedings of CHI*, 1983, pp. 193-196.
- [9] P. M. A. Desmet, C. J. Overbeeke, and S. J. E. T. Tax, "Designing products with added emotional value : Development and application of an approach for research through design," *The Design Journal*, vol. 4, no. 1, pp. 32-47, 2001.
- [10] B.-H. Juang and L. R. Rabiner, "Hidden Markov models for speech recognition," *Technometrics*, vol. 33, no. 3, pp. 251-272, 1991. [Online]. Available : <http://www.jstor.org/stable/10.2307/1268779>
- [11] C. Lee, S. Jung, and G. G. Lee, "Robust dialog management with n-best hypotheses using dialog examples and agenda," *Proc. of the Association for Computational Linguistics*, no. June, pp. 630-637, 2008. [Online]. Available : [http://isoft.postech.ac.kr/publication/iconf/acl08\\_lee.pdf](http://isoft.postech.ac.kr/publication/iconf/acl08_lee.pdf)
- [12] R. D. Mori, F. Béchet, D. Hakkani-Tur, M. McTear, G. Riccardi, and G. Tur, "Spoken language understanding : A survey," in *Automatic Speech Recognition & Understanding, 2007. ASRU. IEEE Workshop on*, 2007, no. 33549. [Online]. Available : [http://ieeexplore.ieee.org/xpls/abs\\_all.jsp?arnumber=4430139](http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=4430139)
- [13] F. Jurčiček, B. Thomson, and S. Young, "Reinforcement learning for parameter estimation in statistical spoken dialogue systems," *Computer Speech & Language*, vol. 26, no. 3, pp. 168-192, Jun. 2011. [Online]. Available : <http://linkinghub.elsevier.com/retrieve/pii/S0885230811000490> <http://www.sciencedirect.com/science/article/pii/S0885230811000490>
- [14] C. Rich, "Building task-based user interfaces with ANSI/CEA-2018," *Computer*, vol. 42, no. 8, pp. 20-27, 2009. [Online]. Available : <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=5197420>
- [15] M. Schröder and J. Trouvain, "The German text-to-speech synthesis system MARY : A tool for research, development and teaching," *International Journal of Speech Technology*, 2003. [Online]. Available : <http://www.springerlink.com/index/JX005727954L74J8.pdf>