

Les interactions fréquentes chez les plus jeunes avec des systèmes de dialogues ont-elles baissé leurs attentes face à ces systèmes ?

Charles Vin

Décembre 2021

1 Introduction

Lors d'un dialogue oral, pour un partage d'information efficace en terme de temps et de compréhension, chaque interlocuteur doit prendre en compte les connaissances de l'autre et s'adapter à celui-ci [1][2]. En effet, si ce n'est pas le cas, les conversations pourraient être trop détaillées et donc longues, ou alors à l'opposé, trop précises et donc complexes pour la compréhension de l'autre. Ainsi, les participants vont implicitement s'accorder sur le sens des termes employés en utilisant des indices linguistiques et physiques de l'environnement [3]. Cette connaissance implicite fait référence au terrain commun, définie par l'ensemble des connaissances que deux « entités » partagent et ont conscience de partager [1][3][4].

Pour mesurer ce phénomène, on utilise le très connu paradigme de Clark et Wilkes-Gibbs [5] où deux participants interagissent autour de figures de tangram abstraites. Chaque participant a un rôle prédéfini, l'un est appelé l'exécutant et l'autre le directeur. L'exécutant doit ordonner les figures de tangram dans un ordre prédéfini connu uniquement par le directeur. Les interactions sont libres et sans interventions de l'expérimentateur. Cette opération est répétée 6 fois avec les mêmes figures à chaque essai (uniquement l'ordre imposé varie). Typiquement, au premier essai, le directeur décrit les figures avec beaucoup de références indéfinies "la première carte ressemble à un genre de renard, mais avec une patte en moins". Au fur et à mesure des essais, les descriptions seront de plus en plus courtes avec les deux interlocuteurs s'accordant sur des références communes "Le renard amputé". L'utilisation de référence définie augmentera également avec le temps.

Le terrain commun se construit par une collaboration entre les deux participants. Le premier présente une information à intégrer au terrain commun avec des indices montrant son incertitude ("un ours", "une sorte d'ours"). En retour, le deuxième va pouvoir indiquer qu'il a compris, qu'il accepte l'information présentée. Par exemple en paraphrasant ce qu'il a compris ou en disant "oui" ou "ok". Si ce n'est pas le cas, il peut le signaler à son interlocuteur qui pourra ré-expliquer. Et ce jusqu'à ce que les deux parties se soit comprises.

Dans les interactions humain-système, le même principe s'applique. L'utilisateur a les mêmes attentes et croyances que dans le dialogue humain-humain. Par exemple, lors d'une interaction avec un système de dialogue à propos d'objets du quotidien, l'utilisateur va réutiliser les mêmes références, les mêmes mots que le système. Assumant que celui-ci doit être capable de les comprendre.

Ces interactions avec des systèmes de dialogue sont de plus en plus fréquentes de nos jours. Des assistants vocaux sont installés par défaut dans nos téléphones, des automates utilisant le langage naturel sont présents dans les gares et les aéroports pour nous permettre d'acheter des billets. Malheureusement, en attribuant les mêmes attentes que lors d'un dialogue humain, les interactions peuvent être frustrantes. En effet les systèmes sont rarement conçus pour prendre en compte les connaissances de la personne en face d'eux et construire un terrain commun.

1.1 Problématique

Même si l'utilisation de ces systèmes tend à se démocratiser pour tous et notamment dans l'aide à la personne. On peut supposer que les moins âgés d'entre nous en ont une utilisation plus fréquente que les plus âgés. Ils ont ainsi fait face à de plus nombreuses déceptions face à ces systèmes que des personnes âgées. Dans cette étude, nous allons examiner l'idée que cette frustration a diminué les attentes en terme de dialogue et d'utilisation du terrain commun chez les jeunes lors des interactions humain-système.

Pour cela, nous avons utilisé le paradigme de Clark et Wilkes-Gibbs décrit plus haut, lors d'une interaction avec un système de dialogue factice, en faisant varier l'âge des participants et le design du robot. En effet, les indices présents dans l'environnement peuvent agir sur les suppositions faites à propos des connaissances communes [6]. Faire varier ce paramètre aux deux extrêmes permet d'obtenir de plus gros effets dans les résultats tout en contrôlant son possible effet parasite (l'utilisateur a associé la frustration uniquement avec un certain type de système).

1.2 Hypothèses

En faisant varier l'âge, les plus jeunes étant habitués aux systèmes de dialogue de la vie quotidienne ne s'adaptant pas ou très mal à leur interlocuteur. On peut s'attendre à ce qu'ils construisent un terrain commun plus lentement que des personnes plus âgées.

En faisant varier le type d'agent, on peut penser que face à un robot humanoïde avec une modalité orale, les participants aient plus tendance à présupposer un terrain commun et donc à le construire plus

rapidement.

On peut également s'attendre à une corrélation négative entre la fréquence d'utilisation des systèmes de dialogue et la vitesse d'installation d'un terrain commun.

2 Méthode

Nous avons donc adapté le paradigme de Clark et Wilkes-Gibbs dans le cadre d'une interaction humain-système. Comme nous étudions uniquement l'aspect interaction et communication dans le dialogue, le robot n'a pas besoin d'être doté d'une intelligence. Ce sera ici un expérimentateur qui répondra et interagira avec le participant à son insu, à la manière du magicien d'Oz. L'important est que le sujet pense qu'il interagit avec une machine.

Pour cette expérience, le robot joue le rôle de l'exécutant, il doit retrouver les cartes pour les remettre dans l'ordre. Et ce à partir de la description faite par le participant, qui lui, joue donc le rôle de directeur. Le robot est tout à fait capable de poser des questions et de répondre à celles du sujet. L'échange est très naturel en dehors du délais de réponse du robot d'environ 3 secondes.

Traditionnellement, les deux sujets vérifient leurs réponses à la fin de l'essai. Ici une fois la carte devinée par le robot, celui-ci peut l'afficher dans l'écran de la salle d'expérimentation et en débattre avec le participant. Il y a 6 cartes par essai et 6 essais.

2.1 Plan expérimental

En faisant varier l'âge et le type d'agent, nous sommes donc face à un plan carré latin avec 2x2 conditions :

- Âge des participants :
 - Groupe 20 (G20)
 - Groupe 50 (G50)
- Type d'agent :
 - Humanoïde
 - Non-Humanoïde

Pour mesurer la construction du terrain commun, les dialogues seront enregistrés, transcrits. Les variables suivantes seront extraites pour caractériser le dialogue :

- Nombre de mot produit en moyenne par essais
- Probabilité d'utiliser un mot Hedge
- Probabilité d'utiliser des références définies

Finalement, un questionnaire permettra de mesurer le taux d'interaction quotidienne avec des systèmes de dialogue de chaque participants. Ce score permettra une analyse intra-participants avec le calcul d'un coefficient de corrélation. (Voir [Annexe](#)).

2.2 Participant

Nous avons recruté n participants qui ont tous pour langue maternelle le français.

Pour le premier groupe (G20), les participants auront une moyenne d'âge de 20 ans avec un nombre d'hommes et femmes équilibré, typiquement des étudiants.

Pour le deuxième groupe (G50), les participants auront une moyenne d'âge de 50 ans avec un nombre d'hommes et femmes équilibré.

2.3 Matériel

Six images de tangram ont été choisies puis placées aléatoirement sur une grille de six colonnes et six lignes. Les cases de chaque ligne étaient numérotées de un à six. Cette grille fût imprimée. Le sens d'utilisation de la feuille était indiqué par une phrase qui rappelait la consigne dites à l'oral. Pour des exemples d'images de tangram se référer à la figure 1.

Quant aux robots, notre choix s'est porté sur Nao de SoftBank Robotics. Très utilisé dans le milieu de la recherche et de l'éducation, il est peu cher et facile à programmer. Il sera utilisé dans la condition humanoïde avec ces animations et émotions par défaut. Dans l'autre condition, nous avons utilisé un système plus proche de celui des assistants vocaux. Il prenait vie sous la forme d'une sphère animée en même temps que sa voix sur l'écran de la salle d'expérimentation. La Figure 2 illustre les deux robots.

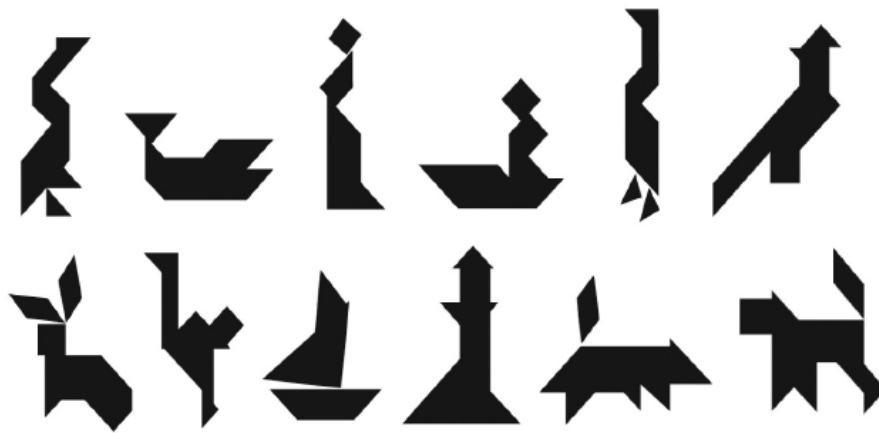
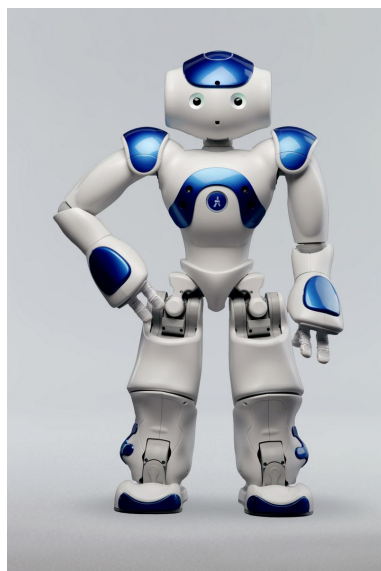
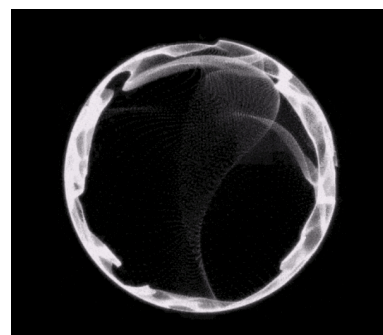


Figure 1 – Exemple de figure de tangram utilisé par les participants



(a) Le système de dialogue dans la condition humanoïde



(b) Le système de dialogue dans la condition non humanoïde

Figure 2 – Les robots utilisés

2.4 Procédure

L'expérience s'est déroulée dans une salle de passation de notre laboratoire, assis, face à un écran d'ordinateur dans la condition non humanoïde. Sinon assis face à Nao avec l'écran sur le côté droit.

Les participants avaient pour tâche de communiquer à voix haute avec le robot. Ils étaient informés qu'ils participaient à une expérience sur le dialogue humain-système et que le robot était doté d'une intelligence artificielle très performante. Mais, comme dit précédemment, en réalité derrière le robot se cachait un expérimentateur.

La consigne pour le participant était résumé en haut de la grille de tangram mais également dites à l'oral. Il était demandé au sujet de :

- Parler distinctement à voix haute
- De démarrer la conversation avec la description de la première image
- Puis de continuer naturellement la conversation avec les réponses du robot

3 Résultat attendu

On s'attend à ce que les résultats confirment les [hypothèses](#) en suivant le paterne de la Table 1 et de la Figure 3. C'est à dire que la courbe du G20 est systématiquement au dessus de celle du G50. De même pour le robot humanoïde qui faciliterait l'installation d'un terrain commun en comparaison avec le non-humoïde. Tous ces effets seraient significatifs.

Les mesures tels que le nombre de mot par essai reflètent l'efficacité de la conversation et donc quantifient l'utilisation du terrain commun. Lorsqu'une condition utilise en moins de mot qu'une autre condition sur tous les essais, cela indique une installation plus rapide du terrain commun. Ainsi le G20 construirait plus lentement un terrain commun que le G50. On présume que cela vient de l'interaction plus fréquente avec des systèmes, éventuellement des déceptions plus fréquentes, et donc une baisse naturelle des attentes sur les capacités des systèmes à prendre en compte leur interlocuteur.

On peut penser que la forme humanoïde d'un système est un indice conditionné qui va influencer les attentes en terme de dialogue de l'humain. Ainsi un système humanoïde provoquerait une utilisation du terrain commun plus naturelles et s'installant plus facilement. Dans nos résultats, cela se refléterait de la même manière qu'avec la condition âge.

Pour le questionnaire, on retrouverai probablement une corrélation positive entre l'utilisation des systèmes de dialogue dans la vie quotidienne et la vitesse d'installation du terrain commun. Qui appuierait l'interprétation faites dans les deux paragraphes précédant.

	Youtg	Old
Humanoïde	-	++
Chat Bot	- -	+

Table 1 – Vitesse d'installation symbolique du terrain commun en fonction des différentes conditions

4 Annexe

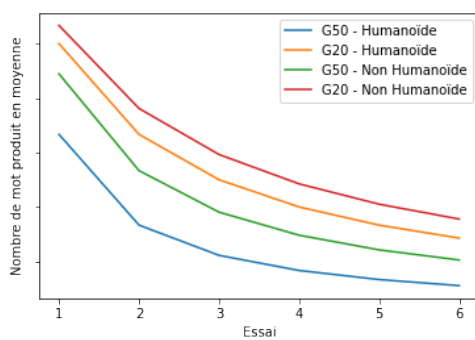
4.1 Exemple de question du questionnaire

Le questionnaire introduit la définition des termes utilisés dans les questions.

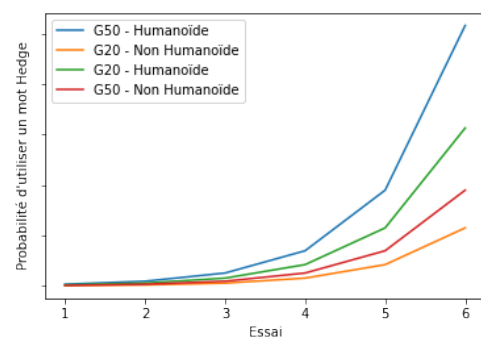
- Avez-vous une enceinte connectée chez vous ? A quelle fréquence vous en servez-vous ?
- Utilisez-vous l'assistant vocal de votre téléphone ? Vous a-t-il déjà frustré ou énervé ?
- Les systèmes de dialogue dans les lieux publiques (gare, aéroport) vous semblent-ils utiles ?

Références

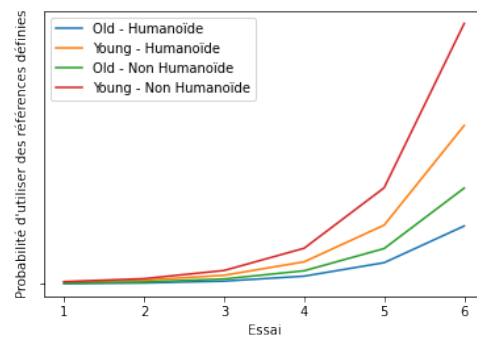
- [1] Herbert H. Clark. *Using Language*. 'Using' Linguistic Books. Cambridge University Press, 1996.
- [2] Christina Turner and Dominique Knutsen. Audience design in collaborative dialogue between teachers and students. *Discourse Processes*, 58(8) :703–725, 2021.



(a) Nombre de mot produit en moyenne par essais



(b) Probabilité d'utiliser un mot Hedge par essais



(c) Probabilité d'utiliser des références définies par essais

Figure 3 – Résultats attendus

- [3] Herbert H. Clark and Catherine R. Marshall. Definite knowledge and mutual knowledge. In Aravind K. Joshi, Bonnie L. Webber, and Ivan A. Sag, editors, *Elements of Discourse Understanding*, pages 10–63. Cambridge, UK : Cambridge University Press, 1981.
- [4] R. Stalnaker. Assertion. In P. Cole, editor, *Syntax and semantics*, volume 9, page 315–332. Academic Press.
- [5] H.H. Clark and D. Wilkes-Gibbs. Referring as a collaborative process. *Cognition*, 22(1) :1–39.
- [6] Aaron Powers, A.D.I. Kramer, S. Lim, J. Kuo, Sau-Lai Lee, and Sara Kiesler. Eliciting information from people with a gendered humanoid robot. pages 158 – 163, 09 2005.