Universitat Autònoma de Barcelona

From Telegraphic to Natural Language: an Expansion System in a Pictogrambased AAC Application

Joan Pahisa Solé

UNIVERSITAT AUTÒNOMA DE BARCELONA

DOCTORAL THESIS

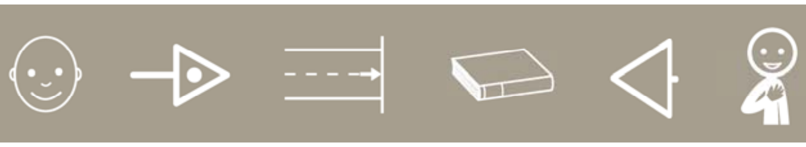
From Telegraphic to Natural Language: an Expansion System in a Pictogram-based AAC Application

Author
Joan Pahisa Solé

Supervisor
Dr. Jordi Herrera-Joancomartí

A thesis submitted in fulfilment for the degree of Doctor of Philosophy
Doctor per la Universitat Autònoma de Barcelona / PhD in Computer Science

Departament d'Enginyeria de la Informació i les Comunicacions



2017

Als meus pares

Abstract

In this doctoral dissertation, we present a compansion system that transforms the telegraphic language (utterances with only uninflected content words) that comes from the use of pictogram-based Augmentative and Alternative Communication (AAC) into natural language in Catalan and Spanish. The system has been designed to improve the communication of people who rely on AAC, who usually have severe speech or motor impairments and use pictogram-based communication methods in their daily life.

The compansion system has two main components: a syntactic-semantic dependency parser and a generator that constructs the final sentence. The system has been technically evaluated and results show that 99,66% of the sentences generated by it, taking into account the restrictions of a constrained grammar, were considered correct by three independent annotators.

Furthermore, a user interface with a pictogram prediction system has also been researched and implemented during the thesis in order to test it with end-users. The system as a whole was tested with 4 participants with severe cerebral palsy and ranging degrees of linguistic competence and intellectual disabilities. During tests, participants were able to learn new linguistic skills while using the compansion system, which proved a source of motivation. The system can also be adapted to the linguistic competence of each person and required no learning curve during tests when none of its special features were used. Finally, qualitative and quantitative results showed a mean communication rate increase of 41,59%, compared to the same communication device without the compansion system, and an overall improvement in the communication experience when the output is in natural language.

Keywords: Augmentative and alternative communication, AAC, natural language processing, compansion (compression-expansion), telegraphic language, Catalan, Spanish, pictograms, dependency parsing, machine translation, tests.

Acknowledgments

For me, a doctoral dissertation is not just the years since this research began, but all the years, all the effort, all the studies, all the options and all the decisions that have led me to this point in my life. Most things are part of a chain and when you put so much effort in a project, what is important is the entire path that you have done towards it.

Therefore, the acknowledgements are for all the people that have accompanied me throughout this journey, for all the people that have helped me reach this point and that have given me the tools to become the person that I am.

First of all, the best of my thanks go to my parents, as even though they could have, they never were overprotective, which I think is the first key to personal freedom. My mother has grown with me and has given me all the love that she could have and being loved instills the necessary confidence to face anything. My father has always been a reference of what hard work, constant studying and dedication could achieve.

To my sisters, Marta that has helped so many times during my studies, including this thesis, for her great counseling and for her selfless work, and Anna, for her artistic advise, for her liveliness and for being the necessary uncertainty that everyone needs.

To my partner, Quim, for being my home away from home and without whom these past few years would have been definitely tougher.

Then, aside from my family, I want to truly thank my supervisor, Jordi, because, even though this research entered some fields that were not of his expertise, even though the thesis did not follow the usual path, he believed in this project from day one, he believed in its potential to help people and has given me support and trust throughout the whole of it. Thank you.

As for institutions, I would like to thank the Autonomous University of Barcelona, the University of Algarve, the Erasmus Mundus Consortium, the FPU scholarship and the Talentum grant from Telefonica and the Adecco Foundation. Without your logistic and financial support this thesis would not have been possible either.

More thanks go to my masters' thesis supervisor, Jorge Baptista, as he guided the first steps of this research, to Jordi Escoin, for sharing his knowledge in AAC, to Jordi Lagares, for the time spent adapting Plaphoons, to Tim Barton, for his translations, and to all my grantees, Helena, Blanca, Xavi, Oscar, José Manuel, Jordi, Jorge, Rafa, Raúl and Héctor, for all the effort, all the bugs and all the joyful moments that we shared. A part of this thesis is also yours.

A very special thanks goes to Prodis Terrassa and their incredible human team for being a light in our path. They have given their all for this project and they were the best center where we could have conducted the tests with persons who rely on pictogram-based AAC. Marta, Àngels, Jose, Xavi and Xavi, Sonia, María José, Brenda and more, you are an essential part of Jocomunico. Your energy is contagious.

I would also like to apologize too to all the people, family and friends that, during these years, I could not spend the time that I would have liked with them and that they deserved. Thank you for understanding and for being supportive.

Finally, to all the teachers that I have had since elementary school, to all my friends, to the rest of my family, to everyone that for sure I have forgotten to mention, to all the good and not so good people that have crossed this path, as, without them, this project would have been a different one.

It has been a really long journey and many times I have thought of not reaching the end, but what are a few crossroads when you can make a positive impact in other people's lives? Everything has been worth it.

CONTENTS

Abstract .... **v**
Acknowledgments .... **vii**
Contents .... **ix**
List of Tables .... **xiii**
List of Figures .... **xv**
Abbreviations .... **xvii**
1. Introduction **1**
1.1. Definition of AAC 1
1.2. Motivation 3
1.3. Project objectives 4
1.4. Main contributions 5
1.5. Thesis outline 7
2. Background **9**
2.1. Brief history of AAC 9
2.2. Learning process of AAC 11
2.3. Current State of AAC 12
2.4. Future of AAC 15
3. State of the Art **17**
3.1. Natural language processing in AAC 18
3.2. Dependency parsing review 21
4. Compansion System **27**
4.1. Design considerations 27
4.1.1. Controlled grammar 30
4.1.2. Vocabulary 35
4.2. Input system and process of constructing a sentence 38
4.3. Parser 40
4.3.1. General characteristics 44
4.3.2. Parsing algorithm: Steps and rules encoded in the parser 48
4.4. Generator 51
4.4.1. Slot ordering module 52

4.4.2. Word order module, prepositions module and agreement module..... 52
4.4.3. Articles' module 52
4.4.4. Verb conjugator module 54
4.4.5. Cleaning module........ 54
5. Compansion System Evaluation and Results 59
5.1. Evaluation 59
5.1.1. Selection of the set of sentences for the test........ 60
5.1.2. Evaluation of the input........ 61
5.1.3. Evaluation of the parser........ 62
5.1.4. Evaluation of the generator 64
5.2. Results 66
5.2.1. Results of the evaluation of the input 66
5.2.2. Results of the parser 69
5.2.3. Results of the generator 70
5.3. Evaluation and results of the tests conducted in English........ 73
6. User Interface........ 79
6.1. Characteristics and design aspects 80
6.1.1. Device selection 81
6.1.2. Board customization........ 82
6.1.3. Scanning and selection methods........ 87
6.1.4. Historic and thematic folders 89
6.1.5. Addition of new vocabulary........ 90
6.1.6. Voices and voice synthesizer engines........ 91
6.2. Pictogram prediction system........ 92
6.2.1. Prediction system design aspects 92
6.2.2. Prediction system algorithm examples........ 94
7. Tests with People who Rely on AAC 109
7.1. Objective of the tests........ 110
7.2. Methodology........ 112
7.2.1. Description of the sessions........ 113
7.2.2. Quantitative indicators........ 114
7.2.3. Model for comparing results with and without compansion system.. 116
7.2.4. Session evaluation sheets........ 123
7.2.5. Sentence types........ 124
7.2.6. Participant profiles........ 126

7.3. Tests conducted 130
7.3.1. Procedure for the test sessions 130
7.3.2. Vocabulary 133
7.3.3. Language-related restrictions and the system's behavior 134
7.3.4. GUI 135
8. Results of the Tests 137
8.1. The companion system 137
8.1.1. General quantitative analysis 138
8.1.2. Quantitative analysis based on the number of pictograms 139
8.1.3. Quantitative analysis by sentence type 142
8.1.4. Qualitative analysis of changes in participants' language skills 143
8.1.5. Qualitative analysis of the communication experience 147
8.2. The graphical user interface (GUI) 148
8.2.1. Hit rate 149
8.2.2. Communication rate 150
8.3. Discussion of the results 151
8.3.1. The companion system 151
8.3.2. The graphical user interface (GUI) 155
9. Conclusions 157
9.1. Conclusions 157
9.2. Future work 161
References 165
Appendix A 175
Appendix B 177
Appendix C 183

LIST OF TABLES

Table 1: Examples of accepted and not accepted sentences by the parser. 31
Table 2: Examples of accepted and not accepted sentences regarding coordination by the parser 32
Table 3: Examples of accepted complements for nouns, verbs and adjectives. 33
Table 4: Amount of vocabulary and patterns in the compansion system before the tests with end users and after the tests. 37
Table 5: Comprehensive list of the sentence type modifiers in the compansion system 40
Table 6: Example of a pattern of the verb “to give”. 43
Table 7: Definite articles' sub-module algorithm: Conditions, algorithm and examples 53
Table 8: Lists of the algorithm 54
Table 9: Operators used in the algorithm 54
Table 10: Transformations of feeble pronouns in Theme and Receiver slots 55
Table 11: Preposition plus article contractions in Catalan 55
Table 12: Prepositions plus article contractions in Spanish 55
Table 13: Differences in the input of the sentences by the annotators 66
Table 14: Types of differences in the input 67
Table 15: Good, bad and identical sentences depending on the type of input 68
Table 16: Rates of the parser outputs 69
Table 17: Parser average score, % of good parses and Kappa 70
Table 18: Rates of the generator outputs 70
Table 19: Types of minor errors in the generator outputs 72
Table 20: Generator average score, % of good parses and Kappa 73
Table 21: Rates of the translated sentences 75
Table 22: Types of errors in the translated sentences 77
Table 23: Translated sentences average score and % of good and bad sentences 77
Table 24: Communication partner data 128
Table 25: Summary of the profiles of those who participated in the tests 129
Table 26: Summary of the total number of sessions and sentences formed by participants. 130
Table 27: Number of sentences formed per participant and per session 131

Table 28: The number of panels and pictograms that each participant had at the start and end of the tests 132
Table 29: Summary of the test sessions and the systems evaluated 133
Table 30: Core vocabulary words defined before the tests and used during the tests133
Table 31: Overall averages for the 333 sentences created by participants during the tests 137
Table 32: Average performances of the compansion system by number of pictograms per sentence 140
Table 33: Average performances of the compansion system for prepared and unprepared sentences 143
Table 34: Comparison of the number of pictograms used by each user in prepared and unprepared sentences 147
Table 35: Averages for the 81 sentences formed using the prediction and compansion systems together vs. averages for the sentences formed without the systems...149

List of Figures

Figure 1: Example of input with a tense modifier. 29
Figure 2: Components of the companion system. 29
Figure 3: Example of a sentence with an order modifier using telegraphic language. In this sentence there is a reduction of the subject and a default subject for an order is used instead. 30
Figure 4: Screenshot of the alternative input system. Lists of words are grouped into different categories using a common colour legend for pictogram-based AAC. 39
Figure 5: Example of the insertion of a generic expression. 41
Figure 6: Example of a special default subject and imperative verb form. 44
Figure 7: Example of a 2-level tree that represents a simple sentence structure. Labels have not been added to ensure legibility. Dotted dependencies can appear multiple times. 46
Figure 8: Example of fusion of patterns for inputs with two verbs. The slot that needed a verb disappears and is replaced by the slots of the pattern of the Secondary verb. 48
Figure 9: Hierarchy of noun classes in the system. A noun can have multiple classes. 49
Figure 10: Rules that define how the parser scores the fit of a word with a certain slot. 50
Figure 11: Recreation of the parse tree of an example input sentence in Catalan. 51
Figure 12: Screenshot of an example of a final output of the companion system 56
Figure 13: Screenshot of the output and the evaluation form with the non-accessible interface 65
Figure 14: Example of a board in our user interface with its main elements highlighted 83
Figure 15: Screen to edit boards. In this example, it is the main board of a user that has the vocabulary clustered in both syntactic categories (e.g. verbs, adjectives, etc.) and semantic categories (e.g. people, places, fruits, etc.) 84
Figure 16: Settings to customize the user interface on the Configuration screen 86
Figure 17: Configuration settings for the available selection methods (mouse/finger, 1-switch scanning and 2-switch scanning) 89

Figure 18: Configuration screen where users can activate and deactivate the prediction system as well as the number of displayed pictograms..95
Figure 19: Predicted pictograms for all three examples when there are no selected pictograms yet...96
Figure 20: Predicted pictograms for Example 1 when there is 1 pictogram in the sentence bar 97
Figure 21: Predicted pictograms for Example 2 when there is 1 pictogram in the sentence bar 99
Figure 22: Predicted pictograms for Example 3 when there is 1 pictogram in the sentence bar 100
Figure 23: Predicted pictograms for Example 1 when there are 2 pictograms in the sentence bar 102
Figure 24: Predicted pictograms for Example 2 when there are 2 pictograms in the sentence bar 103
Figure 25: Predicted pictograms for Example 3 when there are 2 pictograms in the sentence bar 104
Figure 26: Predicted pictograms for Example 1 when there are 3 pictograms in the sentence bar 106
Figure 27: Predicted pictograms for Example 3 when there are 3 pictograms in the sentence bar 107
Figure 28: Flow and functions for comparing test results with and without the *companion* system 117
Figure 29: Screenshot of the auto-scan options 135
Figure 30: A prediction column (left) with a customized pictogram (top) within the interface 136
Figure 31: Words per click per user with and without the compansion system (% increase) 139
Figure 32: % rate increase vs. number of pictograms per sentence 141
Figure 33: no. of pictograms used and rating per sentence by user language skills ..142
Figure 34: Evolution of the communication rate in *wpm* for each user and session ...151

ABBREVIATIONS

AAC	Augmentative and Alternative Communication
ALS	Amyotrophic Lateral Sclerosis
API	Application Programming Interface
ARASAAC	Portal Aragonés de la Comunicación Aumentativa y Alternativa
ASD	Autism Spectrum Disorder
BY-NC	Creative Commons License: By and Non-Commercial
CACE	basic set of pictographic vocabulary for AAC developed by UTAC
CCN	Complex Communication Needs
CPU	Central Processing Unit
Compansion	Compression-Expansion is a technique that expands uninflected content words into syntactically and semantically well-formed sentences
cpm	clicks per minute
CR	Communication Rate
EBP	Evidence-Based Practices
GPS	Global Positioning System
GUI	Graphical User Interface
HMM	Hidden Markov Models
HTS	HMM-based speech synthesis
iOS devices	iPhone, iPad and iPod Touch
ms	milliseconds
MS	Multiple Sclerosis
NLP	Natural Language Processing
NP	Non-Deterministic Polynomal Time
PCS	Picture Communication Symbol
PESICO	Patient Environment Stakeholders Intervention Comparison and Outcome
PhD	Doctor of Philosophy

PICO	Patient Intervention Comparison and Outcome
RAE	Real Academia Española (Spanish Royal Academy)
RI	Rate Index
SAW	Special Access to Windows
SGD	Speech Generating Device
SMT	Statistical Machine Translation
SOV	Subject-Object-Verb
SPC	Sistema Pictográfico de Comunicación
SR	Selection Rate
SVO	Subject-Verb-Object
TBI	Traumatic Brain Injury
TDG	Topological Dependency Grammar
TTS	Text-To-Speech
UAB	Universitat Autònoma de Barcelona
UTAC	Unitat de Tècniques Comunicatives de Comunicació of the University of Barcelona
UTF	Unicode Transforming Format
VSD	Visual Scene Display
XDG	Extensible Dependency Grammar
WCR	Word-Click Rate
wpb	words per bit
wpc	words per click
wpm	words per minute

CHAPTER 1

INTRODUCTION

1.1. Definition of AAC

Augmentative and Alternative Communication (AAC)[1] (Larranz, 2006) is the set of ways, methods and strategies used by people with certain disabilities that prevent them from communicating normally through natural language or speech. In order to maximize communication, intervention programs are designed to enhance the communication capacities of the user. Residual speech, gestures, communication through graphic signs (pictograms), the use of special communicators or other systems to facilitate access to computers are some of the most common methods for AAC users to communicate. All these interventions also need to take into account the preferences and priorities of the user and his motor, sensory, cognitive, linguistic, psychological and behavioral skills.

Note that people from all ages can have communication impairments, from birth to adulthood. The source of the communication disorders (Kent, 2007) is also vast: cerebral palsy, autism, mental deficiency, aphasia (Kraat, 1990), dysphasia, dementia, multiple sclerosis (MS), amyotrophic lateral sclerosis (ALS), Parkinson, Alzheimer and traumatic brain injury (TBI), among others. While some AAC devices, techniques and strategies require certain language or technical skills and abilities, AAC interventions can help support communication to individuals across all skill levels and ages (Blackstone, 2007).

The main goals of AAC are the following:

---

[1] What is AAC? http://everyonecommunicates.org/aacintro.html & International Society for Augmentative and Alternative Communication. https://www.isaac-online.org/english/what-is-aac/ [August 1st, 2016]

- Provide alternative means of communication until speech is fairly restored.
- Provide alternative means of communication for lifetime, when speech is non-existent or severely affected.
- Support the development and restoration of speech and other linguistic skills.

There is the wrong belief that AAC might hinder the development of a person or the return of his speech. Actually, it is the opposite; in many cases, intelligible speech improves after AAC is introduced to an individual (Blackstone, 2007). Therefore, when an AAC intervention begins, it does not mean that professionals are giving up on speech, contrary to what was thought before research on AAC began.

In turn, communication partners also play a critical role in AAC. First of all, many of the AAC users require a speech language pathologist or a member of their family to set up and customize the AAC device or application of the user. Then, in low-tech communicators (Augé & Escoin, 2003), such as paperboards, the intervention of the partner is crucial, because they play the role of the voice synthesizer in high-tech devices and need to make sure that what they are saying is in fact what the user is pointing and, in other words, what the user wants to say. Usually, primary communication partners play multiple roles on the lives of persons who rely on AAC. Apart from conversational partners, they can be AAC facilitators, technicians of the AAC communicator, trainers of other partners or even caregivers. Therefore, AAC systems need to take into account how to support interactions with a wide variety of communication partners, ranging from experts to people unfamiliar with AAC.

To sum up, AAC technologies allow individuals with complex communication needs[2] (CCN) to expand their social network and to help them fulfill their roles into society as family members, friends or employees at work.

[2] The estimated number of severely speech impaired people living in the European Union is two million. Studies reveal that around 1.5% of the population experience communication disorders (Golinker, 2009).

1.2. Motivation

Regarding scientific research in AAC, the potential use of natural language processing (NLP) or machine learning techniques in order to improve communication has still many possibilities to be explored. There are studies that reveal that the mean ratio of AAC systems is between 5 to 20 words per minute (wpm), much lower when direct letter or pictogram selection cannot be used. These values fall way behind the ratio of a standard conversation, which is between 150 and 200 wpm (Copestake, 1997). Obviously, achieving normal ratios in AAC is impossible. Apart from trying to directly increase the ratio to improve the speed of communication through better hardware or through better selection techniques, reducing the number of keystrokes or gestures that a user needs to perform in order to communicate, by means of prediction techniques, would also increase the ratio and reduce the amount of effort of the user. Also, *compansion* (language compression and expansion) techniques that post-process the user input expanding the telegraphic language[3] that often results from AAC, can also increase the amount of words entered per keystroke, improving the communication ratio and also the overall communication experience by both the user and the communication partner. Furthermore, compansion techniques can also be beneficial for literacy purposes in AAC users (Pennington and McCoy, 1998).

To the best of our knowledge, when this thesis began, no AAC application that expands telegraphic language into natural language still existed for Catalan or even for Spanish. Currently, an AAC application that does language expansion named Liberia Community[4] exists in Catalan and Spanish, although it is a closed communicator (without the possibility of configuring its interface or its vocabulary) and it uses a different expansion approach, powerful but with a very restricted controlled grammar and less flexible than the expansion system researched in this thesis.

Furthermore, due to telegraphic language and other social prejudice, persons that are not familiar with people with CCN might not take AAC users seriously or might

[3] Telegraphic language is characterised by containing only meaningful words (Covington, 2001).

[4] Liberia Community. http://liberiacommunity.net/ [August 1st, 2016]

use patronizing attitudes. In turn, people that have full or nearly intact linguistic competence, but that have speech impairments, also feel frustrated when the AAC systems that they use cannot produce natural language.

It is important to note that, despite improvements in recent years, as the mentioned above, unfortunately, there is still a wide gap between research studies and the actual AAC systems for the end users.

In view of all these facts, taking into account the social and psychological benefits of being able to communicate using natural language, both for users and their communication partners, the goals of this thesis will be centered around researching a solid language expansion technique and, on a second term, a user interface that includes prediction techniques, making special emphasis in applying them to a working AAC system and testing it with real AAC users.

1.3. Project objectives

The main goal of this thesis is the creation of a telegraphic language expansion system that transforms it into natural language, which will improve communication, both communication ratio and qualitatively, in daily activities of people who rely on pictogram-based AAC. A secondary goal is to create a user interface adapted to the users' needs that includes a pictogram prediction system that will take advantage of the language model built for the expansion system. Finally, both systems will be presented by means of a Picture Communication Symbol (PCS)[5] (Augé & Escoin, 2003) application in order to test the research with end-users.

Here are the research objectives summarized in more detail:

1. The creation of a compansion system that will expand telegraphic language, reordering and inflecting words and adding articles, prepositions and other necessary grammatical elements/words, within a controlled grammar with the following main characteristics: composed of two components, a

[5] PCS Software. http://www.widgit.com/aboutus.htm & Picture Communication Symbols. http://en.wikipedia.org/wiki/Picture_communication_symbols [August 1st, 2016]

dependency parser that will be mostly language independent, ready to accept languages with different sentence structures, and a language-dependent generator that will be developed in Catalan and Spanish; a core of vocabulary specifically chosen for AAC communication that will be scalable; controlled grammar that allows for simple sentences, but that also accepts certain types of subordination; modifiers to allow for several verb tenses and different types of sentences (declarative, order, question, etc.).

2. The development of a user interface tailored to the needs of people who rely on pictogram-based AAC (with the minimum accessibility requirements to be able to conduct tests with end-users) that will include features taken from works done by other researchers, such as a historic of sentences accessible by the AAC users themselves and a set of customizable thematic folders with sentences to prepare conversations in advance.
3. As well as the aforementioned features, another goal related to the user interface is the creation of a prediction system that will learn from the usage of each user, taking into account global frequency of usage and frequency of usage in different time frames, and that will also take into account the controlled grammar, the language model of the parser and the semantic information in the annotated corpus of vocabulary within the system.
4. Tests with persons who rely on AAC to see the performance of the systems in real situations, to validate the research done and to help better assess future lines of work.

1.4. Main contributions

All the goals of the thesis have been met and have turned into contributions. Regarding scientific publications, a detailed description of the compansion system in its early stages, just for Catalan and without many of its current features, can be found in the bulletin paper:

> Pahisa-Solé, J. (2012), 'Compansion' system for a pictogram-based AAC application in Catalan, *Bulletin de Linguistique Appliquée et Générale (BULAG)*, num. 36/2012, Université de Franche-Comté, Besançon.

A later description of the compansion system, already with generators for Catalan and Spanish and with nearly all the features, can be found in the following conference paper and also in a brief graphical summary of how it works in a conference poster (https://goo.gl/ryaCaG) with the same name as the paper:

> Pahisa-Solé, J. and Herrera-Joancomartí, J. (2014), Pictogram AAC prototype that expands telegraphic language into natural language in Catalan and Spanish. *In Proceedings of the 16th Biennal Conference of ISAAC: Discover Communication*. Canada. ISBN: 978-0-9881189-2-8.

Finally, a description of the tests conducted with people who rely on AAC with a varying degree of linguistic competence, which also contains a summary of the complete compansion system, can be found in the following journal paper:

> Pahisa-Solé, J. and Herrera-Joancomartí, J. (2017), Testing an AAC System that Transforms Pictograms into Natural Language with Persons with Cerebral Palsy. *Assistive Technology* (pending volume number). doi: 10.1080/10400435.2017.1393844

Nevertheless, the full updated description of all the systems and of all the conclusions from the tests with persons who rely on AAC can only be found in this doctoral dissertation.

The final contribution, aside from scientific publications, is a working pictogram-based AAC application that includes all the conducted research. It is a free open-source application named Jocomunico (www.jocomunico.com) and the first version of its code is available under a Creative Commons BY-NC license in the following GitHub repository: https://github.com/narum/jocomunico/. The second version is at: https://github.com/narum/jocomunico2/. As of October 2017, more than 1000 users, hospitals, special education schools and centers for people who rely on AAC have started using it.

1.5. Thesis outline

The rest of this doctoral dissertation is structured as follows.

Chapter 2 presents a historical background on AAC. Being a technological thesis, a brief overview on the history and the current state of AAC would help better understand the context of our research. Chapter 3 then presents a literature review on the state of the art regarding natural language processing applied to AAC, specially focusing on research centered in transforming telegraphic inputs into natural language. An overview on dependency parsing, which our compansion system is based upon, is also presented in the same chapter.

In Chapter 4, we discuss the methodology used to design and develop the compansion system, as well as describe its characteristics and its algorithm in detail. In Chapter 5, we then present the evaluation methods of the system and the obtained results.

Chapter 6 discusses all the components and the accessibility requirements of the user interface that we developed in order to do the tests with end-users. The chapter includes a description of the features of the prediction system and its algorithm.

Chapter 7 explains the tests that we conducted with people who relied on a regular basis on pictogram-based AAC in order to communicate. The chapter defines the goals of the tests and details the methodology used and the profile of the participants that took part in them. Finally, it describes and quantifies the tests carried out. The qualitative and quantitative results of the tests, as well as its conclusions, are in Chapter 8.

Finally, Chapter 9 summarizes the work and the conclusions extracted during the whole thesis and it also discuses possible future lines of research derived from our doctoral dissertation.

Chapter 2

BACKGROUND

This is a multidisciplinary doctoral dissertation and probably, many readers with a computer science or a linguistics background, will not be familiar with Augmentative and Alternative Communication. Therefore, we decided to add this chapter that will give a brief overview on the field of AAC in order for readers to better understand the scope of this thesis, as many of the research done and the decisions were taken within the framework of the AAC field.

2.1. Brief history of AAC

Before 1960, the only population that was eligible for alternative communication strategies were literate individuals, without any other language disorder apart from speech problems, which necessarily had the ability to point. The AAC methods available were limited to typewriters or to boards with the alphabet written on them (Higginbotham, 2007).

During the 60s, professionals realised that they were not obtaining the expected results using traditional methods. That is when symbols were first introduced as an AAC technique (Larranz, 2006). The focus shifted from speech and language itself, to the ability of communicating regardless of the means. Also, thanks to the electronic innovations, the first dedicated communication devices were created (namely the POSSUM communicator).

Moreover, also in the 60s, scanning was introduced (Lesher et al., 1998). By pressing a button or a switch, a user could move a cursor from one item (be it a letter, a word, a sentence or a message) to another. This innovation was the key to opening AAC to almost anyone, regardless of the severity of their physical disabilities.

On the 70s, communication systems based on graphical symbols, like Bliss (Yovetich & Young, 1988) or Makaton (Grove & Walker, 1990), appeared. Through the combination of these symbols, these systems create a language of their own in a way similar to how single Chinese ideograms combine to form new meanings. Consequently, the user needs to go through a learning process before fully using the system. Despite the learning curve, which is still now being discussed by the AAC professional community, these systems obtained encouraging results (Bliss is still widely used nowadays), which led to a spreading of AAC systems. Also in the 70s, the previously developed communication devices started to be manufactured, thus the AAC industry was born.

During the 80s, manufacturers introduced rate enhancement techniques (Garay-Vitoria & Abascal, 2004), like abbreviation expansion and linguistic prediction techniques. The appearance of the microprocessor revolutionised mainstream technology. This change was soon reflected in the type of AAC devices that became available. The first prototypes of AAC computer programs were also developed and more pictographic languages were created. These, like Minspeak (Mathisen et al., 2009), relied on multi-meaning drawings, which depend on the order of the sequence in which the drawings appear.

In the social aspect, in some countries, laws that demanded civil rights for disabled people were passed and universal access to communication began to be a public policy issue.

Eventually, in the 90s, new improvements on AAC technologies, such as synthesised speech, dynamic displays, highly accessible symbol sets and smaller and more powerful AAC devices, extended the access to AAC communication devices. Unfortunately, most of these speech-generating devices (SGDs) were very difficult to use and had a steep learning curve, even for family members and AAC facilitators.

2.2. Learning process of AAC

The evolution in AAC research reflects the process of learning the use of AAC systems for individuals with CCN. First, communication learning starts with tangible objects, specially adapted games and toys for children. Then, patients move to photographs, which are related to their environment, and afterwards to images. At this stage, visual scene displays (VSDs) can be introduced to users. VSDs are specially indicated to children (Reichle & Drager, 2010) and to users with certain types of aphasia (Peña-Casanova, 1991).

Capable individuals continue by integrating the images, pictograms or graphic symbols into communication boards that usually display these images on a grid ordered by semantic groups or by syntactic categories (like nouns, verbs, adjectives, etc.). In most modern PCS systems, to better distinguish these categories, each of them has a colour associated, be it the background colour of the image or the colour of the frame of the image on the grid. Basically, the aforementioned VSDs, instead of placing pictographic symbols, use photos of experiences, situations or contexts familiar to the individual. The colour caption palette usually used is the following:

- Yellow for nouns, proper nouns and pronouns that refer to people.
- Green for verbs.
- Orange for common nouns.
- Pink for vocabulary or expressions related to social relations.
- Blue for adjectives.
- White for numerals, colours, days of the week, holidays (Christmas, Easter, etc.), modifiers and other miscellaneous words and expressions.

Communication boards can be paper or carton made boards or even electronic boards with their own sound output. Some of these basic communicators have pre-recorded sentences that are accessed through a combination of pictograms. However, the number of sentences is limited and, although the device is able to output proper natural language sentences, which is a positive aspect that we will later discuss, the system does not let the user create new sentences, which is also a big drawback.

Finally, the last step is the transition from static boards, be it manual or electronic, to dynamic boards, which allow to drastically increase the amount of vocabulary available to the user. Most of these dynamic boards are part of AAC software developed for specific AAC electronic devices or for regular computers or other portable devices available for the general public. The complexity of electronic communication devices and their AAC software varies widely based on the needs and the cognitive abilities of the user. These devices can have fully synthesised voices, contrary to the previously mentioned electronic boards that just had prerecorded words or sentences. Specific means of accessing these systems are adapted to the needs of each individual and may involve direct use of the keys or of the touch screens, use of head-mounted pointers, use of gaze-tracking technologies or use of external switches to control the devices.

2.3. Current State of AAC

At the moment, the leading AAC commercial software, which can be adapted to many of the individuals with CCN, is a software for Windows-based computers called The Grid 2 (Encarnação et al., 2014) manufactured by Smartbox Assistive Technology, which just released an updated version of the program named The Grid 3[6]. The Grid 2 provides a fully customizable dynamic board interface and is available in many languages, including English, Spanish, French or Portuguese. The user can configure each pictogram of the board so that it represents a single letter, a word, an action (i.e. erase the previous selected character) or so that it represents a full set of pictograms and leads to another board. This way, the user or the facilitator can choose the vocabulary according to the needs of the user and embed boards into boards and freely organise them. However, the synthesised output of the sentences that The Grid 2 offers does not produce natural language sentences, in other words, if the user utilises pictogram-based communication and uses a synthetic or telegraphic language, the system does not transform this telegraphic language into properly constructed sentences.

On the other hand, another leading software, far less spread, Speaking Dynamically Pro (Zygo Industries, 2004), gives the possibility of language

[6] The Grid 3 - https://thinksmartbox.com/product/grid-3/ [Last accessed, October 24th, 2016]

expansion, although it has to be completely programmed by the user and, overall, is not as user friendly as The Grid 2. This software, apart from incorporating all the main characteristics of The Grid 2, allows the user to give features to each pictogram (for example, gender or tense) and associate actions to them. It even permits to set features depending on the features of the other pictograms on the same sentence. Every time that a certain pictogram is used, the software checks its features and its dependencies and applies the programmed actions (i.e. according to the gender, adding the corresponding article, in romance languages, or according to the tense specified by a temporal expression, inflecting the verb). In its new version, The Grid 3, the facilitator can manually associate some articles, prepositions and verb conjugations to the pictograms, as in Speaking Dynamically Pro, although the possibilities are more limited and it has still the same drawbacks. However, to the best of our knowledge, as these programs have been developed by private companies, no scientific research has been published on how both Speaking Dynamically Pro and The Grid 3 approach natural language generation from pictogram-based communication.

Regarding accessibility, both programs are accessible using a wide array of external devices, ranging from regular keyboards and mouse to external switches, head or eye-pointer and any other devices that can substitute the functions of the mouse. Moreover, both systems incorporate word prediction, for individuals that use keyboards, which communicate through regular written language, be it physical keyboards or on-screen keyboards. Aside from AAC communicators, these programs also help the user access other features of the computer, such as Internet browsers, social media, music libraries and more.

Another important aspect of AAC that has greatly improved during this last decade are voice synthesisers. Whilst voice synthesisers used to sound robotic-like as they completely neglected prosody and sometimes were hardly understandable (and even less by people with language disorders), nowadays the most advanced ones feel fluent and natural. Synthesising methods evolved from diphone voices, to HTS voices, which use Hidden Markov Models (HMM), to the more natural clunit voices (Bonafonte et al., 2009), which are based on concatenative speech synthesis. Clunit voices sound more natural than HTS. On the other hand, HTS voices are usually smoother and more stable, while clunit voices can produce rather frequent concatenation errors. HTS voices also require far less memory and CPU resources.

Concerning pictogram-based communication, apart from the aforementioned symbol languages, like Bliss or Minspeak, other, more straightforward, and intuitive pictogram systems are available, like the widely spread SPC (Sistema Pictográfico de Comunicación) system (Augé & Escoin, 2003), which can use different symbol sets like Widgit-Rebus (Detheridge, 2005) or the open source ARASAAC set (Cabello-Luque & Bertola-López, 2012). In SPC, each pictogram or drawing represents a single word, whose meaning is much easier to guess without nearly any learning process than in highly symbolic languages like Bliss.

Finally, regarding the AAC scene in Catalan, during the 20th century, before the appearance of the first computer-based AAC systems, boards and other AAC methodology were primarily developed for Spanish and could be easily ported to Catalan. However, since then research progressed much slower in Catalan, as most of the effort put in developing AAC software was focused in Spanish, until the appearance of Plaphoons [7] (Maia & Cruz, 2010) a free open-source AAC application for Windows, available both in Catalan and Spanish. Plaphoons allows for embedded boards like The Grid 2 or Speaking Dynamically Pro, but lacks word prediction or any other advanced linguistic features, although it allows external switches, head-pointers and similar peripherals to access it. Since late 2009, The Grid 2 and Speaking Dynamically also have basic support, in its non-linguistic features, for Catalan.

Following the steps of Plaphoons, later freeware, like eMintza [8] and inTic [9] (González Rus & Liébana, 2013) and, the aforementioned in the previous chapter, Liberia Community have also been developed for Spanish and Catalan.

---

[7] Plaphoons – Download. http://projectefressa.blogspot.com.es/2016/01/plaphoons-download.html [September 12th, 2017]

[8] e-Mintza – Fundación Orange. http://www.fundacionorange.es/aplicaciones/e-mintza/ [September 12th, 2017]

[9] In-TIC – Fundación Orange. https://www.proyectosfundacionorange.es/intic/ [September 12th, 2017]

2.4. Future of AAC

While in the past decade AAC interventions have mostly focused on computers, on software for laptops and for specific AAC portable devices, such as DynaVox[10] (McCleod, 2011), with the fast growth of tablets in the last couple of years, the tendency has changed. Multi-modal devices that are able to perform different tasks (mailing, web browsing, agenda, calculator, accessing social networks, calling, text messaging, gaming, online shopping, listening to music, etc.) are taking over the world. As people with CCN should not be left apart, AAC devices need to be able to perform all these tasks and be compatible with mainstream technologies. That is why computers, although less portable than specific portable AAC devices, were prefered over the latter. But with the sudden rise of tablets and multi-purpose cellphones the picture has changed.

Nowadays, portable devices are gaining ground as AAC technologies for individuals with minimal to mild physical disabilities (i.e. people with Down syndrome, aphasia, some kinds of cerebral palsy, autism, etc.) who want extremely portable devices, with multiple functionalities and a trend factor (Blackstone, 2009). As time goes by, some peripherals, such as bases to attach devices to a wheelchair, will make them more suitable for users with higher levels of disability, in the same way that it was with desktop computers and laptops. Sound and speech technologies offer natural sounding and amplification, which makes it possible to hear speech generating devices in outdoor conditions, such as classrooms, parties or meetings.

Laptops are too frail for many AAC users, some of whom do not have good motor control. They are not very usable outdoors, where sunlight, weight and short battery life are huge problems. Desktop computers are still widely used too as AAC communicators, but they are located in specific places and can only be accessed in adequate circumstances. Furthermore, in most occasions, desktop computers are placed next to the wall, making face-to-face communication even more difficult. People with CCN do not want to be tethered to a wall or attached by wires (Escoin, 2006).

---

[10] Tobii DynaVox. https://www.tobiidynavox.com/en-us/ [September 12th, 2017]

The key to solving all these issues seems to lie in tablets. A good sign in this direction is that tablets like iPad already offer multiple accessibility options for disabled users, like VoiceOver, that can also benefit future AAC users. A few years ago, tablets still did not possess the processing capabilities of desktop computers nor laptops, and did not allow for fully-fledged AAC applications to be developed for them yet. At that time, basic AAC applications for smartphones and tablets appeared, such as Proloquo2Go (Sennot & Bowker, 2009), for iOS devices, or TapToTalk (Islas et al., 2013), for iOS devices and Android.

Nowadays, as stated in McNaughton and Light (2013) and in Shane et al. (2015), many options for aided communication are tablet-sized computers that can either be AAC dedicated or apps within a multi-purpose device. These dynamic AAC solutions allow for flexibility in display design, vocabulary selection and organization and navigation between boards. Furthermore, most high-tech AAC technologies allow for some kind of prediction to reduce the amount of keystrokes necessary for selection. Some current AAC solutions, like the previously mentioned The Grid 3, include both the desktop version and an app version for tablets.

Whilst some people may be reluctant to start AAC intervention as there is the believe that it may hinder the future acquisition of oral communication, regarding to apps and AAC intervention, emerging research has demonstrated their efficacy in teaching many communication skills to people with autism spectrum disorder (ASD), statement that can be translated to other people with CCN (Reichle et al. 2016).

As seen up until now, AAC is undergoing a major evolution that is trying to embrace the true meaning of communication. The first steps of AAC focused on expressing basic needs and exchanging greetings. While it is an important step to functional use of SGDs, true communication requires that the user has means to generate language at any time, anywhere and with anyone. Communication is not just the goal of AAC interventions in itself, but rather a way to many ends. To participate actively in the community or inside the family, to pursue personal interests and goals, allowing the user to lead a high quality lifestyle and achieve self-realisation, this is the true purpose of modern AAC (Blackstone, 2009).

Chapter 3

STATE OF THE ART

As stated in McNaughton and Light (2015), a review on the research done in AAC the past 30 years, there is a large number of areas in AAC that still need high-quality research. Research in AAC has historically focused mostly on interventions from a speech language pathologist's, a teacher's or a therapist's point of view and technical specifications of the systems used are rarely found. These systems have been mostly developed by private companies, thus technical literature is even harder to find. In the previous Background chapter, it can be seen that most references regarding systems, are from studies that used these systems and not actual reviews on the technical aspects of these systems.

Nevertheless, in the last 15 years, although AAC is still a small field, more researchers from different fields have started to do research on AAC. Despite that, regarding language expansion techniques there is not a large amount of literature available, as most of the researched systems have, at most, been prototypes and have not been thoroughly tested with end-users in order to become hot topics that researchers continued to focus on. Hopefully, this explanation will help readers, which might not be familiar to the field, keep in mind that in this specific area of AAC there is not a large amount of literature available in comparison to other areas in the computer science field.

Having said that, in the first section of this chapter we will present the most relevant research related to natural language processing applied to AAC. We will focus both on research centered in transforming telegraphic inputs into natural language and also on research related to pictogram prediction. Meanwhile, in the second section, we will present a review on dependency parsing, as the parser of our expansion system, which is its main component, is based on dependency grammar.

3.1. Natural language processing in AAC

In terms of scientific research, applying effective natural language processing techniques in AAC for improving the message output has still many possibilities to be explored (Copestake, 1997). The pace of communication using AAC devices is much slower than natural speech (McNaughton et al., 2003), even more when selection is not done directly (Horstmann Koester & Arthanat, 2017). *Compansion* (language compression and expansion) techniques (Demasco & McCoy, 1992) that post-process the user input expanding the telegraphic language that often results from AAC (e.g. pictogram-based AAC), which is characterized by containing only meaningful words, can increase the amount of words entered per keystroke, improving the communication rate and, more importantly, the overall communication experience by both the user and the communication partner. Besides, telegraphic language can be ambiguous and it can require several questions by the communication partner to uncover its meaning that further slow communication.

Furthermore, people that have severe speech impairments, feel frustrated when their AAC systems cannot produce natural language (Lee et al., 2006). Also, people not familiar with AAC can associate a lack of speech (or a telegraphic speech) with a lack of intelligence (Pennington & McCoy, 1998). Moreover, for people that have affected linguistic competence, a system that produces natural language sentences can be a tool that supports literacy and language learning or that even helps in language rehabilitation (Blackstone, 2007), as correct feedback would result from an ill-formed input (Pennington & McCoy, 1998).

The first description of a compansion system is by Demasco and McCoy (1992). They present a text-based compansion system that uses the semantic case frames of the verb (Fillmore, 1976) present in the input, as well as syntactic and semantic information of the other words in the input to expand telegraphic language into well-formed sentences. In 1998, Pennington and McCoy present a further development of this system that focuses on providing feedback of the correctness of the sentences produced in order to promote literacy for persons who rely on AAC. However, results of tests with end-users for either of the systems were not reported.

The first compansion system tested with persons who rely on AAC can be found in Vaillant (1997). Valliant presents a language expansion prototype with 300 pictograms for French that achieved an 80,5% of well-expanded sentences and a 73,5% of perfectly expanded sentences in technical tests. These results proved to be insufficient as tests with end-users showed that the 20% error rate and the lack of vocabulary made them feel frustrated (Vaillant, 1997).

Other research on compansion techniques that did not undergo tests with end-users can be found for Greek in Karberis and Kouroupetroglou (2002) and for English in Waller and Jack (2002). The first describes a prototype with a syntactic oriented constraint grammar, in other words, a simplified grammar that assumes that input sentences would reflect the basic word order of the desired output. Waller and Jack also make this assumption, but their prototype uses a statistical approach based on consecutive words, instead of a grammar driven approach like the previous systems, to translate Bliss symbols into English. While this approach allows for a broader set of vocabulary and more flexibility in allowed sentence structures, as the aforementioned works focused in simple sentences rather than on complex sentences (e.g. subordinate constructions), preliminary results show that translations are slower and have more probability of being inaccurate.

Another approach to achieve natural language in AAC are cogeneration systems (e.g. Lee et al., 2003 and 2006; P. García et al., 2015) where the person selects a certain sentence template (e.g. question, imperative, etc.) and then has to fill the slots that the interface provides in order to build the whole sentence. A hybrid approach that uses a cogeneration technique, but that as templates uses the semantic case frames of the input verb, which is selected first, is the RSVP-IconChat (Wiegand & Patel, 2014). The structure of case frames is also used in the user interface instead of grid-like boards of pictograms. This prototype, with an icon set of 106 items specifically chosen for a controlled environment, underwent successful tests with persons who rely on AAC without linguistic impairments. The main advantage of cogeneration systems over compansion systems is that they do not face the ambiguity of the lack of information in telegraphic language, but, on the other hand, they are not suited to many persons with complex communication needs for their dynamism and for the required linguistic knowledge to use them.

This approach is the one used in Avaz[11], a new AAC app aimed mainly at children with ASD.

Finally, another line to achieve natural language communication in AAC, although not related to pictogram-based communication, is conversational modeling, which can be applied to a multi-lingual translating AAC system (Alm et al. 2002), but that also requires high linguistic competence by the user.

To sum up, there exists little applied scientific research on systems that can expand the telegraphic language that comes from the use of pictograms into natural language. To the best of our knowledge, there is no successful reported research in the field with persons who rely on AAC on a daily basis and with ranging degrees of linguistic competence.

Concerning prediction, word prediction is a field that in recent years has been thoroughly researched paralleling the rise of text messaging and this research has also translated in text prediction for AAC communicators. However, iconographic or pictogram prediction have not been deeply explored (Garay-Vitoria & Abascal, 2004).

Moreover, despite saving keystrokes, the way in which predicted words are displayed, sometimes cause more mental effort for AAC users and it can make them to become tired faster. That is why, apart from prediction, the way of presenting the information also needs to be taken into account. In non-AAC literature, even if several authors have tested systems with a large number of proposals, no more than five to seven proposals are usually offered, which is the estimated number a user can perceive at a glance. This is even more relevant when users can have severe motor impairments (Taylor et al., 2001).

As far as text prediction techniques go, the most common word prediction methods are prediction using frequencies, prediction using word probability tables, which can be extended to support n-grams, syntactic prediction using probability tables, syntactic prediction using grammars and semantic prediction (Garay & Abascal,

[11] Avaz AAC app - http://www.avazapp.com/ [Last accessed, 24th October 2016]

2006). All these methods can be combined, along with machine learning techniques that can extract context knowledge from the user, to create more complex predictors. Examples of such context aware predictors that have been applied to pictogram prediction can be found in Kim et al. (2009) and in L.F. Garcia et al. (2016). The first one describes a pictographic card prediction system that defines the context as a combination of location and goal from the user. However, this information is not gathered dynamically and it's the user that has to provide it. Then this context is used to set the initial structure and the relations of the pictographic cards for communication, afterwards following a conversation modelling approach. On the other hand, L.F Garcia et al. (2016) describes a location aware pictogram and pictogram sentence prediction prototype that detects location using GPS capabilities. Moreover, as far as we know, it is important to note that this is the first study on pictogram prediction with reported performance results. Their research tests done with software simulations compare the location aware model with an all-around one that does not use location information. Results show that the location approach only outperforms the all-around approach if pictogram users reuse more than 50% of their sentences.

Finally, to conclude with prediction approaches, there are hybrid systems, as in Eunsil Lee et al. (2003 and 2006) and P. García et al. (2015), which combine cogeneration with pictogram prediction to show in the user interface only the vocabulary that can fill the different slots required at each moment by the cogeneration approach.

In this doctoral dissertation, we took a similar approach. Our designed prediction system takes advantage of the information encoded in the expansion system. However, in our case, it is not a cogeneration system, but a compansion system and the encoded data that the prediction system feeds on is mainly semantic instead of syntactic.

3.2. Dependency parsing review

As we said at the beginning of this chapter, as the compansion system that we have developed is based on the concept of dependency grammar, we thought that it was necessary to add a small literature review on dependency parsing to this dissertation.

Dependency parsing is the computational implementation of syntactic analysis using dependency representations, which represent relations between words. In dependency parsing, as opposed to parsers based on constituency analysis, the relation between the theoretical frameworks and the computational implementation can be quite shallow. This is probably due to the low degree of formalization of dependency grammar theories.

Also, when talking about dependency parsing, usually two different approaches are distinguished: grammar-driven parsing and data-driven parsing. Due to the high difficulty of obtaining enough data, taking into account the limitations and the low communication rates of pictogram-based AAC that we will discuss in the following chapters of this thesis, to train a data-driven parser, which is the method mostly used nowadays and with most recent literature about it, in this review, we will only talk about grammar-driven parsing, which is the approach that we have taken.

The first works on dependency parsing were done by Hays (1964) and Gaifman (1965). Their approaches were close to the formal theories of dependency grammar. Most of the basic notions of dependency grammar (Nivre, 2005) were present, such as the single-head constraint, the representation resulting in a rooted tree and the applicability of the projectivity constraint, among others. These first parsers did not use any dependency types to classify dependency relations, thus dependencies remained unlabelled.

The results that Gaifman (1965) obtained, which closely related his dependency system to context-free grammars, discouraged the further study of dependency grammar for parsing. However, Järvinen and Tapanainen (1998) disclosed that the conclusion that dependency grammar is only a small variant of context-free grammar reached by Gaifman was erroneous.

Following the framework that Gaifman and Hayes first proposed for dependency parsing, the algorithms that appeared used dynamic programming in the same way that parsing algorithms for context-free grammars did. Most of the frameworks that later appeared also implemented the notion of projectivity, except for some parsers that introduced non-projective structures after a post-processing step (Sleator & Temperley, 1991). Most of these frameworks can also be included under the concept of bilexical grammar introduced by Eisner (2000). The basic parsing algorithm proposed by Eisner also uses dynamic programming.

The second principal tradition in grammar-driven dependency parsing is based on the method of eliminative parsing. In this method, sentences are analysed by sequentially eliminating representations that do not fit constrains until only a valid representation remains. As we can see, this notion is closely related to the notion of constraint grammar.

The first dependency parsers using this idea came from the Constraint Grammar framework (Karlsson, 1990). Afterwards, Maruyama (1990) extended the idea using a system that tagged dependencies with both a syntactic label and an identifier for the head node. This type of representation for dependencies is fundamental for many approaches to dependency parsing, as it reduces the parsing problem to a problem of classification or tagging.

In the eliminative system, parsing is a constraint satisfaction problem, where a good analysis is the one that does not break any of the constraints of the grammar. In general, this type of constrain satisfaction problem is NP complete. In this type of problems a given solution can be quickly verified, but, on the other hand, there is not a known efficient way to find a solution. The time to solve these problems rapidly increases as the size of the problem grows, thus, when implementing algorithms that solve this kind of problems, you need to be cautious, so that it is fast enough and does not take too much time.

In more recent approaches, the TDG framework (Duchier 2003) faces this problem using constraint programming, which ensures the finding of a solution to the parsing of a given input in a reasonable computing time. The TGD framework also introduces several levels of representation in dependency parsing (notion that we already saw), exposing that a single constrain can point to different levels (i.e. syntactic and semantic) at the same time (Duchier & Debusmann, 2001). This view is also extended in the Extensible Dependency Grammar framework (XDG)

(Debusmann et al., 2004) where many levels or dimensions can be defined in the grammar.

The two main problems that the parsing frameworks that we have seen up until now are: first, there might not be an analysis, for a given input, that fits all the constraints in the grammar and, second, there might be more than one analysis that fits all the constrains, in other words, there can be a disambiguation problems. To face both issues, the notion of weighted constraints appeared (Menzel & Schröder, 1998). In this approach, a weight that indicates how serious is the violation of a given constraint, is assigned to each constraint. Therefore, the best analysis is the one for which the sum of violated constraints is minimized.

Finally, there is a third grammar-driven parsing tradition that combines dependency grammar with a deterministic parsing strategy. The basic strategy of this tradition is to accept the words one by one, starting at the beginning of the sentence, and to try linking each of these words as head or dependent of every previous word (Covington, 2001).

This strategy is compatible with many different types of dependency grammar. The only thing that is required is for the grammar to define a function that for any two words it returns a Boolean, true if the first word can be the head of the second word or false otherwise. Covington (2001) demonstrates that this method can be used to obtain dependency representations that satisfy conditions like uniqueness (having a single head for each node) and projectivity by adding pertinent constraints on the linking process. Covington (1990a, 1990b, 1994) also showed that this method can be adapted to languages with different types of word order, be it free, flexible or rigid.

To conclude with grammar-driven dependency parsing, we have seen three different traditions. The first one is more based on the formalization of dependency grammar theories that are mostly restricted to projective dependency representations. The first tradition uses dynamic programming algorithms to implement the parsers. The second one is based on the formalization of constrains, which do not need to be restricted to projective structures, and the parsing method uses a successive eliminative approach. Finally, the third one is based on the combination of dependency grammar with deterministic parsing strategies and the parsing method is the one described by Covington (2001). This last strategy is the one that is most similar to the one that we have taken for the parser of the companion system, as we will now see in the following chapter.

CHAPTER 4

COMPANSION SYSTEM

This chapter will discuss the considerations that we took into account in order to design and develop the compansion system and it will also describe its parsing and generating algorithms.

4.1. Design considerations

Apart from documenting ourselves through research papers, books and other literature on AAC, at the beginning of the thesis we also met with three associations for disabled people[12] that use AAC strategies and devices and with the person in charge of accessibility for disabled students of the State Department of Education of Catalonia (Departament d'Educació de la Generalitat de Catalunya). Later on, we visited more special education schools and the Unit of Augmentative Communication Techniques (UTAC in Catalan), an external service offered by the University of Barcelona together with the Catalan Government. With their help and advice, we better learned how pictogram-based communication worked and, with the experience acquired, we were able to determine the needs of the final application, to limit the types of sentence structures that the compansion system would accept and to define the methodology to follow in order to build it.

Nevertheless, even though we would be using a controlled grammar, instead of accepting all possible sentence constructions in Catalan and Spanish, we

[12] Centre d'Educació Especial Pont del Dragó http://www.bcn.es/pontdeldrago/ca/index.html & Prodiscapacitats Fundació Privada de Terrassa. http://www.prodis.cat/ & Associació Pro-Disminuïts Físics i Psíquics de Sant Cugat del Vallès. http://asdi.cat/ & Unitat de Tècniques Augmentatives de Comunicació http://www.utac.cat/ [May 20th, 2017]

decided to build a system capable of handling more linguistic complexity than the usually required by persons that use pictogram-based AAC. The decision was made thinking of persons that might have reading and writing potential, but that did not have the proper conditions to develop them. Thus, the system could help them better communicate increasing their linguistic skills. Furthermore, as the compansion system would be adaptable to the linguistic competence of each person by removing or adding vocabulary and available gender, number, verb tense or sentence modifiers, as we will later discuss, it was better to err on the side of excess in order to avoid frustration and later abandoning of the system, as seen in Vaillant (1997). The importance of making compansion systems flexible to the linguistic competence of each person is stated in Pennington and McCoy (1998).

Before discussing the methodology followed to build the compansion system, we would like to better define some elements that influence it.

In short, our research has to do with the transformation of telegraphic language[13] used in pictogram-based AAC into natural language utterances in Catalan and Spanish. The telegraphic language that will need to be expanded will result from the input of pictograms, which only contain drawings for meaningful words (in other words: verbs, nouns, adjectives, a small set of adverbs, possessive determiners, quantifiers and some set expressions, plus sentence modifiers like tense modifiers, order or permission modifiers, word modifiers (e.g. the plural modifier, etc.).

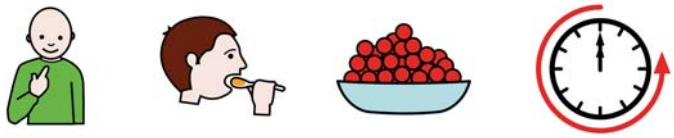
The main characteristics of this telegraphic language, apart from only containing content words (and some modifiers), as we have just seen, are the following:

- Words can appear in any order. Thus, it is a free word order language.
- There can also be reduction of some content words (like the subject of the sentence or even the main verb of the sentence).

[13] Telegrpahic language is characterised by containing only meaningful words (Karberis and Kouroupetroglou, 2002)

These characteristics are what make translating from pictograms into natural language more difficult than translating between two natural languages. The fact that there is no set word order, the absence of words like prepositions that mark the function of other words and the reduction of semantic roles, like the subject or even the main verb of a sentences, implies that the system has to face ambiguity and, thus, has to infere all this missing information.

An example input could be:



CAT	Jo	Menjar	Molt	Passat (mod)
ES	Yo	Comer	Muy/mucho	Pasado (mod)
ENG	*I*	*Eat*	*A lot*	*Past (mod)*

Figure 1: Example of input with a tense modifier.

The input in Figure 1 would expand to the sentences “Vaig menjar molt” o “Comí mucho” (I ate a lot). Therefore, as just seen, the first part of the compansion system needs to be an input system.

After the input system, the core of the compansion system comes into play. First of all, there is the parser, which takes the input words from the input system, which are in the form of telegraphic language, and tries to assign them their correct function in the sentence. Then, the generator takes the parse tree built by the parser, plus the sentence modifiers that might have been input as well, and, using a generation model, expands the sentence transforming it into a natural language sentence (Figure 2).

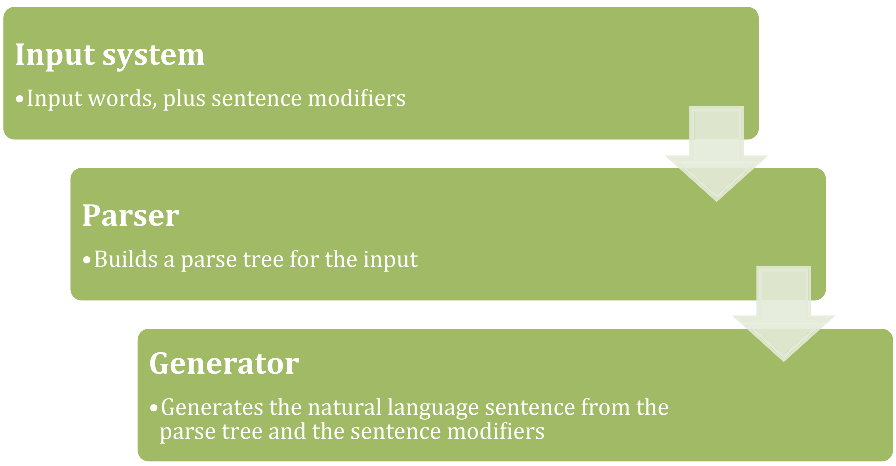


Figure 2: Components of the compansion system.

In order to expand the telegraphic language into natural language, the parser uses a semantic approach similar to the one that can be found in Pennington and McCoy (1998). The basic idea is that the verb is the centrepiece of the sentence, which perfectly adapts to dependency grammar where the verb is the root, and the rest of the complements of the sentence that need to be filled are the semantic roles (or slots) that that verb accepts. The semantic roles accepted by each verb of the lexicon are described in a set of patterns. The patterns also have default values that are used in case there is a reduction of one of the mandatory slots for a given verb (Figure 3).

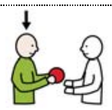
		
CAT	Donar	poma
ES	Dar	manzana
ENG	*To give*	*apple*
CAT	Dóna'm una poma, si us plau.	
ES	Dame una manzana, por favor.	
ENG	*Could you give me an apple, please?*	

Figure 3: Example of a sentence with an order modifier using telegraphic language. In this sentence there is a reduction of the subject and the default subject for an order is used instead.

4.1.1. Controlled grammar

Also, as previously mentioned, trying to accept all the constructions allowed in Catalan and Spanish grammars would increase the amount of possible relations (i.e. dependencies) between pictograms. This would increase the ambiguity faced by the parser in terms of choosing which pictogram potentially fulfills each semantic role (e.g. which pictogram acts as subject, which acts as main verb, which potentially acts as secondary verb, if any acts as a noun complement, etc.). Therefore, to avoid a higher error rate by the parser, we decided to narrow down the target constructions using a controlled grammar, as it is commonly done when developing dependency parsers. Here are the main restrictions of the system, along with some examples in Catalan (grammar constraints in our system are the same for all languages):

1. The grammar only allows one type of subordinate clause in sentences. It only allows for verbs that directly depend on other verbs and does not allow any other type of subordinate clauses, like relative clauses or

subordinated clauses leaded by a subordinate conjunction (Table 1). In our system, in order to make this kind of sentences, the sentence needs to be split in two.

a. In turn, a verb that already depends on another verb cannot be the head of another verb. Therefore, the system only allows for a maximum of two verbs in a sentence (unless sentence modifiers, like the Desire or Permission modifiers, are used).

Example in Catalan	**Example in English**	**Is it allowed?**
Estic cansat de jugar.	I'm tired of playing.	Yes.
Ajuda'm a baixar les escales.	Help me go down the stairs.	Yes.
M'agrada cantar.	I like singing.	Yes.
Espero que vinguis a veure la pel·lícula.	I hope that you come to watch the movie.	No (there are three verbs in the sentence).
Vull anar a comprar.	I want to go shopping.	Yes (if the *Desire* modifier is used).
L'home, que passeja el gos, és simpatic.	The man that walks the dog is very nice.	No (there is a relative clause. It should be split in: 1- The man walks the dog. 2- The man is very nice).
Era intel·ligent, però va cometre un error.	He was smart, but made a mistake.	No (there is a subordinate conjunction. It should be spit in: 1- He was smart. 2- But made a mistake).

Table 1: Examples of accepted and not accepted sentences by the parser.

2. Apart from subordination, the system does not accept coordination of sentences either.

   a. Nevertheless, the parser accepts coordination between two or more nouns or between two or more adjectives (Table 2).

Example in Catalan	**Example in English**	**Is it allowed?**
La Maria és llesta i alegre.	Mary is smart and cheerful.	Yes.
L'home i la dona fan	The man and the woman	Yes.

pastissos.	make cakes.	
Ell canta i balla.	He sings and dances.	No (it should be split in: 1- He sings; 2- (And) He dances.)

Table 2: Examples of accepted and not accepted sentences regarding coordination by the parser

3. Nouns that are the head of a slot (e.g. the head of the subject or the head of the theme in the sentence) can have the following complements or modifiers (Table 3):
    a. Another noun, just one (or two or more coordinated nouns).
        - A noun that is complementing another noun cannot have any more complements (but can be modified by for adjectives).
    b. An adjective, just one (or two or more coordinated adjectives).
    c. Quantifiers. There are the following quantifiers in the system: “molt”, “poc”, “més”, “menys” (meaning *very/much*, *a few/little*, *more* and *less*). Two quantifiers can appear together: “molt més alt” (*much taller*).
    d. Numerals[14].
    e. A possessive.
    f. A locative adverb (if the noun is the head of a locative slot).
    g. All the previous at the same time (except for numerals and possessives and except quantifiers, numerals or possessives that cannot go with adverbs).
4. Adjectives, Adverbs and Quantifiers can only have quantifiers as complements (any number of them).
5. Verbs can have sentence modifiers, which are described two pages below, such as the negative or the imperative modifiers.
6. Pronouns cannot have any complements.
7. Expressions are thought to be the only element in the input, as they alone constitute a sentence on its own. Only a few expressions, like “Si us plau” (*Please*), can appear along with other words in the input.

[14] Although they are also modifiers, as they are treated differently by the parser, that is why we put them in two separate categories.

Example in Catalan	**Example in English**	**Is it allowed?**
L'home de ferro.	The man of steel.	Yes.
El ninot de neu i gel.	The doll of snow (snowman) and ice.	Yes.
Moltes més pomes.	Many more apples.	Yes.
La casa blanca de pedra.	The white house of stone.	Yes.
La casa de pedra blanca.	The house of white stone.	Yes.
Tres dones altes.	Three tall women.	Yes.
La meva germana més alta.	My more tall (tallest) sister.	Yes.
A sota la taula.	Under the table.	Yes.
No vinguis.	Don't come.	Yes.
Els meus tres germans.	My three brothers.	No (there is a possessive and a numeral).
A sota la meva taula.	Under my table.	No (there is a locative adverb and a numeral).

Table 3: Examples of accepted complements for nouns, verbs and adjectives.

Initially we thought that these constraints were not a big limitation to the system as most sentences built using AAC devices have very few words, so complex constructions are extremely rare and are usually split into simpler sentences. As we will see in Chapters 7 and 8, this was later confirmed in the tests conducted with persons who rely on AAC.

Furthermore, telegraphic language, even using simpler constructions, can sometimes have several interpretations due to lack of non-content words. These non-content words, such as prepositions or conjunctions, can be important to determine the function of each word in a sentence. Therefore, to address the lack of these words, in order to further narrow the options and in order to allow users to build more complex sentences, we decided to add the following optional modifiers to telegraphic communication. Among these modifiers, there are both pictogram or sentence modifiers:

1. Gender and number modifiers: they can be applied to nouns and adjectives to turn them into their feminine and plural forms. For example, transform *gat/gato* (in English, *cat*), which in Catalan and Spanish has both feminine and plural forms, into *gata/gata* [feminine], *gats/gatos* [masculine-plural] or *gates/gatas* [feminine-plural].
2. Sentence-type modifiers: there are a total of 9 and can be applied to sentences to turn them into questions, desires, imperative sentences, answers, negative sentences, asking for permission, conditional statements, declarative or exclamatory sentences. For example, if you add a permission modifier to an input like *I-GO-TOILET*, which is usually by default generated in its declarative form "*I go to the toilet*", it is transformed into "*Can I go to the toilet, please?*". The same sentence with the imperative modifier would be "*Could you go to the toilet, please?*".
3. Verb tense modifiers: apart from time adverbs, like *yesterday* and *tomorrow*, that automatically change the verb tense of a sentence, verb tense modifiers can be applied manually to change the verb tense of the generated sentence to produce sentences in past and future tenses. In total, our system has got three past tenses, a present tense and a future tense.
4. Conjunction *and*: needs to be specified in order to coordinate nouns. For example, to say *the bar and the beach*; otherwise, the system would generate *the bar at the beach*.

Other common characteristics of telegraphic language that our system can solve without the use of modifiers are the following:

1. Verb subject reduction: When the verb subject is not present in the input pictograms, the system, in order to generate a natural language sentence, chooses a default subject. For example, if the input pictograms were *EAT-APPLE*, the system would choose *I* as the default subject and produce "*I eat an apple*". This default subject can also depend on the type of sentence, for instance, for questions the default subject is the second person singular *you*.
2. Verb reduction: For some nouns and for all adjectives, if not found among the input pictograms, the system automatically puts a verb which is

associated to them. For example, for adjectives the copulative verb *to be* is used. In other words, if the input pictograms were *DOG-HAPPY*, the generated sentence would be "*The dog is happy*".

In turn, the vocabulary for the system is also constrained, as the vocabulary used is based on a vocabulary selection specially designed for basic AAC communication by the University of Barcelona, named CACE[15].

It is also important to mention that the compansion system assumes that the input will have a correct parse, in other words, that the words that the user entered intend to build a correct, both syntactically and semantically, sentence. As a result, the system does not check for possible input errors in the sentence.

Having seen the basics of the system, lets start defining all its elements in detail.

4.1.2. Vocabulary

All the vocabulary in the system needs to have related semantic and syntactic information annotated. Therefore, having a large corpus of annotated vocabulary would have required a huge amount of time to build. We checked if there were similar corpuses already built or if we could reuse corpuses built for other purposes, like WordNet[16], but as the information that we required in order to expand the telegraphic language into natural language was very specific, we could not automathize the process of building the corpus. Obviously, for some of the characteristics annotated for each word, which we will later see, we used different kinds of dictionaries, but most of the annotation had to be done manually (except for verb tenses).

In consequence, our first idea was to expand general greetings, such as introducing oneself, and another daily activity such as shopping or going to a restaurant. In order to select one of these daily life activities, we had several meetings with associations for disabled people around Barcelona that used

[15] UTAC-CACE. http://www.utac.cat/noticies/utac-cacejadisponibleperadescarregarenquatreversions [May 19th, 2017]

[16] WordNet. https://wordnet.princeton.edu/ [May 19th, 2017]

different kinds of AAC systems. We realised that the interest for each activity highly depends on the preferences of the person and their daily routine, which can widely vary depending on their disability. Some other important external factors should also be taken into account. For example, people with complex communication needs (CCN) are at high risk of suffering some kind of abuse, victimisation or crime. Consequently, AAC systems ought to have the vocabulary needed to face these situations.

As selecting an adequate vocabulary in a proper way would need a detailed study of word and sentence usage of people with CCN in different contexts, we decided to follow the recommendations from the associations that we met and to use an already available set of specific vocabulary for Catalan and Spanish AAC users named CACE (Soro et al., 2007). In fact, CACE is a set of ready-to-use boards that conform a pictographic communicator in Catalan and Spanish, which can be integrated into several types of AAC software for computers (The Grid 2, Boardmaker, which in turn integrates with Speaking Dynamically, Plaphoons and SAW[17]).

The most important characteristics of the pictographic communicator CACE, built by UTAC (Unitat de Tècniques Augmentatives de Comunicació), a specialized group in AAC techniques from the University of Barcelona, are the following:

- CACE includes a proposal of basic vocabulary[18] (853 vocabulary items and 132 phrases) that needs to be customised for each user.
- It is organised in categories, 20 in total, in several pages and navigable boards.
- It also includes morphological modifiers (tense modifiers, possessives, etc.).
- It can be extended, while maintaining its general structure.

[17] Special Access to Windows. http://informaticaparaeducacionespecial.blogspot.com/2007/07/programa-saw-o-cmo-emular-un-teclado.html [May 19th, 2017]

[18] List of vocabulary of CACE in Catalan. http://www.utac.cat/descarregues/cace-utac/L%C3%A8xicCACE.pdf [May 19th, 2017]

First, we included all the vocabulary in CACE, except for set expressions, which users can add themselves through the user interface, into our compansion system. Later on, in order to conduct the tests with persons who rely on AAC, we expanded this initial core vocabulary to fit the needs of the participants of the tests. In Table 4 there is a summary of the amount of vocabulary that we had initially encoded into the system and the amount of vocabulary that we later added for the end users tests. More details on the added vocabulary will be given on Section 7.3.2.

Type	Initial core vocabulary	Core vocabulary during the tests
Verbs	88	147
Patterns	119	180
Nouns (including pronouns)	571	704
Adjectives	95	129
Adverbs	21	27
Set expressions	39	59
Question words	11	12
Modifiers (possessives, quantifiers, etc.)	15	19
Sentence modifiers (sentence type, tense, etc.)	16	17
Total pictograms	**856**	**1114**

Table 4: Amount of vocabulary and patterns in the compansion system before the tests with end users and after the tests.

For each word in the vocabulary, it is necessary to annotate a certain set of features that vary depending on the type of word (for a comprehensive list of annotated features see Digital Appendix Z.1.1; all digital appendixes can be downloaded from https://goo.gl/ZfQpGk). For example, for each verb, several patterns representing each of the most common senses or usages of the verb in Catalan and Spanish have to be described. Each pattern tells which semantic roles are mandatory for that verb sense, which are optional, which type of word and class within that type of word is the ideal head for the slots, default values for mandatory slots, etc. More information on verb patterns can be found in the description of the parser below or in Digital Appendix O, where a sample of the patterns of the system in Catalan can be reviewed. For other words, such as nouns, information on gender, number, the articles that usually precede it or even verbs that commonly go with that noun need to be encoded.

Even if at 1100 items the set of vocabulary encoded might seem small (all the annotated words available in the system can be found in Digital Appendix P), as the system has been designed and built in order for it to be easily scalable (regarding the vocabulary available), it is not an issue in terms of testing the system to see its accuracy as it already represents the full potential and functionality of the system.

4.2. Input system and process of constructing a sentence

Now that we have seen the main decisions taken regarding the initial vocabulary for the system, we can continue to the first element that constitutes the compansion system, the input system. In order to be able to test the sentence generator, the initial input system did not use pictograms, but the words they referred to instead. This initial input system (Figure 4) was the one used in the technical tests of the compansion system that are presented later on in this chapter.

In order to send an input that the compansion system can generate, these steps need to be followed:

1. Think of the sentence that you want to generate.
2. Choose the content words for the sentence (and select the pictograms).
    a. Apply feminine, plural or coordination modifiers to these words, if needs be.
3. Choose the desired verb tense from the following verb tense modifiers (or leave the default value):
    a. Default: If a verb tense is not selected, the generator will decide the best tense for the sentence depending on the type of sentence chosen, the verbs selected and the time expressions inputted, if any.

b. Present: This forces the main verb in the sentence to be in present tense.
c. Past: This forces the main verb to be in past tense.
d. Immediate past: The main verb will be in a tense similar to the present perfect tense.
e. Distant past: The main verb will be in a tense that is used both in Catalan and Spanish for actions that used to take place a long time ago.
f. Future: The main verb will be in the future tense.

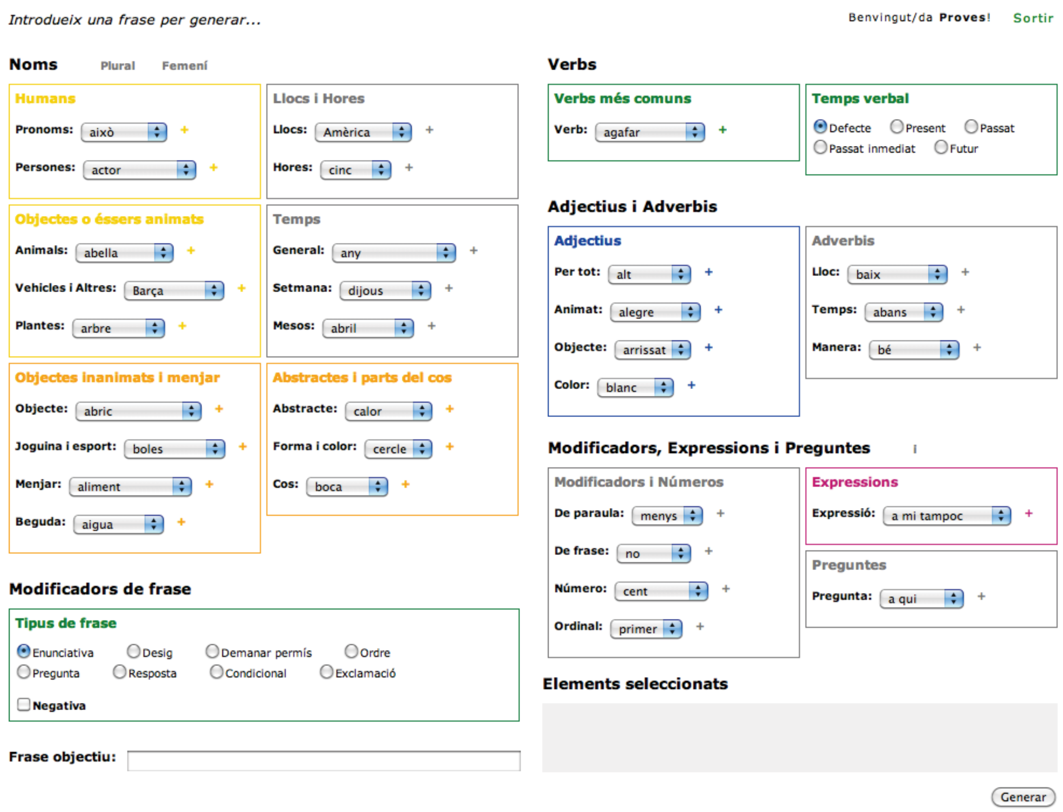


Figure 4: Screenshot of the alternative input system. Lists of words are grouped into different categories using a common colour legend19 for pictogram-based AAC.

4. Choose the desired type of sentence from the following sentence type modifiers: default, declarative, desire, ask permission, order, question, answer, conditional or exclamatory (see Table 5 for a comprehensive description of each modifier).

[19] Yellow for nouns that are humans or human pronouns; orange for other types of nouns, such as objects and locations; green for verbs; and blue for adjectives; colours assigned to adverbs, modifiers and other types of words tend to vary depending on each individual.

5. Send the sentence to the parser.

Sentence Type	Description
Default	If it's not specified differently on the verb pattern that the parser selects to generate the sentence, it's a declarative sentence.
Declarative	Used to override a different default behaviour than usual in order to turn the sentence into a declarative sentence.
Desire	For sentences like "I want…", the system will add automatically "Vull" (I want) at the beginning of the sentence. When using this modifier, sentences with three verbs can be built. E.g. "Vull anar a comprar" (I want to go shopping).
Ask permission	For sentences like "Can I…", the system will add automatically "Puc" (I want) at the beginning of the sentence. When using this modifier, sentences with three verbs can also be built. E.g. "Puc anar a comprar?" (Can I go shopping?).
Order	If selected, sentences will use the imperative tense.
Question	If a question particle has already been selected, the sentence is already a question and the type of sentence selected becomes overwritten.
Answer	It is mainly for sentences without a verb ("verbless"). If this option is selected, the system will give priority to patterns without a verb to produce sentences like "An apple, please".
Conditional	The generator will add the "si" (if) particle at the beginning of the sentence.
Exclamatory	The generator will add the exclamation mark (!) at the end of the sentence.
Negative	Can be selected along with any other type of sentence. Transforms the sentence into its negative form.

Table 5: Comprehensive list of the sentence type modifiers in the compansion system

Keep in mind that, apart from the previous steps, the system expects the user to try to build a sentence that makes sense and to take into account the restrictions in the sentence structure that we have detailed in the controlled grammar section (Section 4.1.1).

4.3. Parser

The aim of our system is to transform telegraphic language that will eventually result from selecting pictograms into well-constructed natural language sentences. This is known as cogeneration (Copestake, 1997). The job of a cogeneration system is to order text units from an input, add inflections and insert extra words (both function and content words, as in Figure 5). In fact, our system would not exactly do cogeneration, but compansion (Pennington & McCoy, 1998) (compression-expansion).

Compansion, unlike cogeneration, only expands uninflected content words (compressed, synthetic or telegraphic utterances) into syntactically and semantically well-formed sentences.

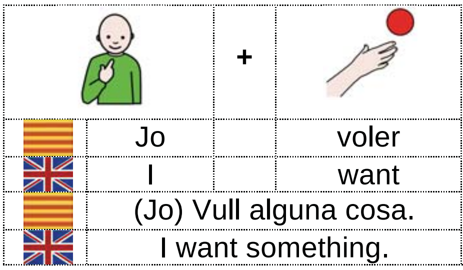


Figure 5: Example of the insertion of a generic expression.

Usually, in cogeneration systems, the user selects a certain sentence template (question, request, order, etc.) and then has to fill the different slots that the interface provides in order to build the whole sentence. In our system the templates will be selected dynamically on-the-fly depending on the previous words selected by the user (mainly depending on the verbs selected, but also on sentence modifiers or question particles). Also, the templates or patterns will be invisible by the end user and, if a certain verb selected can have several patterns associated, the system, in this case the parser, will select the best one according to the rest of the words in the input.

Following the approach of Pennington and McCoy (1998) on compansion, the core of our compansion system is a semantic parser that interprets input based on the use of case frames or slots. These slots are conceptual structures that represent the meaning of a sentence by describing the semantic cases or roles that each of the content words has in relationship with the others. When running, the semantic parser designates the verb as the main component of the sentence: all other words in the input are used to fill semantic roles with respect to the main verb that is chosen and the patterns annotated for that verb.

For each semantic role (e.g. subject, theme, receiver, beneficiary, location, etc.) the system has encoded its priority (whether it's optional, mandatory or not applicable), the type of word and class that better fits it and the preposition that usually precedes it, among others. For each verb pattern, there is also general information encoded, such as if the verb is pronominal[20], pseudoimpersonal[21] or copulative, its default subject, if it is not found in the input, or the default type of sentence and verb tense that will be applied to that pattern. Finally, there is also a feature that marks if any of the semantic roles can be fit by another verb, in other words, it says if the pattern accepts another verb as a complement or not.

Donar / Dar (*to give*)

Pronominal	0
Pseudoimpersonal	0
Copulatiu	0
Tipusfrase	ordre (*order*)
Defaultense	imperatiu (*imperative*)
Subj	human
Subjdef	2
Theme	1
Themetipus	noun
Themedef	ho / lo (pronouns meaning *this*)
Themeprep	
Themeart	
Receiver	1
Receiverdef	mi / mí (*me*)
Receiverprep	a / a (*to*)
Benef	0
Beneftipus	
Benefdef	

[20] Pronominal verbs have the property that the receiver is always the same as the subject. In Catalan and Spanish, these verbs have a pronoun that agrees with the subject and that always goes attached to the verb, either before or after the verb (depending on the verb form), e.g.: “Amagar-se” / “Esconderse” (*to hide oneself*), e.g.: “M’amago a l’armari” / “Me escondo en el armario” (*I hide in the closet*).

[21] In “pseudo-impersonal” verbs, subjects and receivers are flipped. An example of it would be the verb “agradar” / “gustar” (*to like*): “M’agraden les pomes” / “Me gustan las manzanas” (*I like apples*). Here, in Catalan and Spanish, the subject is “pomes” / “manzanas” (*apples*) and it usually goes after the verb, although: “Les pomes m’agraden” / “Las manzanas me gustan” would also be correct.

Benefprep	
Acomp	0
Acompdef	
Acompprep	
Tool	opt
Tooldef	
Toolprep	amb / con (*with*)
Manera	opt
Maneradef	
Maneratipus	adv
Locto	0
Loctotipus	
Loctodef	
Loctoprep	
Locfrom	0
Locfromtipus	
Locfromdef	
Locfromprep	
Locat	opt
Locatdef	
Locatprep	a / en (*at*)
Time	opt
Expressio	si us plau / por favor (*please*)
Subverb	0

Table 6: Example of a pattern of the verb "to give".

In Table 6, we can see an example pattern of the verb "Donar" / "Dar" (*To give*). In this case, we can see that the default subject is the second person singular. We think that it is more reasonable in the context of use of the AAC device and in normal life contexts as well for the subject or agent of the verb "to give" to be the second person singular "you" and for the verb to be in an imperative form (see Figure 6). Therefore, if the user wanted to use the first person singular, he/she should explicitly use the pronoun "jo" / "yo" (*I*) in the input. Nevertheless, in most verbs in our system, we considered, taking into account the most common use of language, the default subject to be the first person singular.

When defining a pattern, if we needed a type of semantic role that is not in the features list, we could always use the features from an unused slot, which does not have special syntactic properties (i.e. the subject, which can be reduced and which usually appears before the verb, or the receiver, that can be pronominalized), that would act as the desired slot instead. This is possible because in the system, we have prioritised functionality and performance over linguistic fidelity.

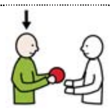
		
CAT	Donar	poma
ES	Dar	manzana
ENG	*To give*	*apple*
CAT	Dóna'm una poma, si us plau.	
ES	Dame una manzana, por favor.	
ENG	*Could you give me an apple, please?*	

Figure 6: Example of a special default subject and imperative verb form.

Also, when defining a pattern, it's important not to have two slots with the same values of type, as the system would not be able to solve the ambiguity of which word fills each slot. Only the subject role would be an exception as it can be potentially disambiguated using syntactic information, such as whether it appears before or after the verb (although word order in telegraphic language is not relevant, but it's a feature that can be used to solve ambiguity between two words that can act as a subject if the rest of the features are unable to do so).

4.3.1. General characteristics

Having seen the basics of the system and the structure of the verb patterns (a sample of the annotated patterns can be found in Digital Appendix O), we will now start defining the algorithm of the parser.

Initially, we considered using a statistical machine translation (SMT) (Koehn, 2009) approach. However, all SMT algorithms need to be trained with a large corpus of sentences in order to learn how to do translations automatically. In our case, to the best of our knowledge, there does not exist a corpus of sentences produced by persons who use pictogram-based AAC devices and, even less, a corpus with its corresponding counterpart in natural language, which would be necessary to build the parallel corpus[22] required to train the algorithm. Therefore, this approach was ruled out from the early stages of this thesis and we decided on a rule-based dependency parsing machine translation approach.

As we already explained at the beginning of this chapter, the parser uses a controlled or constraint (in the less technical sense of the term) grammar dependency parsing method. Controlled grammar because the compansion system does not accept all the existent constructions in Catalan and Spanish, but a restricted set of it. The approach that we use, as SMT was not possible, is a rule-based grammar-driven eliminative approach (Karlsson, 1990), but, instead of just using syntactic information for the constraints, we use a mix of syntactic language-dependent information, plus semantic features, annotated in the features for each word in the database, meaning that our system combines different levels of analysis, similar to the notions that can be found in Extensible Dependency Grammar (Debusmann et al., 2004).

The system also does labelled dependency parsing, as the dependencies from head to dependent are tagged using the semantic roles on the first level of the tree. Then, on the rest of the levels, the system uses tags such as “NC” for noun complement, “ADJ” for adjectives that act as complements, “ADV” for adverbs and “MOD” for other modifiers.

As we can see in Figure 7, the idea is to limit to two levels the dependency tree, to ensure simple sentence structures and to limit possible errors that would lead to wrongly expanded sentences. The two-level constraint does not take into account extra levels that can result from adding prepositions or articles and other determiners to nouns that appear on the second level during the generation. It does not take into account either a third extra level that can appear in sentences where a second verb acts as the nucleus of a thematic role and thus depends on the root verb.

---

[22] A parallel corpus is a text in one language paired with its translation into another language (Koehn, 2009). In our case, it would have been a corpus of sentences built by persons who rely on AAC using pictograms paired with a corpus with its translations in both Catalan and Spanish.

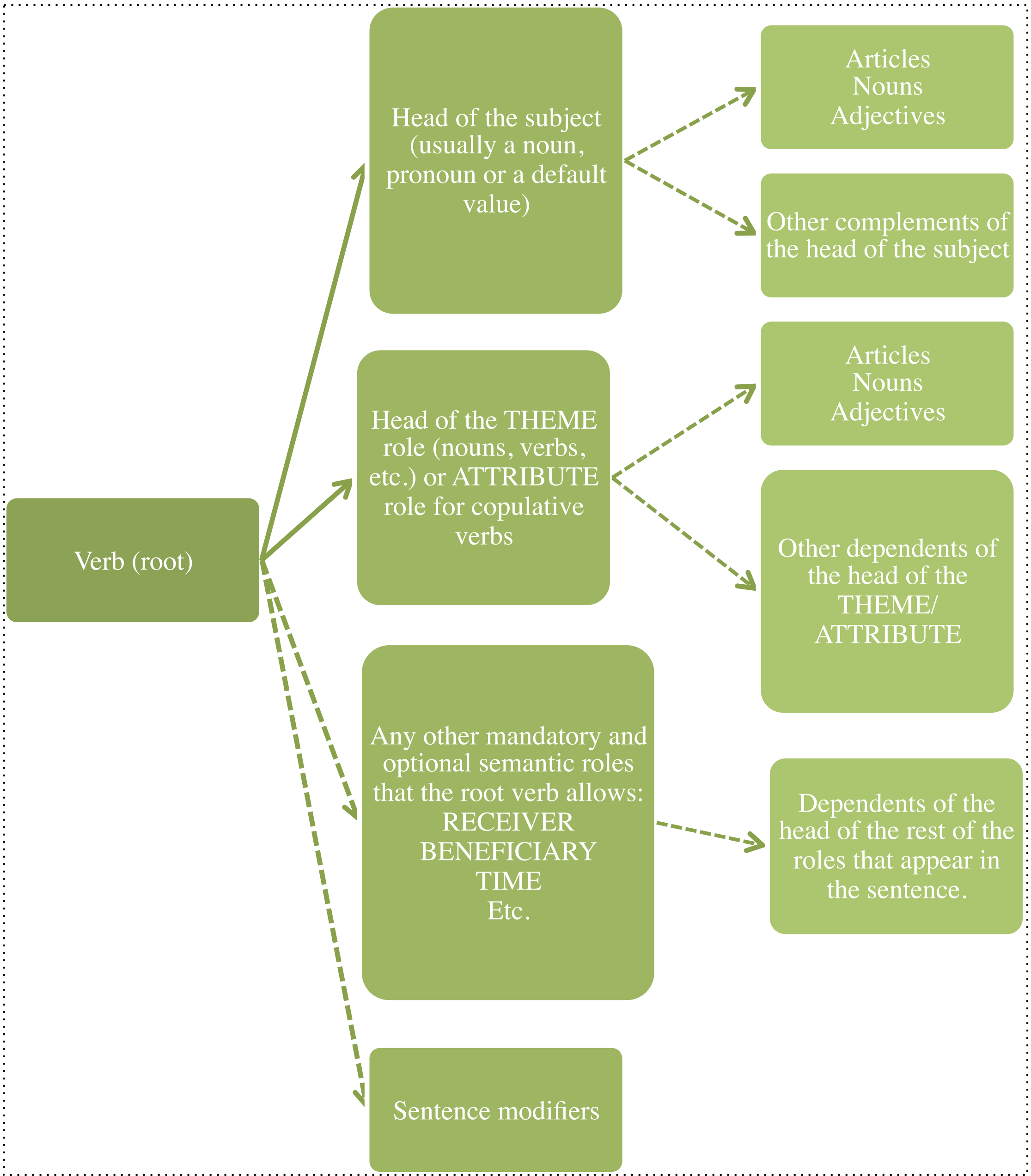


Figure 7: Example of a 2-level tree that represents a simple sentence structure. Labels have not been added to ensure legibility. Dotted dependencies can appear multiple times.

Also, the method to label dependencies is deterministic, similar to the one that Covington (2001) explains. In general, our system starts searching the possible dependencies of the root. Once it has all of them, it proceeds to the second level of the tree looking for possible dependencies, constrained by the grammatical categories of the heads of the first level and so on with the rest of the levels.

Finally, in case a word can depend on several heads or can fit several slots or semantic frames (Demasco and McCoy, 1992), in order to disambiguate, semantic features (Karberis & Kouroupetroglou, 2002) of the head and the words that can modify it (in the first case) or of the slot and the words that can fit it (in the second case) are used. After applying these features, if there are still ambiguities left, only then language-dependent word-order syntactic information is used. Also, to disambiguate different correctly parsed sentences that come from different possible patterns for the same verb, a weighting method is used.

To better illustrate how the algorithm solves ambiguities, here is an example. If the input were *BEACH-GO-I*, the algorithm knows that usually an animated being is the Subject that goes to a Location, thus the Subject is *I* and the Location is *beach*, not the other way around. However, if the input is *YOU-LOVE-I*, as both the Subject and the Receiver can be animated human beings, the word order will be used to solve the ambiguity and the algorithm would decide that *you* is the Subject and *I* is the Receiver, later transforming it to “You love me”.

When it is impossible to solve an ambiguity, be it by the parser or by the generator, the algorithm tries to choose the sentence that minimizes the changes in meaning the most or the sentence that preserves the ambiguity the most (e.g. choosing an indefinite article over a definite article in some cases).

Related to the language-dependent word-order syntactic information, in fact, the algorithm of the parser is designed to support languages with different relative word order in the structures of the sentences, such as Subject-Verb-Object (SVO) languages, like romance languages or English, SOV languages, like Japanese, or Name-Adjective (NAdj) languages, like Catalan and Spanish, or AdjN languages, like English. In order to do so, for each language present in the compansion system (at the moment just Catalan and Spanish, although tests have been conducted setting the parser for relative word-order in English), the following language features are annotated:

- Relative order for the subject, the verb and the object of the sentence (e.g. “SVO” or “SOV” languages, such as “I eat apples” in English versus “私はりんごを食べます” in Japanese).

- Relative word order for an adjective that complements a noun (e.g. “NAdj” versus “AdjN” languages, such as “casa blanca” in Spanish versus “white house” in English).
- Relative word order for a noun that complements another noun (e.g. “NNC” versus “NCN”, such as “la casa de pedra” in Catalan versus “the stone house” in English).

4.3.2. Parsing algorithm: Steps and rules encoded in the parser

After seeing the main characteristic of our dependency parsing approach, we now proceed to present a summary of the parsing algorithm and its rules. The full algorithm step by step can be found in Digital Appendix Z.1.2.

First of all, the algorithm retrieves the verb patterns, taking into account the number of verbs in the input. If there is more than one verb in the input, the algorithm fuses the patterns as shown in Figure 8.

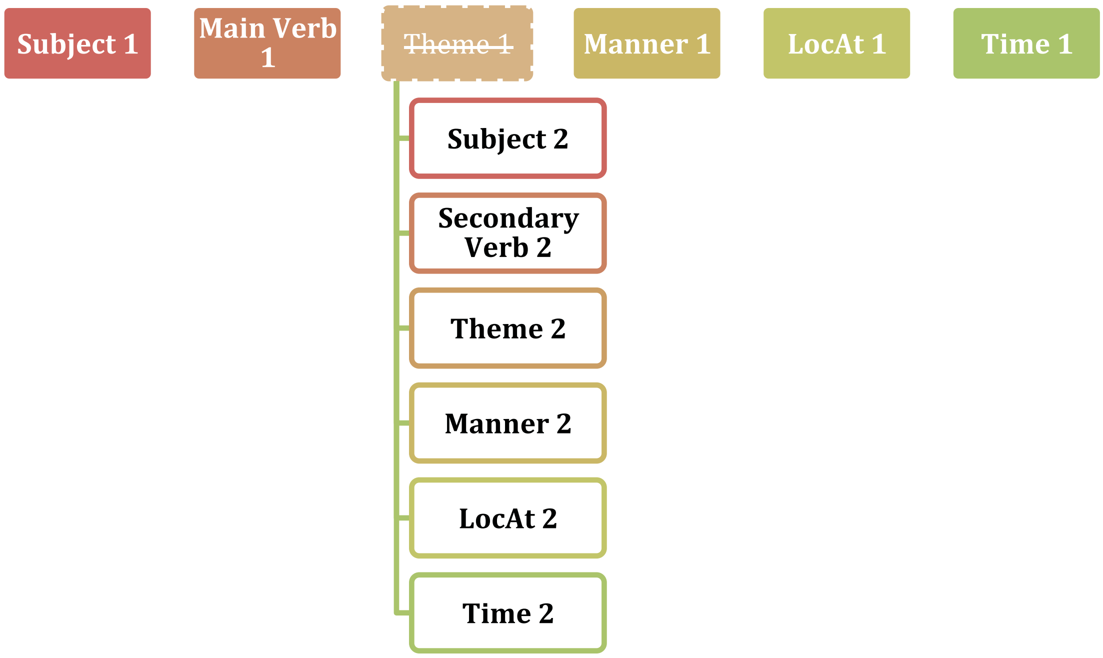


Figure 8: Example of fusion of patterns for inputs with two verbs. The slot that needed a verb disappears and is replaced by the slots of the pattern of the Secondary verb.

Then, for each pattern, the parser tries to fill its slots or semantic roles with different types of words present in the input in the following order: nouns, adverbs, adjectives and modifiers, including quantifiers.

For each of these types of words, the parser checks if they can fill the slot according to the preferred class encoded in the verb pattern. There is a weighted graph that compares classes between them to see how well a word can fill a slot. For instance, a slot that expects an animate noun will be better fit by a human than by an object. In Figure 9 you can see all the classes that a noun can have in our system.

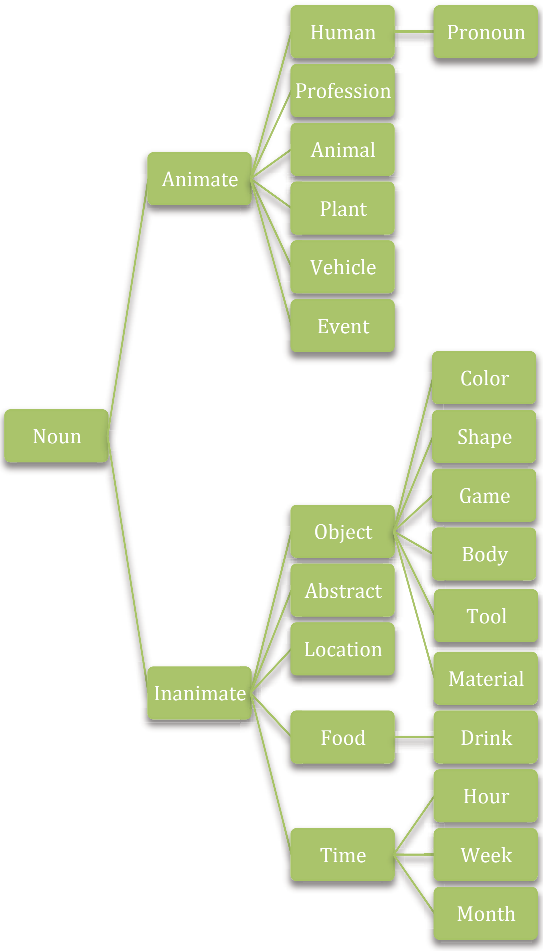


Figure 9: Hierarchy of noun classes in the system. A noun can have multiple classes.

For types of words that can complement other words, the algorithm also checks, by means of set rules and another graph, how well they complement the words that already act as the head of a slot. These rules and graphs help the algorithm determine the best fits and all the dependencies between the words in the input. Figure 10 shows a summary of these rules.

Finally, once the system has the resulting parse tree for each pattern, the algorithm gives an overall score to each of them and chooses the one with the highest score.

The scoring works as follows:

- If the slot is mandatory, its initial score is higher (more than three times higher than optional slots). This is in order to fill mandatory slots before optional slots. Only if the fit with the mandatory slot is terrible, optional slots will be filled beforehand.
- The worse the fit is, in other words, the further the class that the slot prefers is to the class of the word that fits it, the more points are subtracted to the initial score. This information is encoded in a weighted directed graph.
- If it is a Subject slot and the word that fills it is not a terrible fit[23] and it was inputted sequentially before the verb (only for SVO languages), extra points are given to the fit in order to level it to the importance of mandatory slots.
- As a result of the previous rule, for SVO languages, words that appear before the verb have a higher chance of fitting a Subject slot.

Figure 10: Rules that define how the parser scores the fit of a word with a certain slot.

Once the system has selected this pattern, it will apply the generation algorithm to it. An example of the result of the parsing algorithm could be the following. From the input:

(1) Ahir donar voler nena ós peluix vermell.

(1) Ayer dar querer niña oso peluche rojo.

(*Yesterday give want girl bear teddy red*)

A recreation of the parsed output is on Figure 11 (dependency labels are in brackets).

[23] The parser uses a threshold for terrible fits. Terrible fits are encoded in the values of the graph that represents how semantic classes fare with each other (for more details go to Digital Appendix Q, file “Mymatching.php” in folder “libraries”, to see the matrix that represents the graph).

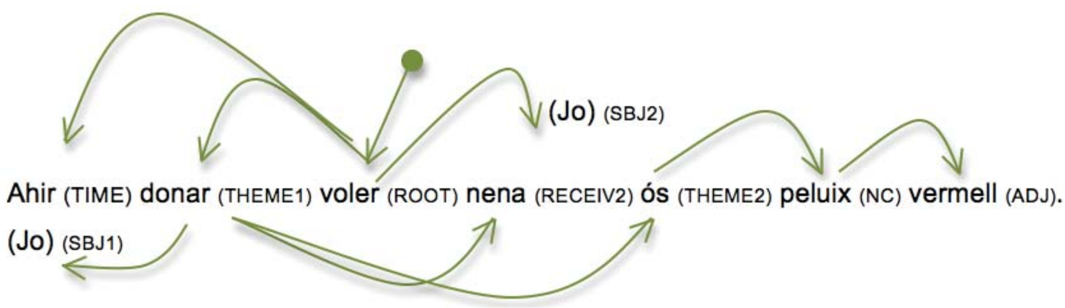


Figure 11: Recreation of the parse tree of an example input sentence in Catalan.

4.4. Generator

Even though at this point the parsing process has already finished, in order to obtain the full natural language sentence, a generation algorithm is needed.

The generator in the system is built using independent modules, each with a specific task, that progressively expand the parse input step by step. As opposed to the parser, the generator uses mainly syntactic information to expand the parsed input into a natural language sentence in Catalan and Spanish. Therefore, the generator is language dependent and, even though, some of the algorithm's modules are similar, there is a different version of each of the modules for each language. Most of the modules are also tailored to the controlled grammar of our compansion system. Nevertheless, they have been designed in such a way that they can be improved in the future and that they would be able to accept new structures. In order to add a new language, all the modules of the generator have to be programmed to reflect the specific grammar of that language.

In the next subsections, we will summarize the aim of each of these modules and also, for some of them, we will briefly see how do they work and some of the grammatical rules that they apply in the generation process. As well as with the parser, the full algorithm step by step can be found in Digital Appendix Z.1.3.

4.4.1. Slot ordering module

This module orders the components (slots) of the sentence according to the sentence type (declarative, imperative, desire, etc.) of the sentence that it is generating. For example, for questions, the semantic role with the question particle would go at the beginning of the sentence for both Catalan and Spanish.

4.4.2. Word order module, prepositions module and agreement module

Inside each filled thematic role, this module will put the head and its dependents in the correct order, depending on the grammatical category of the head and the dependents (including quantifiers and possessives). It also includes coordinated words that were transparent to the parser, adds the necessary prepositions and makes sure that words agree in gender and in number (e.g. attributes with the subject in copulative verbs, modifiers with their heads, etc.).

4.4.3. Articles' module

This module adds all the necessary articles to the sentence and apostrophizes them if it needs be (in Catalan). The part that puts definite articles and that apostrophises them, if necessary, is a module on its own. This sub-module, as opposed to others that are tailored for the constraint grammar of this system, can be used for any of the words in Catalan as it takes into account all the rules and exceptions to apostrophise nouns in Catalan. The sub-module was built following the micro-systemic approach described in Cardey and Greenfield (2008). The specific rules to apostrophise articles in Catalan can be found in Appendix A. The algorithm is described in Table 7, the lists necessary for the algorithm are in Table 8 and the operators used are in Table 9.

Concerning the choice of articles, it is important to keep in mind that in most of the cases, articles can either be definite or indefinite for the same noun. Both of the articles can give a small variation to the sense of the sentence. We could have added a word modifier that would be applied from the input that could allow the user to define if the article attached to a noun was definite or indefinite. Nevertheless, as in AAC applications, saving as many selections as possible is very important, we decided not to implement this solution. Therefore, what the algorithm tries to do in this module is to reflect the most common occurrences for the articles in a given situation in natural language, taking into account both syntactic and semantic features.

Conditions	
Id	**Condition text**
a	the word is a noun in the sentence
b	the word is in plural
c	the word begins in consonant except for “h”
d	the word is masculine
e	the word begins in “a”, “e”, “o” or “ha”, “he”, “ho”
f	the word belongs to the lists B, D or F
g	the word begins in “i”, “u” or “hi”, “hu”
h	the “i” or “u” is unstressed
i	the “i” or “u” is non-vocalic
j	the word belongs to the lists A or E
k	the word belongs to the list G
l	the word belongs to the list C
m	the word is not in the list H
n	the word is a proper noun

Algorithm with examples		
Level	**Condition > Operator**	**Example**
0	a (	
1	b (	
2	d > R	conills [rabbits] -> els conills
2	¬d > S)	festes [holidays] -> les festes
1	¬b (	
2	n & d > T	Pere (Peter) -> en Pere
2	¬(n & d) (	
3	c (	
4	d > O	molí [mill] -> el molí
4	¬d > P)	cadira [chair] -> la cadira
3	¬c (	
4	e (	
5	f > P	ema [em] -> la ema
5	¬f > Q)	abella [bee] -> l’abella
4	¬e (	
5	g (	
6	h (	
7	d (	
8	i & m > O	iogurt [yoghurt] -> el iogurt
8	¬(i & m) > Q)	ignorant [ignorant] -> l’ignorant
7	¬d > P)	universitat [university] -> la universitat
6	¬h (	
7	j > P	u [u] -> la u
7	¬j (	
8	k (	
9	d > O	hippy [hippy] -> el hippy
9	¬d > P)	Harriet [Harriet] -> la Harriet
8	¬k > Q))	illa [island] -> l’illa
5	¬g (	
6	l > Q	11 [11] -> l’11
6	¬l > O))))))	300 [300] -> el 300
0	¬a > exit	

Table 7: Definite articles’ sub-module algorithm: Conditions, algorithm and examples

Lists	
Symbol	**Set**
A	{una, ira}
B	{host}
C	{1, 11}
D	{a, e, o, efa, ela, ema, ena, erra, essa}
E	{i, u}
F	{hac}
G	{hippy, Harry, Harriet}
H	{ió}

Table 8: Lists of the algorithm

Operators	
Symbol	**Operator**
O	el
P	la
Q	l’
R	els
S	les
T	en

Table 9: Operators used in the algorithm

4.4.4. Verb conjugator module

This module conjugates the verbs of the sentence according to their subject and the verb tense of the sentence, which can be given by the type of sentence, a time expression, a tense modifier or by the default tense encoded in the pattern.

4.4.5. Cleaning module

This last module does the final changes to the generated sentence and gets it ready to output. The most significant changes that it performs are:

1. Reduce the subjects with the pronominal forms “jo” / “yo” (*I*) and “tu” / “tú” (*you*).
2. Transform the themes or receivers in pronominal forms into the correct pronominal form depending on the conjugation of the verb. Depending on the conjugation, pronouns go before or after the verb. See all the

transformations and the correct placing of the resulting pronouns in Table 10.

		Before the verb		After the verb	
		Infinitive or Positive Order		Else	
Catalan	Spanish	Catalan	Spanish	Catalan	Spanish
jo / mi	**yo / mí**	em	me	me	me
tu	**tú**	et	te	te	te
ell / ella	**él / ella**	el / la / li	lo / la / le	lo / la / li	lo / la / le
nosaltres	**nosotros**	ens	nos	nos	nos
vosaltres	**vosotros**	us	os	vos	os
ells / elles	**ellos / ellas**	els / les / els	los / las / les	los / les / los	los / las / les

Table 10: Transformations of feeble pronouns in Theme and Receiver slots

3. Add sentence modifiers that have not been added yet by other modules.
4. Join prepositions with articles (see Table 11 in Catalan and Table 12 in Spanish).

Preposition+Article	Contraction
de+el	del
de+els	dels
a+el	al
a+els	als
per+el	pel
per+els	pels

Table 11: Preposition plus article contractions in Catalan

Preposition+Article	Contraction
de+el	del
a+el	al

Table 12: Prepositions plus article contractions in Spanish

5. Add time expressions and set expressions, if any.
6. Add sentence punctuation depending on the type of sentence (declarative, question, exclamation, etc.).

After all the modules are applied, the sentence is ready to be output. The system saves the output sentence and the parse tree in the database and sends the results to the user interface.

The result for the example sentence (1) used previously:

(1) Ahir volia donar l'ós de peluix vermell a la nena.

(1) Ayer quería dar el oso de peluche rojo a la niña.

(*Yesterday, I wanted to give the red teddy bear to the girl*)

Figure 12 shows another example (in Catalan) of the output in the alternative interface for another input sentence. On it, we can see the input words, the output of the final sentence and the output of the parse tree.

Figure 12: Screenshot of an example of a final output of the compansion system

The input words are: "meu / germana / gran / no / anar / casa / ahir" (*my* / *sister* / *old* / *no* / *go* / *home* / *yesterday*).

The generated sentence is: “Ahir la meva germana gran no va anar a casa” (*Yesterday my older sister didn’t go home*). From it, we can see that the possessive agrees in gender with “germana” (*sister*). Also, the verb has been conjugated in the past tense due to the time expression “ahir” (*yesterday*) and the articles and prepositions have been correctly included.

Finally, the parse tree shows that the subject, “germana”, has got two complements, one is the possessive modifier, “meu”, and the other is the adjectival modifier “gran” (*old*). Also, the whole sentence has a negative modifier, “no”. The rest of the parse tree is straightforward, “casa” (*house*) fits the LocationTo slot and “ahir” is a time expression.

CHAPTER 5

COMPANSION SYSTEM EVALUATION AND RESULTS

This chapter presents the evaluation method and the results of the technical tests that we conducted in order to determine the error rate of the compansion system and also the results of tests done in order to see if existing machine translation engines could be used to add generators for new languages to the system.

5.1. Evaluation

In order to evaluate the system, we gave a set of 100 natural language sentences in Catalan and Spanish to 3 different annotators (a linguist, a translator and a computer scientist). The annotators had to input these 100 sentences for each language and then evaluate separately the output of the parser and the final output of the generator. With this, our aim was:

- To evaluate if the system could take different inputs for the same target sentence, as telegraphic language coming from pictogram-based AAC does not have a set word order.
- To evaluate the two main components of the system, the parser and the generator.

For future research purposes, we did also tests with just one of the annotators generating the sentences in English. As the generator of the compansion system cannot expand in English, we set the parser to parse in English and the generator[24]

[24] As explained in the first step of the cleaning module subsection of the generation algorithm in Chapter 4, the generator had a special modification when the input had been parsed in English, which was not to omit subjects, such as *I* and *you*, in order to better translate the resulting Spanish sentences into English. For

to generate in Spanish. In other words, the input pictograms were chosen having in mind the desired final sentence in English, but the sentence was generated in Spanish. Then, using Google Translate we translated the resulting sentences into English. Doing so, we wanted to explore the possibility of using a machine translation system as the generator of our system in order to be able to add new languages to the compansion system without having to program specific generators for each of them. With these tests we also wanted to see whether the parser worked well for languages with different structures other than Catalan or Spanish. Both the evaluation and the results of these tests will be presented in a separate results section at the end of this chapter.

Nevertheless, before detailing the obtained results for the tests conducted in Catalan and Spanish, let us better explain each of the steps of the evaluation process.

5.1.1. Selection of the set of sentences for the test

We contacted two of the associations that we had visited on the first year of research, to see if they could provide us with a representative set of sentences to test our system. Answering our request, one association sent us a set of 120 basic sentences that they commonly use to teach and train new AAC users on the use of their AAC panels or devices or sentences that persons who rely on AAC commonly build.

As our system at the moment has reduced vocabulary, mainly limited verbs and verb senses available, we had to adapt the initial set. Basically, we removed the sentences that used verbs that our system does not have or we substituted them with verbs that had a similar sentential structure (in total 30 sentences). For example, we removed sentences like “Se m’ha trencat la cadira de rodes” / “Se me ha roto la silla de ruedas” (*My wheelchair got broken*), as we have the verb “trencar” / “romper” (*to break*), but not in its pronominal form, and replaced other sentences like “Puc marxar, si us plau?” / “¿Puedo irme, por favor?” (*Can I leave, please?*), for “Puc jugar a pilota, si us plau?” / “¿Puedo jugar a pelota, por favor?”

---

example, Google Translate translates better into English a sentence like “Yo como macarrones” versus the same sentence with the subject only encoded in the verb tense “Como macarrones”.

(*Can I play ball, please?*). We also substituted missing nouns and adjectives and replaced them by words that our system has in its database.

Concerning complex sentence structures, like subordinate clauses or chains of noun complements, we did not have to remove any of them, which confirmed that, although our system has limitations in these areas, as previously explained in the section that describes the controlled grammar (Chapter 4, Section 4.1.1), it mostly fits the communications needs of persons who rely on pictogram-based AAC, as we would later confirm in the tests conducted with end users.

Finally, as there were many similar sentences in structure (e.g. “Estic cansat” / “Estoy cansado”, “Estic content” / “Estoy contento”, etc. – *I'm tired*, *I'm happy*, etc.) and not many complex sentences, we added some sentences that tested some of the most complex features of the system, like the coordination of two nouns or two adjectives, more sentences with two verbs in them and sentences with nouns acting as noun complements.

The final set of 100 sentences can be found in Appendix B.

5.1.2. Evaluation of the input

By evaluation of the input, we do not mean of the user interface, which for these tests was a provisional interface not adapted to the needs of people who rely on AAC, but of the words chosen to construct the sentences, that is the different ways in which the users can enter the words (or pictograms) to build the same target sentence. We think that this is an interesting aspect to evaluate as it can have a direct effect on the parser and, therefore, on the generated sentence.

In order to make this evaluation, we categorized all sentences input by the annotators in the following three groups:

1. Sentences that all annotators input in the same way.
2. Sentences that one annotator did differently.
3. Sentences that all three annotators input differently.

Then, to see the main differences in the input, we classified the sentences that were input differently in the seven following categories:

1. Subject omitted.
2. Receiver omitted.
3. Possessives input in a different order.
4. Quantifiers input in a different order.
5. Different sentence modifiers applied[25].
6. Temporal expressions input in a different order.
7. Other words inputted in a different order.

If a sentence could fall into two of these categories, in the evaluation of the results, we counted the sentence as half in each of the categories.

Finally, to compare and see if the results varied a lot if the sentences were built differently or not, we classified equally built sentences and differently built sentences separately into two categories:

1. Good sentences (the sentence is perfect or it can be understood although there are minor generation issues)[26].
2. Bad sentences.

If we found sentences that one annotator classified as good and the rest as bad or viceversa, we counted the sentence as half in each of the categories.

We also annotated sentences that were differently built but had exactly the same output generated sentence.

5.1.3. Evaluation of the parser

To evaluate the parser, the annotators had to classify the output of the parser for each of the input sentences into 4 different categories:

1. The analysis of the parser is correct.

[25] Category number 5 stands for different verb tenses applied (i.e. leaving the system to choose the default tense for that given sentence or specifying manually the verb tense) and for different sentence type modifiers (declarative, question, order, answer, desire, etc.).

[26] We will better detail this classification when we talk about the evaluation of the generator in Section 5.1.4.

2. The analysis of the parser is not the expected, although the solution reached could also be correct, as it does not change too much the meaning of the sentence.
3. The analysis of the parser is not correct: there is a single parsing error.
4. The analysis of the parser is not correct: either there are several errors or the parsing does not make any sense at all.

Moreover, as we wanted to have more details on the errors made by the parser, category 3 was further subdivided into several subcategories:

a. Error in the detection of the subject.
b. One of the adjectives is not parsed in the correct place.
c. Error in the detection of a noun as a noun complement.
d. The parser has misplaced a noun in a slot where it should not go.
e. The parser has chosen a verb sense, which is not the desired one.
f. The parsed sentence has an error that can be derived from the word order in the input.

Also, to evaluate the interannotator agreement on the parser, we calculated Randolphs' free-marginal multirater kappa (Randolph, 2005; Warrens, 2010) using the Online Kappa Calculator[27]. Brennan and Prediger (1981) suggest using free-marginal kappa when raters are not forced to assign a certain number of cases to each category, which is the case for our tests, as annotators could freely grade each of the sentences separately. That is why we selected a free-marginal kappa over a fixed-marginal kappa. The values of kappa can range from -1.0 to 1.0:

- -1 would indicate perfect disagreement below chance.
- 0 would indicate agreement equal to chance.
- 1 would indicate perfect agreement above chance.

A general rule is that a kappa of 0.70 (or above) indicates adequate interannotator agreement.

To calculate the kappa we took into account categories 1-4 explained above, without considering the different subcategories in 3.

[27] Randolph, J. J. (2008). Online Kappa Calculator. Retrieved from http://justusrandolph.net/kappa/. [June 4th, 2017]

Finally, we also wanted to have an overall score for the parser, so we decided to translate the qualitative scale of categories 1 to 4 into a balanced quantitative scale 0-10, using the following conversion:

- Category 1 would evaluate as a 10.
- Category 2 would evaluate as a 7.
- Category 3 would evaluate as a 3.
- Category 4 would evaluate as a 0.

Like this, if each of the categories had the same number of selections, the final average score of the parser would be a 5.

5.1.4. Evaluation of the generator

To evaluate the generator, the annotators had to use a qualitative scale similar to the one used with the parser. The four categories in which the annotators had to classify the generated sentences were the following:

1. The sentence is perfectly generated.
2. The sentence is well generated and it can be understood, although there are some minor errors.
3. The sentence is well generated, but it cannot be understood, as there are errors that come from the parser.
4. The sentence is wrongly generated and cannot be understood.

Furthermore, like it has been done before with the parser, in order to gather more information on the issues of the generator, categories 2 and 4 are further subdivided. Category 2 has the following subcategories:

a. There are minor errors in the word order of the generated sentence.
b. There are minor errors in the choice of the articles (mainly definite vs. indefinite) of the generated sentence.
c. There are minor errors with the conjugation of the verbs.
d. There are other minor errors that do not come from the parser, which are not described in a-c, above.
e. There are minor errors that come from the parsing.

And category 4 has the following subcategories:

a. The sentence is wrongly generated and cannot be understood, although the parsing was correct.
b. The sentence is wrongly generated and cannot be understood and the parsing was also incorrect.
c. The sentence was not generated at all, due to an internal error of the system.

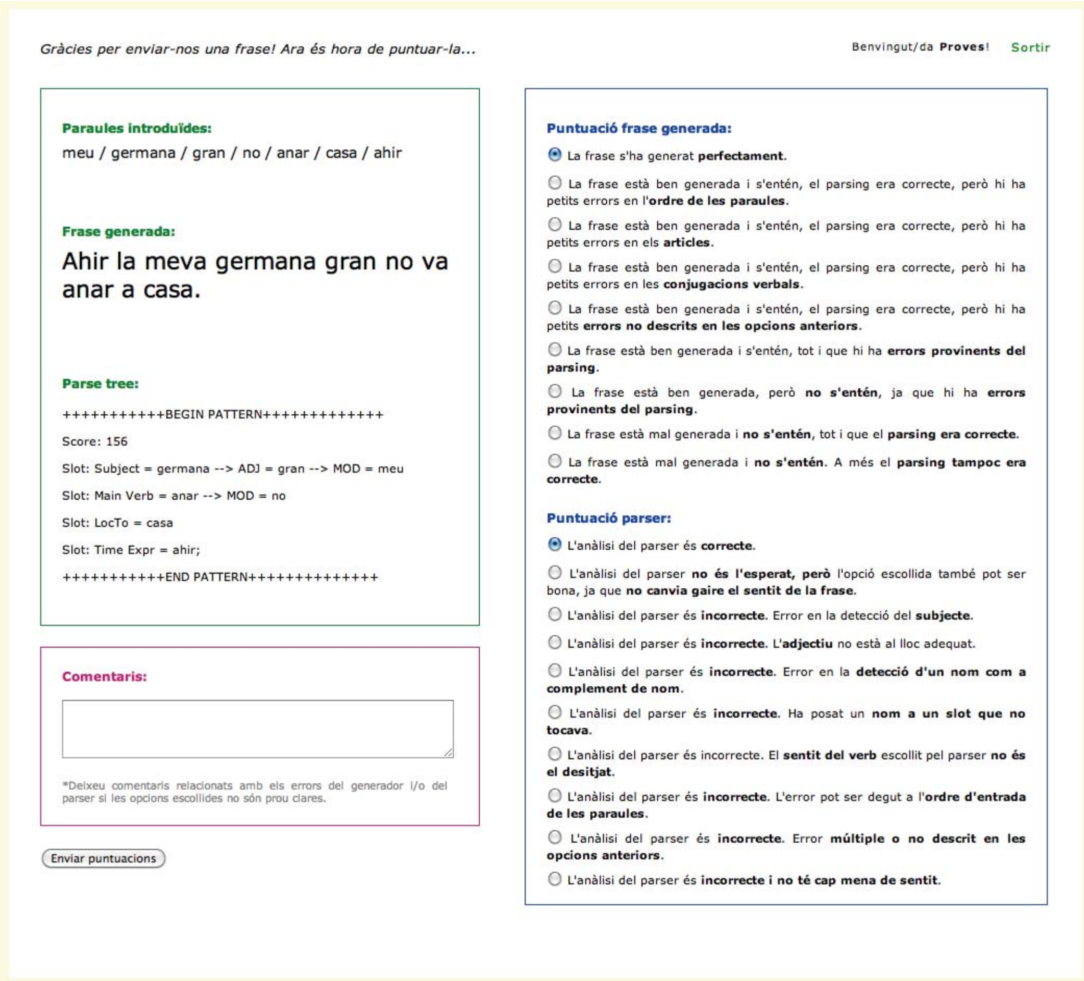


Figure 13: Screenshot of the output and the evaluation form with the non-accessible interface

As with the parser, we also calculated Randolphs' kappa to assess the interannotator agreement when rating the output of the generator. To do so, we used categories 1-4 without taking into account the subcategories of neither 2 nor 4.

Finally, to have an overall score of the performance of the generator, we transformed the qualitative scale of categories 1 to 4 into a balanced quantitative scale 0-10, using the same conversion that we used with the parser:

- Category 1 would evaluate as a 10.
- Category 2 would evaluate as a 7.
- Category 3 would evaluate as a 3.
- Category 4 would evaluate as a 0.

To simplify the analysis of the results, in the following sections, when we talk about “Good sentences” and “Bad sentences”, we considered that good sentences are the ones that fall into categories 1-2, which are either perfect or can be understood; while bad sentences are the ones in categories 3-4, which cannot be understood.

5.2. Results

In this section we will present the results obtained using the indicators that we have just described. The evaluation of the tests done in English will be presented separately at the end of this section. For these tests only one annotator intervened, therefore, some of the aforementioned indicators will not apply.

5.2.1. Results of the evaluation of the input

The raw data with all the input sentences by the annotators, along with their respective results and rates of the parser and generator, can be found in Digital Appendixes RC for Catalan and RS for Spanish.

Having noted this, we will start the analysis of the input of the 3 annotators by looking at the results in Table 13.

Indicator	Catalan	Spanish
% Sentences that all annotators inputted the same	40%	40%
% Sentences that one annotator inputted differently	**47%**	**46%**
% Sentences that all annotators inputted differently	13%	14%

Table 13: Differences in the input of the sentences by the annotators

Table 13 shows that only 40% of the sentences have been input using the exact same words (in the same order) and the same modifiers by all annotators. This confirms one of the concerns in the development of the system, which was that it had to be flexible enough so that it could treat different inputs for the same target sentences. In Table 15, we can see that this problem has not translated in the final results, as 99,17% of the sentences are well generated, even when the input by the annotators was different, against a 100% of the sentences that are good when the input is the same.

In Table 14, we can see that, by far, the most common difference is the omission of the subject in the sentence (which amounts to a 49,17% and a 47,50% of the total differences in the input in Catalan and in Spanish respectively). Another important difference, with a 25% of occurrences, is the different use of modifiers in the sentence. Taking a closer look at these sentences, we can see that it is mostly due to the use of the verb “voler” / “querer” (*to want*) instead of the “Desire” modifier, the use of the word “No” / “No” (*No*) instead of the “Negative” modifier and the use of the “Answer” modifier in short sentences. As stated before, we can see in Table 15 that all these differences have been, in general, correctly treated by the system.

Type of difference	Catalan	Spanish
% Omission of the subject	**49,17%**	**47,50%**
% Omission of the receiver	0,83%	0,83%
% Different order of the possessives	4,17%	4,17%
% Different order of the quantifiers	6,67%	6,67%
% Different sentence modifiers used	25,00%	25,00%
% Different order of the time expressions	5,83%	5,83%
% Different order in other words	7,50%	8,33%
% Omission of the verb	0,83%	1,67%

Table 14: Types of differences in the input

Another interesting data that reinforces this last statement is that a 95% of the sentences that were input differently by the annotators have exactly the same output. In other words, a 97% of the total outputs by the generator (both outputs that came from sentences that had the same input, which are always identical, and outputs of sentences that had a different input) are identical.

Indicator	Catalan	Spanish
% Good sentences when input was the same	**100,00%**	**100,00%**
% Good sentences when input was different	**99,17%**	**99,17%**
% Identical output when input was different	95,00%	95,00%
% Good sentences in total	**99,50%**	**99,50%**
% Identical output in total	97,00%	97,00%

Table 15: Good, bad and identical sentences depending on the type of input

Other general observations that we could extract from the input sentences are that there is a small learning curve to use the compansion system and take advantage of its capabilities. We observed that, in the beginning, most of the annotators stated all the words (including all the redundant subjects *I*, which the system usually takes as default), while in the middle or at the end of their task they only stated the subject in sentences where there could be an ambiguity. The same happened in the use of sentence modifiers. In the beginning, annotators would rather use the verb “voler” / “querer” (*want*), instead of the “Desire” modifier, or they did not use the “Answer” modifier to specify that the sentence could very well be “verbless”, while at the end, annotators used modifiers at its fullest.

Another indicator that there is a small learning curve is that, in the beginning, annotators repeated some of the sentences (the most complex ones, in terms of parsing complexity) once or twice to get the desired result, while afterwards, this was reduced to a minimum. Nevertheless, what confirms that it is a “small learning curve” and not a bigger one, is that after having input around 70 sentences, all annotators seldom had to repeat the input to obtain the desired result.

All the detailed results for the input, sentence by sentence can be found in Digital Appendixes SC and SS and its analysis in Digital Appendixes TC and TS.

5.2.2. Results of the parser

As we can see in Table 17, there has been nearly no disagreement among the annotators (the free-marginal Kappa is 0,978 for both Catalan and Spanish) when grading the parser. Even more, the amount of outputs of the parser that were labelled as correct parses (Category 1) by all annotators goes up to 98%.

We can also see from Table 16, that the subcategories to better detail the errors of the parser in Category 3 have been barely used. The total number of sentences that were classified as good parses (Category 1 or 2) is 99,67%.

Indicator	Catalan	Spanish
% Sentences rated as Category 1	**99,00%**	**99,00%**
% Sentences rated as Category 2	0,67%	0,67%
% Sentences rated as Category 3	0,33%	0,33%
% Sentences rated as Category 4	0,00%	0,00%

Table 16: Rates of the parser outputs

Also, if we look at the detailed rates for the parser sentence by sentence, which can be found in Digital Appendixes UC and US, we can see that there was not a single sentence rated as Category 3 or Category 4 by all annotators. The disagreement comes from sentences that were input differently by the annotators. In all cases, there was at least one annotator that found a way to produce a good sentence. This leads us to think that the parser is indeed ready to support the type of sentences for which it was conceived. Moreover, as the algorithm is the same, the results of the parser are also exactly the same for Catalan and Spanish, despite some slight differences in the inputs (as it can be seen in Table 14), considering all the words in the input as pictograms with the same meaning.

Indicator	Catalan	Spanish
Average Score of the parser (0-10)	**9,96**	**9,96**
% Sentences rated Category 1 by all annotators	98%	98%
% Sentences rated Category 1 or 2 by all annotators (good parses)	**99,67%**	**99,67%**
Percent of overall agreement	0,983	0,983
Free-marginal Kappa	**0,978**	**0,978**

Table 17: Parser average score, % of good parses and Kappa

5.2.3. Results of the generator

In the generator assessment, there has been a little bit more disagreement among the annotators, although the Kappa is still clearly above 0.70. As shown in Table 20, Kappa is 0.876 for Catalan and 0.858 for Spanish, which still indicates good interrater agreement. Here, also as opposed to in the assessment of the parser, the number of sentences rated in Category 1 (perfectly generated sentences) by all annotators has dropped to a 83% for Catalan and to 80% for Spanish (Table 20). Nevertheless, the number of sentences rated either in Category 1 or 2 by all annotators is a near perfect 99%.

Indicator	Catalan	Spanish
% Sentences rated as Category 1	**91,33%**	**89,33%**
% Sentences rated as Category 2	8,33%	10,33%
% Sentences rated as Category 3	0,00%	0,00%
% Sentences rated as Category 4	0,33%	0,33%

Table 18: Rates of the generator outputs

In Table 19 below, we can see the types of minor errors that the annotators pinpointed in generated sentences that fell under category 2. If we further subdivide some of the types of errors found going sentence by sentence (see Digital Appendixes RC, RS, UC, US, VC and VS), we observe the following:

- Most of the word order issues are due to moving time expressions at the beginning or at the end of the sentence (66,67% of cases), when it should be the opposite (the correct order depends on the specific time expression). For example, the generated sentence "Nosaltres anirem a cantar el dimecres" / "Nosotros iremos a cantar el miércoles" (*We are going to sing on Wednesday*), would have been better generated as "El dimecres nosaltres anirem a cantar" / "El miércoles nosotros iremos a cantar" (*On Wednesday, we are going to sing*). Another difference occurs when the generator puts the subject before the verb in "pseudo-impersonal" verbs like "agradar" / "gustar" (*to like*), which goes usually after the verb (i.e. "Les pomes m'agraden" / "Las manzanas me gustas", should be better generated as "M'agraden les pomes" / "Me gustan las manzanas" (*I like apples*), which sounds more natural in both Catalan and Spanish). Nevertheless, even though the word order might not be the usual in these two cases, both result in perfectly well formed sentences in Catalan and Spanish.
- Nearly half of the occurrences of other issues in Catalan (44,44%) are due to pronominal subjects that could have been omitted, but that the generator did not reduce ("ell", "nosaltres", "vosaltres" – *he*, *us*, *you* (plural)). This was a decision made taking into consideration the context of AAC users: redundancy of the subject helps avoid ambiguity and reinforces the role of the addressee who the AAC user is talking about.
- Also, the total number of issues was a bit higher in Spanish (20% of sentences not rated perfect by all annotators versus a 17% in Catalan). This increase was mostly due to a problem accentuating verbs with feeble pronouns attached to them, such as "Dámelo" (*Give it to me*), which the generator outputs as "Damelo". This translates into the higher percentage of other minor errors in Spanish, 45% against 35,29% in Catalan.

Another issue that can slightly alter the meaning of the generated sentence is the inadequate selection of articles, which totals a 23,53% of the occurrences of minor generation problems in Catalan and a 20% in Spanish. The ambiguity of choosing either a definite pronoun or an indefinite pronoun is one of the most difficult to solve, as it requires knowledge on the context that most of the time cannot be obtained from the current input or even the previous inputs in a conversation. A solution would have been to add a modifier to indicate whether an article should be definite or indefinite, but this requires high linguistic competence and also adds extra effort for the users, which is what we are trying to avoid in the first place. As most of the time the meaning is not altered substantially and can still be properly understood, we decided that, for the time being, the algorithm was already working well enough. Improving this algorithm that selects the best article in context for a given noun will remain a task for future work.

Subcategory	Catalan	Spanish
% Subcategory 2.a (word order)	29,41%	25,00%
% Subcategory 2.b (article)	23,53%	20,00%
% Subcategory 2.c (verb conjugation)	5,88%	5,00%
% Subcategory 2.d (other)	**35,29%**	**45,00%**
% Subcategory 2.e (parsing)	5,88%	5,00%

Table 19: Types of minor errors in the generator outputs

Finally, to conclude the analysis of the results of the generator, just add that we do not calculate the overall F-score[28] in our system, as when precision and recall are the same, which is our case (all inputs had an output), the F-score is the same as the percentage of good sentences. As we can see in Table 20, in the generator this measure goes up to 99,66%.

All the detailed results for the generator can be found in Digital Appendixes UC and US, along with the detailed results for the parser.

[28] F-score is used in statistical analysis to measure the accuracy of a test when doing a binary classification. To calculate the F-score, both the precission (*p*) and the recall (*r*) are taken into account: Precission (*p*) is the number of correct results divided by the number of the results returned by the system; Recall (*r*) is the number of correct results divided by the number of all the results that should have been returned by the system.

Indicator	Catalan	Spanish
Average Score of the generator (0-10)	**9,72**	**9,66**
% Sentences rated Category 1 by all annotators	83%	80%
% Sentences rated Category 1 or 2 by all annotators (good sentences)	**99%**	**99%**
% Sentences rated Category 1 or 2 in total (good sentences)	**99,66%**	**99,66%**
Percent of overall agreement	0,907	0,893
Free-marginal Kappa	**0,876**	**0,858**

Table 20: Generator average score, % of good parses and Kappa

5.3. Evaluation and results of the tests conducted in English

The indicators used to evaluate the tests conducted in English were a little bit different from the ones used in the tests in Catalan and Spanish. This was due to, first of all, having conducted the tests with just one of the annotators, thus indicators such as the free marginal kappa did not make sense, and also because the errors present in the generated sentences were a bit different. For example, as we used Google Translate to translate our output in Spanish into English, common errors that are found in machine translation between pairs of natural languages, such as translation of idioms and frozen expressions, appeared.

Therefore, the categories used to evaluate the generation, or rather translation, of the sentences in English were the following:

1. The sentence is perfectly translated.
2. The sentence is well translated and it can be understood, although there are some minor errors.
3. The sentence could be understood, but there are more significant errors, some of which may come from the parser.
4. The sentence cannot be understood; there are severe errors in it.

As with Catalan and Spanish, categories 1 and 2 are considered as “Good sentences” and categories 3 and 4 as “Bad sentences”.

Also, to have an overall score of the performance of the system generating sentences in English with the help of Google Translate, we transformed the qualitative scale of categories 1 to 4 into a balanced quantitative scale 0-10, using the same conversion that we previously used:

- Category 1 would evaluate as a 10.
- Category 2 would evaluate as a 7.
- Category 3 would evaluate as a 3.
- Category 4 would evaluate as a 0.

Furthermore, like it has been done in the previous tests, in order to gather more information on the issues of the translation, categories 2, 3 and 4 are further subdivided. Category 2 has the following subcategories:

a. Although the sentence can be understood, there are minor errors in the translation of certain words.
b. Although the sentence can be understood, there are minor errors with prepositions.
c. Although the sentence can be understood, the auxiliary verb *to do* is missing when asking a question. For example, “You want us to go home?” instead of “Do you want us to go home?”.
d. Although the sentence can be understood, there are minor errors with the conjugation of the verbs.
e. Although the sentence can be understood, there are minor errors with articles.
f. Although the sentence can be understood, there are minor errors in the word order of the sentence.

And categories 3 and 4 have the following subcategories:

a. The sentence cannot be understood or its meaning is different, as there are severe errors that come from the parser (Category 4).
b. The sentence could be understood, but there are significant errors in the word order of the sentence (Category 3).

c. The sentence could be understood, but there are significant errors in the translation of certain words that change the meaning of the sentence (Category 3).
d. The sentence could be understood, but there are significant errors related to sentences being translated into frozen expressions or idioms when they were not or related to words that were not translated into frozen expressions or idioms when they should have (Category 3).
e. The sentence could be understood, but there are missing elements in the sentence (Category 3).
f. The sentence cannot be understood, as there are severe translation problems that may or may not be related to words interpreted as frozen expressions or idioms (Category 4).

The results obtained in these tests have been encouraging, but, as expected, are not as good as with tests in Catalan and Spanish, for which there were specific generators. The number of perfectly translated sentence reached 76%, the number of good sentences reached 92% and the average score of the resulting sentences was 8,87, as it can be seen in Tables 21 and 23.

Regarding the minor errors found (see Table 22), the most common ones were due to badly translated words that, although did not affect the overall understanding of the sentence, they slightly changed its meaning (34,38%). For example, "medicamento" translated into *medication* instead of *medicine*. Other interesting errors to note are errors in the translation of locative prepositions (21,88%), such as *in*, *on* or *at*.

Indicator	Total # of sentences	Percentage
% Sentences rated as Category 1	**76**	**76,00%**
% Sentences rated as Category 2	16	16,00%
% Sentences rated as Category 3	6	6,00%
% Sentences rated as Category 4	2	2,00%

Table 21: Rates of the translated sentences

As for more significant errors that did affect the understanding of sentences, a 50% of them were due to wrongly taking words for frozen expressions when they were not or to wrongly not translating words into expressions when they should have been (subcategories 3.d and 4.f). For example, “El bicho raro se ha escondido detrás de la puerta” was translated as *The weirdo has been hidden behind the door*. Sometimes, “bicho raro” can be translated into *weirdo*, but in this case it should have been either *rare bug* or *weird bug*. This example sentence also had a problem with verb conjugations that in Spanish work in a way, but that should be translated into different forms in English.

Other important errors found were related to problems with the parser or with problems that carried from particular behaviours of the generator in Spanish (30%; subcategories 4.a and 3.b). It seems that for AdjN types of languages, there is a problem when adjectives are combined with noun complements. For example, *DRINK-GLASS-WARM-MILK* was parsed in the following way: *warm milk* was the Theme and *glass* was the Tool instead of *warm milk* being a noun complement of *glass*, as in *drink a glass of warm milk*. Also, for “pseudoimpersonal” verbs, the less common order in Spanish that the generator used, does not usually work in English and Google Translate only changes it in short sentences. For example, the sentence “La nieve me gusta” is correctly translated into *I like the snow*, while the sentence “Leer libros the miedo me gusta” is translated into *(To) Read scary books I like*, which does not sound natural at all.

Subcategory	Total # of sentences	Percentage
% Subcategory 2.a (translated words)	**5,5**	**34,38%**
% Subcategory 2.b (prepositions)	**3,5**	**21,88%**
% Subcategory 2.c (questions)	2	12,50%
% Subcategory 2.d (verb conjugation)	2	12,50%
% Subcategory 2.e (articles)	2	12,50%
% Subcategory 2.f (word order)	1	6,25%

% Subcategory 4.a (parser)	1	12,50%
% Subcategory 3.b (word order)	1,5	18,75%
% Subcategory 3.c (translated words)	1	12,50%
% Subcategory 3.d (frozen expressions)	**2**	**25,00%**
% Subcategory 3.e (missing elements)	0,5	6,25%
% Subcategory 4.f (severe 3.b or 3.d)	**2**	**25,00%**

Table 22: Types of errors in the translated sentences

Finally, important expressions like *No*, were translated into specific usages instead of more general ones, which are correct for wider contexts and which would be a preferred behaviour for AAC. In this specific case, the sentence “No” (*No*) was translated into *Do not*, which would only be correct when telling someone not to do something.

To conclude this section, although the results obtained while generating/translating the sentences in English were quite good (all the detailed results for the tests conducted in English can be found in Digital Appendixes RE and TE), we are not sure if they would be good enough to conduct tests with persons who rely on AAC, as, even though the percentage of good sentences reaches 92% (see Table 23), some of the errors found are quite severe and could very well raise some eyebrows and cause frustration and the later abandoning of the system. Therefore, before advancing to a test phase with end users, we think that we would have to work on a small side generator that would focus on these errors in order to minimise its impact on the final output.

Average Score of the translated sentences (0-10)	**8,87**
% Sentences rated Category 1 or 2 (good sentences)	**92%**
% Sentences rated Category 3 or 4 (bad sentences)	8,00%

Table 23: Translated sentences average score and % of good and bad sentences

Chapter 6

USER INTERFACE

In this chapter we will discuss the user interface, and its main components, developed in order to test the compansion system with persons who rely on AAC. While usually a user interface in order to test a system is not a relevant topic in a doctoral dissertation, in this case the user interface is an essential element for the success of the tests. Persons who rely on AAC usually have, apart from severe speech impairments, physical impairments and/or cognitive impairments that need to be taken into account in order to design and develop an accessible interface that fits their needs. Furthermore, we wanted to include a pictogram prediction system in the user interface that could take advantage of all the semantic information encoded in the verb patterns of the compansion system. We hoped that this prediction system would reduce the amount of effort needed by the users in order to build sentences.

All in all, while developing the user interface, apart from taking into account the tests that we had to conduct, we wanted it to become a full-fledge pictogram-based AAC application that could be used by persons who rely on AAC, thus, if tests were successful, transferring our research into a practical solution that could impact people’s lives.

As it will be described in the next chapter, we conducted the first sessions of the tests integrating our compansion system into Plaphoons, which is a pictogram-based AAC open-source software that was used by the persons that took part in the tests. However, due to technical issues we had to rule out integrating Plaphoons in our system as a definitive solution and we were forced to develop a new GUI.

One of the main problems to integrate Plaphoons was that it was impossible to modify its GUI dynamically. This limitation meant that we could not integrate the prediction system and, although we could incorporate an audio feed of the final sentences generated by the compansion system, we could not incorporate a visual written representation of them. This was not a major problem for our trials, since none of our participants read written texts, but it could be a problem for potential users who read written texts or, especially, for those who have the potential to do so or for those who are developing reading skills.

Another issue was that Plaphoons could communicate only with our compansion system through text and with a different character encoding from the UTF-8 encoding that we were using. Certain diacritical marks would get lost during conversion, which meant that our compansion system could not distinguish properly between certain words and the parser would sometimes assign the wrong grammatical category. Consequently, certain slots would be assigned the wrong pictogram, and later, if any articles or prepositions were needed they would not be generated correctly. Furthermore, because Plaphoons communicated with the compansion system using only text, not actual pictograms, two of the main advantages of pictogram communication were lost: the use of different pictograms for a single polyseme (a word with various related meanings) and for homographs (different words with the same spelling).

During the sessions where we used Plaphoons in the tests of the compansion system with end users we were able to bypass these issues (except the inclusion of the pictogram prediction system) by encoding the identifier for each pictogram hidden in the text. However, this temporary solution was time consuming and made making changes to the boards of pictograms of each participant quite difficult. Furthermore, it could not be used as a permanent solution, as it required a deep knowledge of the database, which Plaphoons could not get access to. All these aspects, made the development of a new user interface necessary.

6.1. Characteristics and design aspects

In this section, the essential design aspects of the user interface will be discussed. We present the main interface features, aided communication design aspects for dynamic communicators (Reichle & Drager, 2010) and other relevant elements that were added to the interface, along with important information that needed to be taken into account for its design. The pictogram prediction system will be presented separately at the end of this chapter in Section 6.2.

All the features of the user interface as well as the final application can be tested at http://jocomunico.com. For a quicker look at the interface and at its main characteristics, a video showing them, as well as the participants that took part in the tests, can be found online at https://youtu.be/AQ5m0ZZO7no.

6.1.1. Device selection

As seen in the Background (Chapter 2), the AAC scene at the moment is switching towards portable devices that enable individuals with complex communication needs to use them in any communicative situation without any previous setup, instead of computers that need to be wired. Nevertheless, due to the wide range of motor impairments of persons who rely on AAC, many individuals, schools and centers still use computers, which still have a wider support for peripherals and accessories needed to access the computer by many individuals that use AAC. That being said, every day more AAC peripherals are already being designed specifically for smartphones and tablets[29], although its variety and accessibility options is still not the same as with computers.

Therefore, we decided to develop a multi-platform user interface that could be used on most devices and on most operating systems. The answer was to build a web-based application. The main advantage is that the same person is able to access their application from different devices. It also means that speech language pathologists, teachers, family members or caregivers can configure and edit the application on a remote device and changes are available directly on the user's device. The main disadvantage is that a working Internet connection is required. In order to minimize the impact of this drawback, we also developed offline versions for Windows and Mac operating systems, which only implied to set up a local server in the computer and a few changes in the code.

---

[29] Some examples of peripherals and accessibility options would be Mouse4All (http://mouse4all.com/en/) or RJ Cooper & Associates, Inc. (http://www.rjcooper.com/). [June 26th, 2017]

We also decided to make the code available for free under a Creative Commons BY-NC license[30], so that developers could make changes, add functionalities or adapt it to other operating systems, such as Linux. All the code developed during the thesis (companion system and user interface with its pictogram prediction system) can be found in a GitHub repository[31] as well as in Digital Appendix Q.

6.1.2. Board customization

Due to their disabilities, persons who rely on pictogram-based AAC do not usually customize their applications. Usually, speech language pathologists, teachers, family members or caregivers are the ones that do so. Therefore, these applications are usually split into two main components: the side used by the users to communicate and the side used to configure the application. The main screen of the user's side of most pictogram-based AAC applications is the board with pictograms. This screen and the rest of the screens that can be accessed by the user are the ones that require to be the most accessible.

As the application will target individuals with different types of disabilities (even among similar kinds of disabilities there can be a huge difference in physical capabilities, like sight problems, motor impairments and so on) it is very important for the interface to be extremely customizable. Some of the layout aspects that need to be adaptable to each of the users are size and number of pictograms displayed on screen, position of pictograms and color palette of the application, among others. As the number of accessibility options can be nearly endless, we decided to implement the ones that were necessary for the participants in our tests to be fully capable of using our interface and also the ones that could accommodate a wider future audience.

Going back to the configuration and distribution of the boards, it is necessary to remind that each user has different needs in terms of vocabulary, clustering of the different pictograms into groups, size of the images and so on. That is why a basic initial configuration will be provided in order to showcase how the application

[30] Creative Commons BY-NC license. http://creativecommons.org/licenses/by-nc/4.0/ [June 26th, 2017]

[31] GitHub jocomunico.com. https://github.com/narum/jocomunico/ [June 26th, 2017]

works; nevertheless, everything, except for some structural items, will be fully customizable. The only fixed items will be:

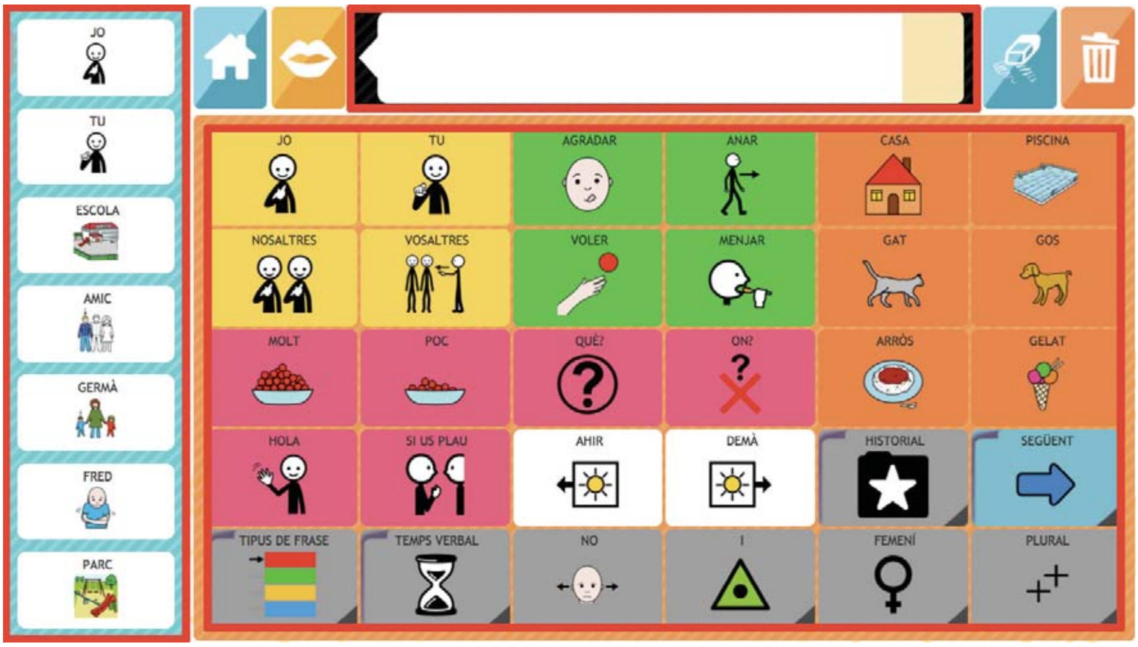


Figure 14: Example of a board in our user interface with its main elements highlighted

- **The sentence bar** (top of Figure 14): it's a horizontal bar where all the selected pictograms chosen to build the sentences appear and where the final generated sentences are shown. This bar can be either at the top or at the bottom of the page. At the end of this bar, there is a colored part where the selected sentence or tense modifiers for the current sentence are shown. If nothing is shown, it means that the compansion system is set to its default values.

  Next to either side of the bar, there are function buttons that can be either active or inactive (they are not shown if they are inactive) that are used to go back to the main board of the user, to generate and read a sentence with the currently selected pictograms, to erase a pictogram in the sentence, to erase all the pictograms, etc.

- **The pictogram prediction system bar** (left of Figure 14): if active, it will appear on the left side of the screen.

- **The board with the pictograms** (bottom-right of Figure 14): it has a grid structure with customizable number of rows and columns.

Furthermore, in terms of the customization of the boards, in all the boards created by a user, there must be a main board, which will be the one that appears first when the application is opened. This board, depending on the distribution of the vocabulary according to the linguistic competence of each user, can sometimes be thought as a navigation hub where most of the rest of the boards of pictograms, grouped into categories, can be accessed. Even though the number of boards that can be embedded into other boards is not limited in our system, the tree of embedded boards should not exceed a depth of 3 levels, as higher values make board navigation a tiresome task.

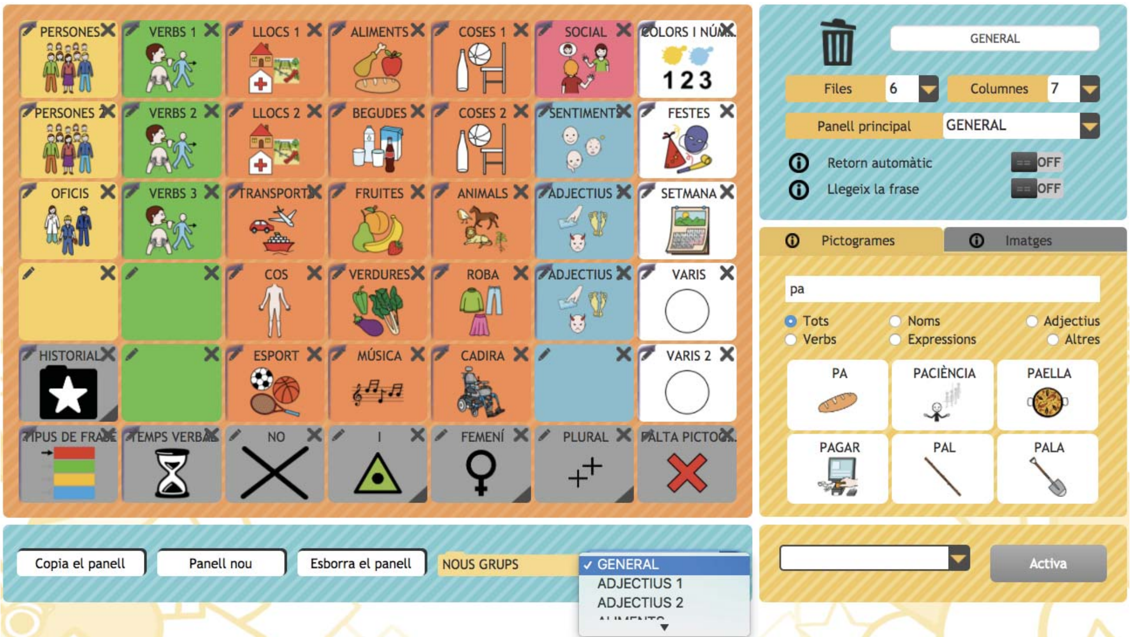


Figure 15: Screen to edit boards. In this example, it is the main board of a user that has the vocabulary clustered in both syntactic categories (e.g. verbs, adjectives, etc.) and semantic categories (e.g. people, places, fruits, etc.)

In Figure 15, we can see the screen to edit boards with one of such boards. In this particular example, it is the main board of an AAC user with high linguistic competence that has the vocabulary clustered in both syntactic categories (e.g. verbs, adjectives, etc.) and semantic categories (e.g. people, places, fruits, etc.).

Apart from choosing pictograms and adding or removing rows and columns of the grid, the edit board screen can be used to set other board characteristics:

- **Automatic return feature**: if activated, when a user makes a selection from this board it will automatically return to the main board. This feature is useful when the main board acts as a communication hub and usually no more than one pictogram is selected from the current board. This feature can reduce the amount of clicks necessary to construct sentences, especially in users that use scanning as selection mode.

- **Automatically read the sentence**: if activated, when a user makes a selection from this board the system will automatically generate the sentence with the currently selected sentences. This feature is designed for panels that are linked mimicking a sentence structure. For example, a group of panels where in the first one you select nouns that can act as subjects (e.g. *I*), in the second one you have verbs (e.g. *to go*) and in the third you have the main complement for that verb (e.g. locations such as *park*). In this example, you would activate this feature in the locations panel, so that after selecting *park* the system would automatically generate "*I go to the park*" without having to press the generate button.

- **Switch the default images of the pictograms**: By default the system uses the open-source ARASAAC[32] set of pictograms. However, persons who rely on pictogram-based AAC might be used to other sets of pictograms, thus, the default images from the pictograms can be changed from this screen by uploading them to the system.

- **Color the cells**: cells can be colored using the colors that are commonly used in pictogram-based AAC. As seen in Chapter 2, many centers of persons who rely on AAC use a common paradigm: yellow for nouns that are humans or human pronouns; orange for other types of nouns, such as objects and locations; green for verbs; and blue for adjectives; colors assigned to adverbs, modifiers and other types of words tend to vary depending on each individual and are usually colored white, pink or grey.

Finally, each cell of the grid can be further edited. Apart from being pictograms, each cell can be a link to other boards or a function, such as the functions that are

[32] Aragonese Portal of Augmentative and Alternative Communication: http://www.arasaac.org/ [September 12th, 2017]

also available from the buttons next to the sentence bar or functions to toggle gender, number, tense and sentence-type modifiers of the companion system. It can also be a link to a folder with sentences or a sentence itself.

Furthermore, from the edit cell view the custom scanning selection method can also be configured. These last three options will be further explained in the following sections.

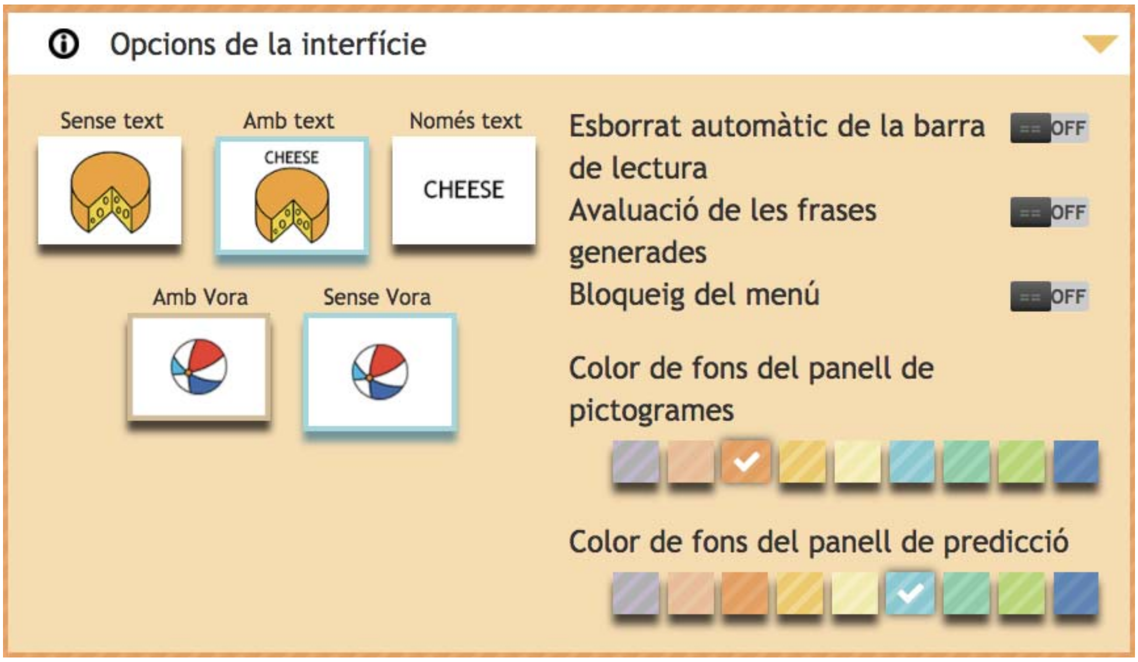


Figure 16: Settings to customize the user interface on the Configuration screen

The other main screen to customize other settings of the pictogram boards is the Configuration screen (Figure 16). Some of the features that can be set are:

- **Text of the pictograms**: whether the text of the word associated to the pictograms is shown or not. Text can be useful for individuals that have reading skills or for persons that are developing them. Text can also be of help to communication partners as reinforcement to communication.

- **Border or background color of the cells**: whether the background of the cells is colored when coloring them or whether just the border of the cells is.

- **Color palette**: patterned background colors for the board of pictograms and for the prediction column.

- **Evaluation of the sentences**: if activated, when users generate a sentence they can choose if they like the generated sentence by the compansion system or not. This feature is specifically designed for testing purposes. It can also be used in sessions by speech language pathologists in order to see if users like the output of the sentences or if users can differentiate between well generated sentences and sentences with generation errors.

6.1.3. Scanning and selection methods

Many people who rely on AAC have got motor impairments that do not allow them to use direct selection methods (e.g. the mouse or direct touch controls in touch screens) to access their computers or their communicating devices. In order for them to access their devices, different selection methods have to be implemented. Apart from peripherals that allow them to mimic the behavior of the mouse, such as eye-trackers, the most common alternative and affordable selection method is scanning. By pressing a button or a switch, a user can move a cursor from one element of the screen to another, be it a pictogram or a group of pictograms, in our case.

There are many kinds of scanning methods (Lesher et al., 1998) and within the same methods there can be many settings to configure them. In our interface, we just implemented the scanning methods necessary for the participants in our tests to use the application, which can also accommodate a wide range of potential users.

The most basic scanning method is linear scanning. In linear scanning, the cursor advances from item to item one-by-one and this way sweeps the entire screen. Another scanning method is row or column scanning, where it advances row-by-row or column-by-column. Once a certain row or column is selected, then linear scanning is usually applied to the selected subgroup, though further row scanning can also be successively applied. In our system, in the first level, we implemented row and column scanning and, in the second level of scanning, when a row or column is selected, we use linear scanning.

Finally, we also implemented group scanning, which we call customized scanning, that works exactly like row scanning, but instead of rows, the cursor advances by groups or clusters of items (square, rectangle or irregularly shaped blocks). The second level of group scanning can either be group scanning again or linear scanning. The third level of scanning, though, can only be linear scanning.

The three implemented scanning methods can advance through automatic scanning; where there is a set customizable speed that makes the cursor advance and the user just needs to select the desired item when it is highlighted; or through manual scanning; where the user controls the cursor speed by pressing a button every time that he/she wants the cursor to advance. If manual scanning is used, in order to select an item, to differentiate it from the advance directive, the user presses another button, on multiple-button switches, or presses the same button during a longer period of time, on single-button switches.

Aside from these options, we also implemented two more features for automatic scanning. The first one is the possibility that scanning only starts after the user has clicked once. This option allows the user to take a look at the board of pictograms and locate their next desired selection before the scanning starts sweeping the screen.

The second feature allows to cancel and reset the scanning cursor to its initial position. While scanning, erred selections can be a common occurrence depending on the skills and the level of attention of each individual, which depending on the type of disability can vary from day to day. An erred selection happens when the user presses the switch either too soon or too late. If the cancelling feature is not available, then the user has to wait until the scanning has swept the entire next element, be it a row, a column or a group, until it is reset to its initial position and starts again. Depending on the speed that the automatic scanning is set to, an erred selection can mean an extra 30 seconds or even more than a 1-minute loss; thus, the importance of this feature.

Furthermore, the color that highlights the scanning and the order in which to scan the main elements of the interface can also be customized. A screenshot of all the aforementioned settings can be seen in Figure 17.

Finally, apart from the different scanning options, a single configuration setting has been implemented for finger or mouse access: the minimum time of selection. It is conceived for users that have got milder motor impairments. With this option activated, the mouse or finger has to stay in the same place for a set interval of time in order to make a selection. This option can also be useful for eye-trackers, as selection is made when the mouse pointer stays within a selectable item for more than the set interval of time.

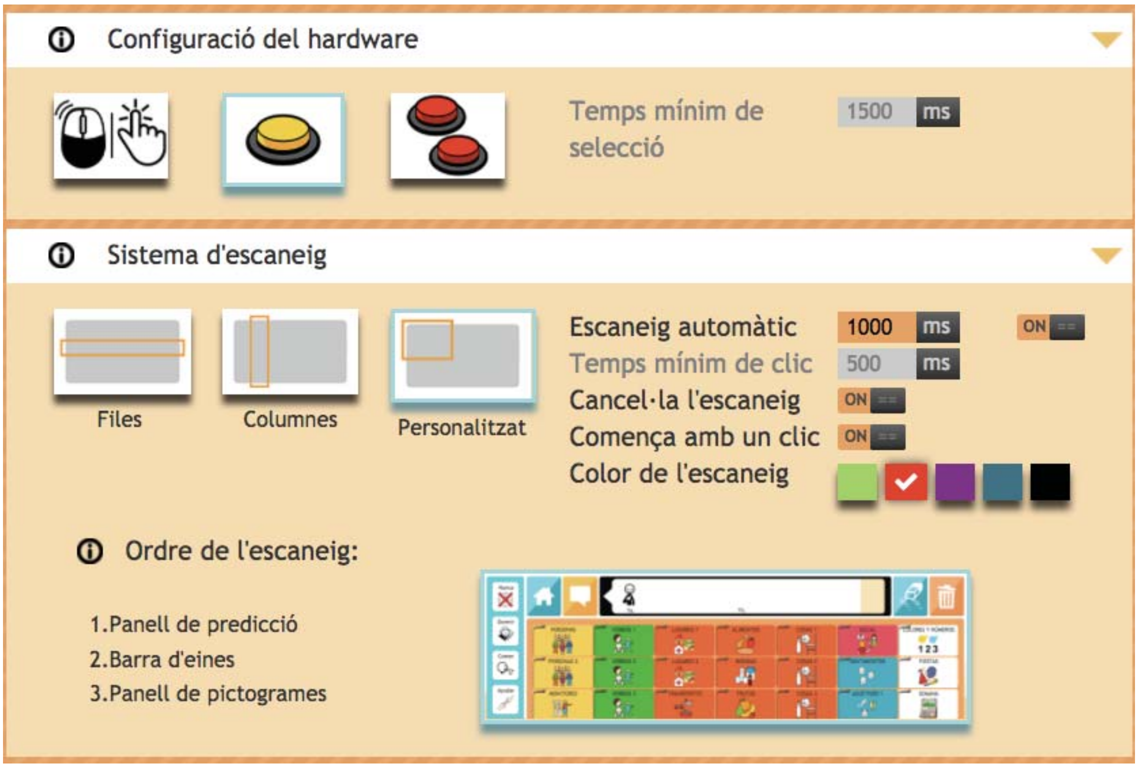


Figure 17: Configuration settings for the available selection methods (mouse/finger, 1-switch scanning and 2-switch scanning)

6.1.4. Historic and thematic folders

Another feature of the system is that it also stores the recently generated sentences (grouped into sentences used today, last week and last month), so that they can be used again without having to build them from scratch, similarly to the history feature of Internet browsers. These sentences can be later saved into custom thematic folders (e.g. “At the hospital”, “At school”, “About me”, etc.). This is an interesting feature, because it allows users to prepare in advance questions, answers or any type of sentences for a certain a conversation. This can be used in a similar way as when people rehearse sentences for meetings or important events that they have to attend to (Copestake & Flickinger, 1998).

Moreover, it is important to keep in mind that most people are not familiar with AAC technologies, that is why it is very advisable to provide a way for the users of the application to explain to their communication partners how they communicate. Furthermore, due to the constraint grammar of the compansion system, some sentences or explanations can be very difficult or can take a long time to make using pictograms, sentences such as introducing yourself in an elaborate way or such as detailing the medication that you need to take. Consequently, the administrators of the system (family members, speech language pathologists, etc.) will have the possibility to pre-record sentences in the system using a keyboard. These sentences will be summarized with a maximum of three custom images so that users that cannot read can differentiate between them. Other sentences that could be pre-recorded this way are recommendations that need to be taken into account when communicating with a person who relies on AAC (Larranz, 2006). These recommendations would be such as the following:

- Ask where you should sit in order to make the communication easier.
- Ask clear and precise questions, so that the answers can be concrete.
- Give enough time to answer. Be patient.

6.1.5. Addition of new vocabulary

As explained in the previous chapter of the compansion system (Chapter 4), the core vocabulary that we have encoded in our system is based on the CACE vocabulary, a set of vocabulary specifically designed for AAC. In order to conduct the tests, we added more vocabulary to CACE. A summary of the total amount of vocabulary available in the system is in Table 4 of Chapter 4.

Nevertheless, we decided to implement the possibility to add more vocabulary through the user interface. Otherwise, it would be really difficult for users, different from the participants in our tests, to use the application as their means of communication. Everyone uses specific vocabulary in their daily lives (e.g. names of family members and friends, vocabulary specific to their hobbies, etc.), so in order for the software to be used aside from the scope of this PhD thesis, we had to implement the option to add vocabulary to the compansion system. However, notice that in order to add new vocabulary to the system all the related syntactic and semantic information has to be provided and, in some cases, it can be quite complex. This is specially the case for verbs and verb patterns, as explained in Chapter 4. Therefore, we decided that, temporarily, we would only provide the possibility to add new nouns and new adjectives.

6.1.6. Voices and voice synthesizer engines

AAC communication devices need at least two different voices. The first one is the voice that becomes the voice of the user and the second one is the voice that the system uses to give indications to the users, the feedback voice (Escoin, 2006) or the interface voice, as we labeled it in our user interface. For example, every time that a user selects a pictogram, if the interface voices has been activated, the application will read the word corresponding to the pictogram aloud to help the user detect possible errors while selecting the pictogram. It is very important for these two voices to be different to avoid confusion for both the user and the communication partner. Usually, if the user is male, the interface voice is set to be a female voice and the other one is a male voice and vice versa.

Regarding voice synthesizer engines (or text-to-speech engines), for the online version of the application we use a high-quality text-to-speech (TTS) service named Vocalware[33] with both male and female voices available in Catalan and Spanish. Before selecting this service, tests were run on the server in order to see its latency. The delay of the response of the web service is between 250 and 3000ms depending on the length of the sentence and on the server's traffic. That means that from the instant where the user asks for a sentence to be synthesized (without taking into account the time that the compansion system takes, which is negligible), until it is read aloud by the system, there might be 3 seconds of delay. Notice, however, that persons who rely on pictogram-based AAC usually take quite a long time to build sentences, as can be seen in the statistics of the tests conducted with end-users (Chapter 8). In our tests, the mean time to produce a sentence was 3 minutes and 30 seconds so, in this case, 3 seconds would represent a 1.43%, which is not a relevant amount of time.

---

[33] Vocalware. Cloud-based Text to Speech API. https://www.vocalware.com [September 12th, 2017]

Concerning the local version of the application for Windows and Mac operating systems, the application has been programed so that it can access the default voices of the operating systems as well as installed voices, such as Microsoft Speech Platform[34] or Acapela[35] voices. Nevertheless, the voice synthesizer engine module has been programmed independently to the rest of the functionalities of the application, so that, if needs be, other engines can be added.

6.2. Pictogram prediction system

This section will discuss the design and implementation of the pictogram prediction system. The results of the system will be presented in the next chapter that discusses the tests conducted with end users who rely on AAC.

6.2.1. Prediction system design aspects

Taking into account the existing research on prediction systems for AAC, our system has the following characteristics:

1. **Learning feature:** The prediction system learns from the usage of the users by monitoring the sentences built and recording in the database the used pictograms in directional n-grams[36] (bi-grams and 3-grams). The algorithm of the system presents the user the most used pictograms depending on the last and the second to last selected pictograms (relative frequency). In case of a tie in relative frequency, then the total frequency (total amount of times that a pictogram has been used by the user) is used to break it. All in all, the more the user repeats the same sentences, the more accurate the prediction system is and the more the effort is reduced.

---

[34] Microsoft Speech Platform. https://www.microsoft.com/en-us/download/details.aspx?id=27225 [September 12th, 2017]

[35] Acapela group – Voice synthesis – Text to Speech | voice solutions. http://www.acapela-group.com/ [September 12th, 2017]

[36] In the fields of computational linguistics and probability, an n-gram is a contiguous sequence of *n* items from a given sequence of text or speech.

2. **Syntactic-Semantic feature:** Apart from reducing, by means of the n-grams, the effort required by the users to build sentences, we wanted the prediction system to be an asset that could help them better structure their sentences. As explained in the previous chapter of the compansion system, one of the main difficulties of the telegraphic language that comes from pictogram-based AAC is the lack of an established word order that relates to the conventional word order of natural language grammar. If the pictogram prediction system helped users better structure the sentences by predicting words that fitted the conventional order of the grammar, it would both improve the users language proficiency and help the compansion system at the same time. To better structure sentences is also a skill that speech language pathologists that work with people who rely on AAC usually work on.

   In order to achieve it, the system uses all the semantic information encoded in the compansion system. Depending on the previously selected items in the sentence, the system takes into account the relevance of the missing slots that need to be filled according to the already selected pictograms (i.e. if a transitive verb is selected, the theme slot is the most relevant one that needs to be filled). Once the most relevant slots are selected, the vocabulary items that can fit these slots are filtered by the same features that the parser of the compansion system uses (i.e. if a verb needs human themes, only nouns or pronouns that correspond to humans are not filtered). Furthermore, the system also takes into account syntactic aspects; such as that, in SVO languages, sentences usually start with subjects, etc.

   This feature is, in fact, the main difference between our prediction system and the other prediction systems presented in the state of the art chapter.

3. **Context feature:** The prediction system uses context-aware information. However, we do not consider location information since it is difficult to automatically map GPS information to the locations where communication takes place without it being inputted by the user and

without using peripherals and sensors in these locations, as done in L.F. Garcia et al. (2016).

On the other hand, as most people have a scheduled life and do similar things in similar times of the day or similar things in similar days of the week, our system takes into account contextual time information (days of the week and 1-hour spans during a day). This scheduled life is even truer for many AAC users that go to special education schools or that stay in daycare centers, thus we think that this feature fits perfectly with AAC. This contextual information is merged with the data obtained by the Learning feature using a weighting method. Moreover, this feature is also combined with syntactic information. For example, as nouns in conversations tend to get repeated (for instance, a proper noun can be the receiver in a sentences and then be the subject in the next sentence), in some case, when the prediction system is looking for nouns, the weighting method gives priority to nouns that have been used recently. This methodology is conceived to mimic the use of actual contextual information from conversations in order to improve predictions.

To sum up, as well as with the algorithm of the compansion system, due to the impossibility of having a large enough user-generated corpus of pictogram-based AAC sentences in a reasonable amount of time, the algorithm of the prediction system does not use advanced machine learning techniques that require a large amount of data. Instead, it uses a linear approach, such as n-grams, for the Learning feature, the semantic information encoded in the compansion system for the Syntactic-Semantic feature and total frequencies grouped by hours and days of the week for the Context feature.

6.2.2. Prediction system algorithm examples

In the following section we will describe with examples the execution flow of the algorithm. Depending on the number of pictograms selected in the sentence bar and depending on the types of these pictograms (e.g. nouns, verbs, modifiers, etc.)

the algorithm uses the aforementioned features in different ways in order to choose the pictograms that will be displayed on the prediction bar.

There will be 3 examples. The examples will be made with a fictitious board of a person named John that has been using the system for a few months. Notice that, as one of the purposes of the prediction system is to help persons who rely on pictogram-based AAC to better structure sentences, the prediction system works better when the SVO word order, for SVO languages, is followed. Thus, the examples will also follow the SVO structure.

First of all, before going through the examples in detail, it is important to note that the total number of recommended predicted pictograms displayed on screen for an AAC device is between 1 and 7 (Taylor et al., 2001). Our system can display between 2 and 7 items. This number is chosen by the user in the configuration screen. By default it is 5 (see Figure 18), although, for explanation purposes, we will use a prediction bar of 7 items in the examples.

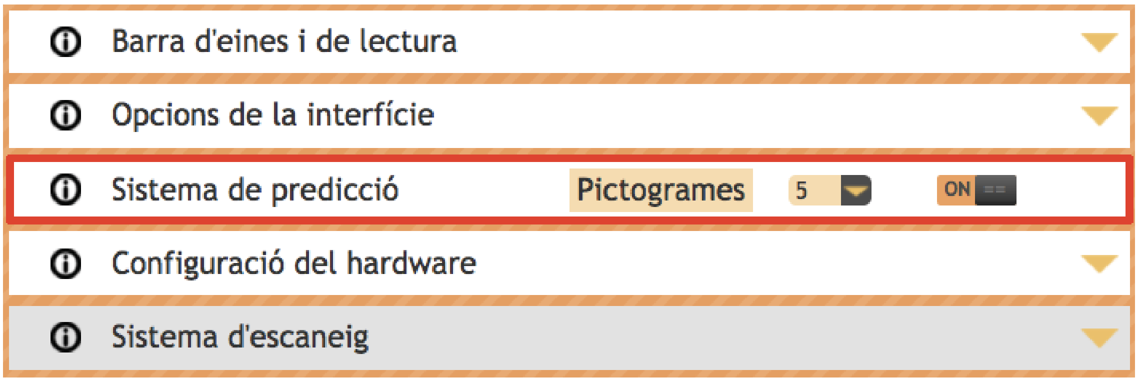


Figure 18: Configuration screen where users can activate and deactivate the prediction system as well as the number of displayed pictograms

The first example will follow the changes in the prediction bar while forming the sentence *I-GO-BEACH-YESTERDAY*. The second example will be *WANT-EAT-RICE* and the third one *MY-MOTHER-BE-TEACHER*. In each of the phases of the sentence building process (each phase will be determined by the number of pictograms currently selected in the sentence bar), for each example, we will explain why did the prediction algorithm select each pictogram, according to a given context that will also be explained. Finally, we will state the next selected pictogram, also for each example, which may be selected either from the pictogram prediction bar or from the pictogram boards.

6.2.2.1. 0 pictograms selected in the sentence bar

The initial phase that is common for all three example sentences is when there is no selected pictogram in the sentence bar yet. In this case, the algorithm only shows words that can act as subjects. Figure 19 shows the following predicted pictograms for our example user John:

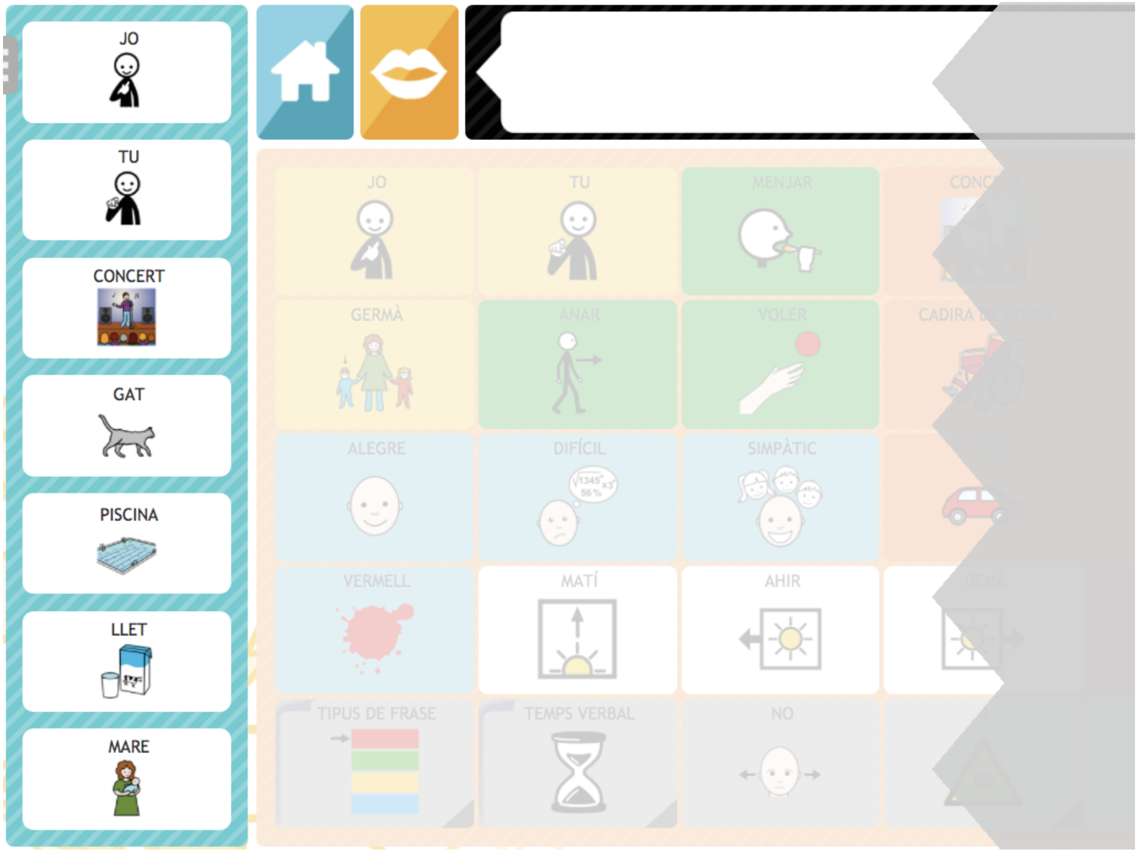


Figure 19: Predicted pictograms for all three examples when there are no selected pictograms yet

- *I* and *YOU*: these personal pronouns are always shown in the first two positions of the prediction bar when there are no pictograms in the sentence bar.
- *CONCERT*: the algorithm shows it because it is the most used noun in the past couple of days. John will be going to a concert in a few days and is very excited about it.
- *CAT*: it is the second most used noun in the past two days. John has a new cat at home and frequently talks about it.
- *POOL*: it is a noun that is frequently used on Mondays. Today is Monday and every Monday John goes to the swimming pool with the daycare center.

- *MILK*: at the moment it's 9am and it's breakfast time. Milk is a common word used around this time of the day.
- *MOTHER*: it is one of the most used nouns by John. John is very fond of his mother and talks about her a lot.

Following each example, on the first one, the selected pictogram is *I*, which was among the predicted items. On the second one, the selected pictogram is *WANT*, which was not predicted, but was among John's boards. Finally, on the third example, the selected pictogram is *MY*, which was not predicted either.

In the following section, we will describe the behavior of the algorithm for each example once there is already a selected pictogram in the sentence bar.

6.2.2.2. 1 pictogram selected in the sentence bar

In this phase, the algorithm changes depending on the type of word of the first selected pictogram.

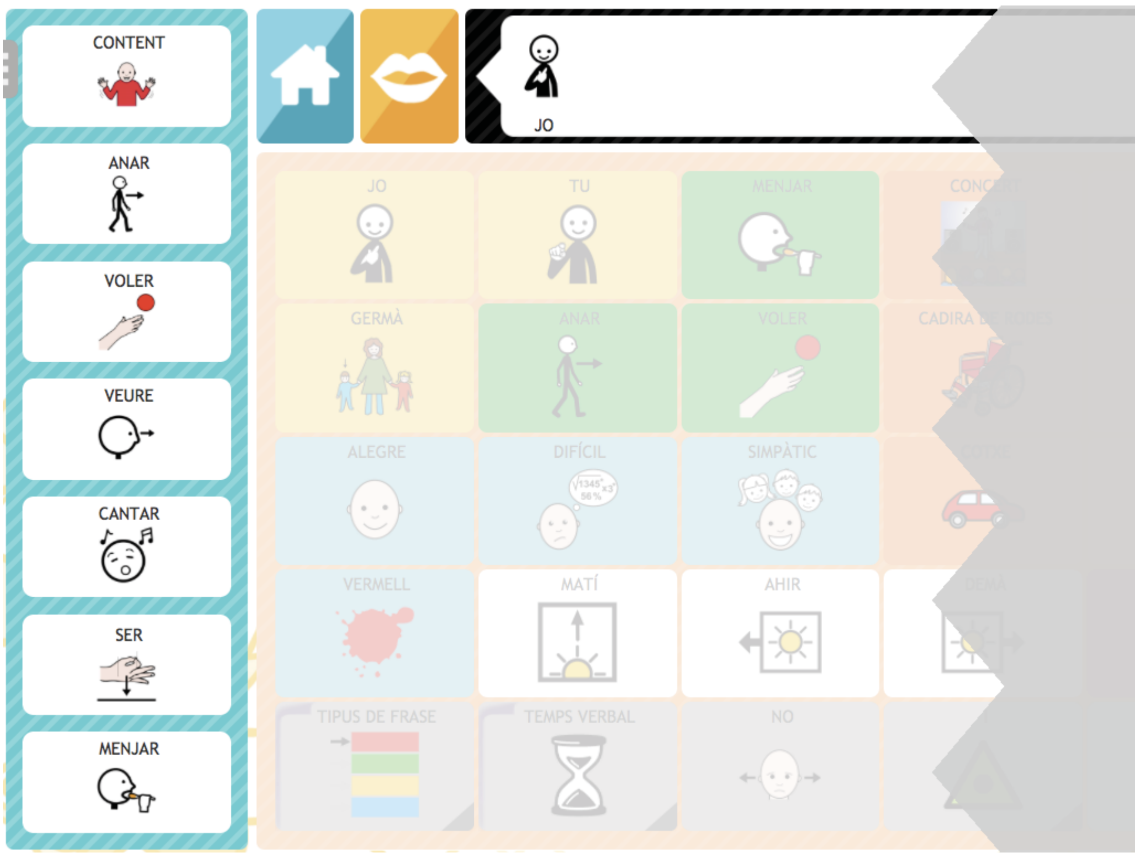


Figure 20: Predicted pictograms for Example 1 when there is 1 pictogram in the sentence bar

Example 1: Figure 20 shows the predicted pictograms when John has selected *I* as the first element of the sentence:

- *HAPPY*: it is the most common pictogram used after the pictogram *I*. John frequently uses this pair of pictograms (bigram) to say *I'm happy*.
- *GO*: it is the second most frequently used pictogram by John after the pictogram *I*.
- *WANT*: it is the third most frequently used pictogram by John after *I*.
- *SEE*: the most used verb by John the past two days is *GO*, as John is really looking forward to go to the concert. However, *GO* is already among the predicted pictograms, so the algorithm chooses *SEE* instead, as it is the second most used verb the past two days.
- *SING*: it is the third most frequently used verb by John the past couple of days.
- *BE*: it is the third most frequently used verb by John. The first two are *WANT* and *GO*, but they are already among the predicted pictograms.
- *EAT*: it is the most commonly used verb around 9am in the morning, as John will soon be eating his breakfast. Actually, the algorithm uses 3-hour spans to look for pictograms, in this case verbs, that are frequently used around a certain time of the day.

From these predicted pictograms, in this first example, the second one, *GO*, is chosen.

Example 2: Figure 21 shows the predicted pictograms when John has selected *WANT* as the first element of the sentence. In the example, John has omitted the subject, but the sentence will still have SVO structure, as the parser of the compansion system will add *I* as the subject of the parse tree. Here the algorithm of the prediction system also considers that the subject is already in the sentence.

- *GO*: it is the most common pictogram used by John after the pictogram *WANT*. John always wants to go to different places.
- *FANTA*: it is the second most frequently used pictogram by John after the pictogram *WANT*. John loves Fanta and is always asking for it.

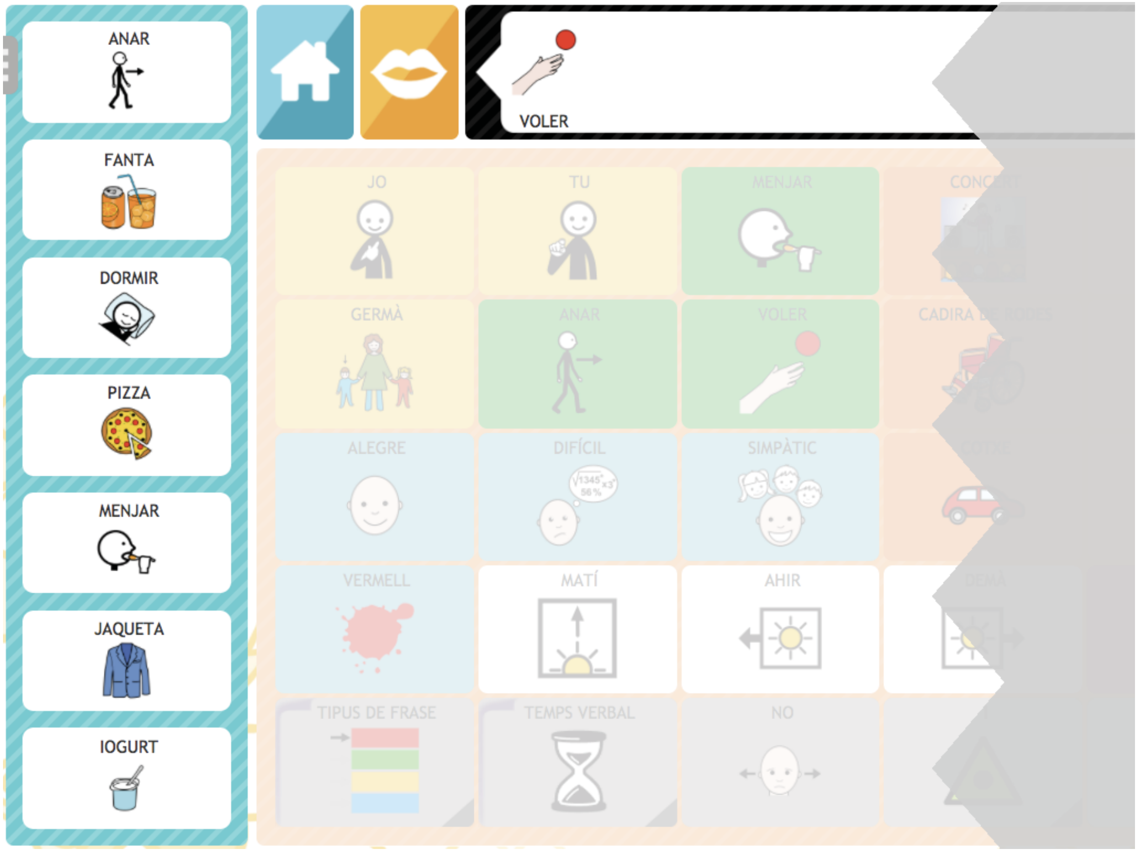


Figure 21: Predicted pictograms for Example 2 when there is 1 pictogram in the sentence bar

- *SLEEP*: it is the third most frequently used pictogram by John after *WANT*. John gets tired often and states when he wants to sleep.
- *PIZZA*: the verb patterns of the verb *to want* need either a noun or another verb as its theme slot, which is a mandatory slot in these patterns. Without including *GO*, *FANTA* and *SLEEP*, *PIZZA* is the likeliest word that may be used according to a weighted score calculated by the algorithm. This weighted score takes into account the total number of times that John has used a pictogram after *WANT* and also after any other pictogram, the number of times that it has been used around that time of the day and finally the number of times that it has been used that day of the week. Only pictograms that can perfectly fit the given slot (e.g. in our example the theme slot) are taken into account.
- *EAT*: it is the second likeliest pictogram that may be used that can fit the theme slot according to the weighted score. As seen before, around 9am John usually wants to eat.

- *JACKET*: it is the third likeliest pictogram that may be used that can fit the theme slot according to the weighted score. John is often cold and, thus, asks for his jacket.
- *YOGHURT*: it is one of John's favorite desserts also commonly used around this time of the day.

In Example 2, the second pictogram selected by John is *EAT*, which was also in the prediction bar.

Figure 22: Predicted pictograms for Example 3 when there is 1 pictogram in the sentence bar

Example 3: Figure 22 shows the predicted pictograms when John has selected *MY* as the first element of the sentence. In this example, the subject has not been selected yet, so the algorithm will still favor nouns that can act as subjects, as in the previous phase.

- *MOTHER*: it is the most common pictogram used by John after the pictogram *MY*.

- *CAT*: it is the second most frequently used pictogram by John after the pictogram *MY*.
- *WHEELCHAIR*: it is the third most frequently used pictogram by John after the pictogram *MY*. John's wheelchair often has technical issues and he talks about them.
- *CONCERT*: it is the most used noun in the past couple of days, although it does not fit very well with *MY*. The algorithm would need more information to be able to check the semantics in order to better filter these occurences.
- *BROTHER*: it is the third most used noun in the past two days. *CAT* is the second, but it is already selected.
- *POOL*: it is a noun that is frequently used on Mondays.
- *MILK*: it is a common word used around breakfast time at 9am. It does not fit well with *MY* either.

In Example 3, the second selected pictogram is *MOTHER*, also included in the prediction bar.

In the following section, we will continue with the description of the behavior of the algorithm now after a second pictogram has been selected in the sentence bar.

6.2.2.3. 2 pictograms selected in the sentence bar

In this phase, the algorithm changes depending on the type of word of the first two pictograms.

Example 1: Figure 23 shows the predicted pictograms when John has selected *I* and *GO* as the first two elements of the sentence. Here the algorithm works in a similar way as just seen with the second example sentence after the verb *WANT* was selected. Instead, in this example, the patterns of the verb *GO* are used.

- *POOL*: it is the most common pictogram used by John after the pair of pictograms *I-GO*.
- *PLAY*: it is the second most frequently used pictogram by John after the pair *I-GO*. John likes to go to play boccia and other sports.

- *HOME*: it is the third most frequently used pictogram by John after the pair *I-GO*. Whether he goes or does not go home, it is a combination that John frequently uses.
- *BEACH*: it is the second likeliest pictogram that may be used that can fit the location slot, which is the mandatory slot in the patterns of the verb *to go*, according to the weighted score. On weekends John sometimes goes to the beach with his family and afterwards talks about it on Monday at the daycare center. The likeliest pictogram was *POOL*, but it is already in the list of predicted pictograms.
- *CONCERT*: it is the third likeliest pictogram that may be used that can fit the location slot according to the weighted score. Remind that John will be going to a concert in a few days.
- *WORKSHOP*: from Monday to Friday John takes part in different workshops at the daycare center. Therefore, it is a common location used by him.
- *ASIA*: it is another frequently used location by John, as he has an uncle that works in Asia and one day he wants to visit him there.

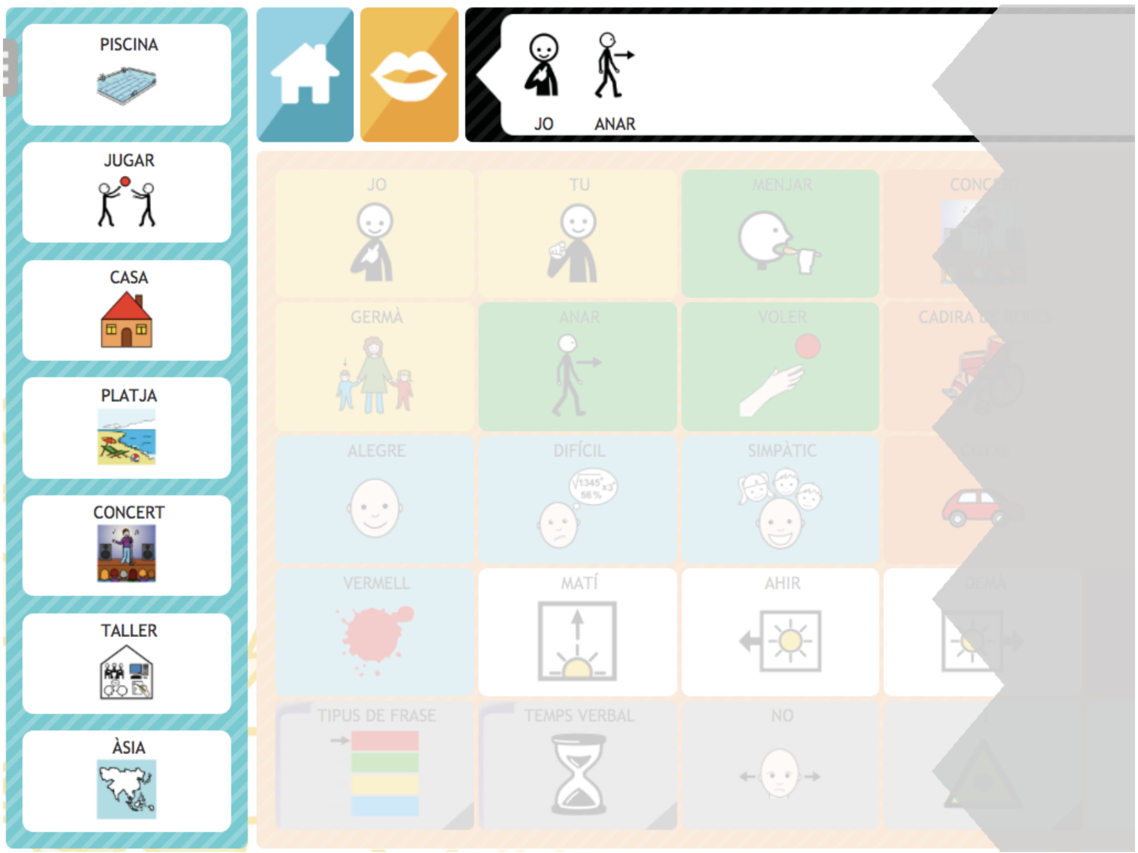


Figure 23: Predicted pictograms for Example 1 when there are 2 pictograms in the sentence bar

The third pictogram selected in Example 1 is *BEACH*, which was also shown in the prediction bar.

Example 2: Figure 24 shows the predicted pictograms when John has selected *WANT* and *EAT* as the first two elements of the sentence. In this example, even if there are two verbs in the input, now the algorithm only takes the patterns of the second verb into account.

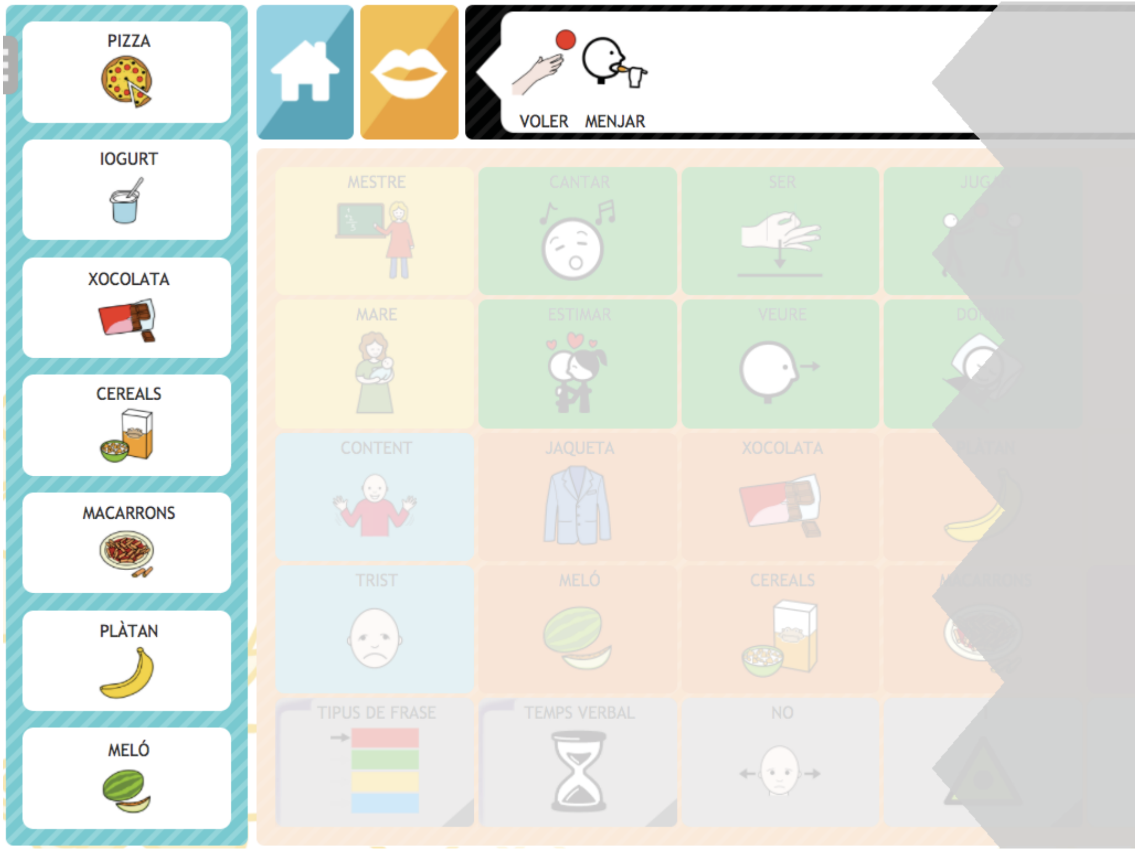


Figure 24: Predicted pictograms for Example 2 when there are 2 pictograms in the sentence bar

- *PIZZA*: it is the most common pictogram used by John after the pair of pictograms *WANT-EAT*.
- *YOGHURT*: it is the second most frequently used pictogram by John after the pair *WANT-EAT*.
- *CHOCOLATE*: another of John's favorite foods. It is the third most frequently used pictogram by John after the pair *WANT-EAT*.
- *CEREALS*: as it is breakfast time, *CEREALS* is the third likeliest pictogram that may be used that can fit the theme slot, which is the first optional slot in

the patterns of the verb *to eat*, according to the weighted score. The theme slot of *to eat* can only be perfectly fit by nouns of the food class. PIZZA and *YOGHURT* were the first and second likeliest pictograms, but are already in the list.

- *MACARRONI*: it is the fourth likeliest pictogram that may be used that can fit the theme slot according to the weighted score. It is also one of John's favorite foods. When there are two or more pictograms selected in the sentence bar, the weighted score takes into account both trios and pairs of consecutive pictograms along with the previously mentioned contextual and synctactic-semantic features.
- *BANANA*: another of John's favorite foods.
- *MELON*: even though it is a commonly eaten fruit where John lives, he hates melons so much that the *MELON* pictogram is often used by him in sentences like *I don't want to eat melon*. Thus the pair *EAT-MELON* is highly scored by the weighting algorithm.

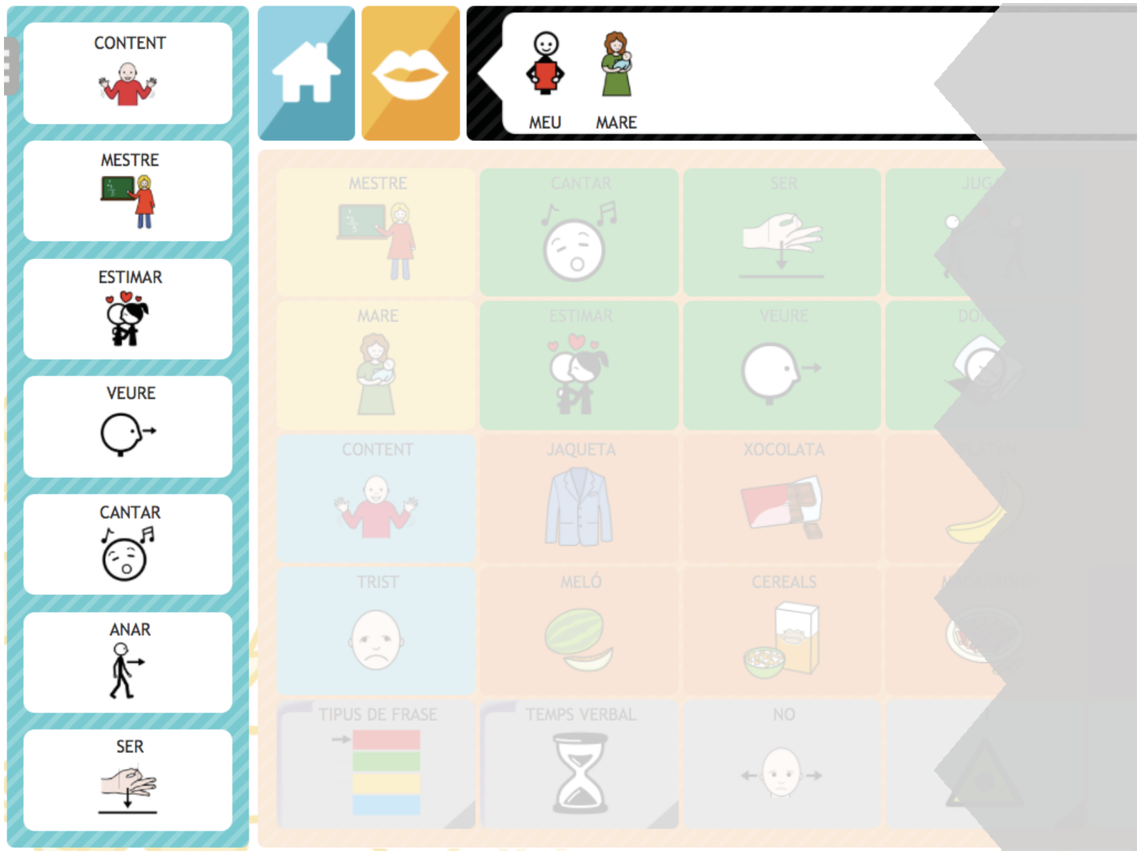


Figure 25: Predicted pictograms for Example 3 when there are 2 pictograms in the sentence bar

The third pictogram for Example 2 is *RICE*. This time the selected pictogram is not in the prediction bar. With this third selection, this example concludes with the sentence *I want to eat rice*.

Example 3: Figure 25 shows the predicted pictograms when John has selected *MY* and *MOTHER* as the first two elements of the sentence. This case is similar to the previously explained occurences where there was not a verb yet in the input.

- *HAPPY*: it is the most common pictogram used by John after the pair of pictograms *MY-MOTHER*. With this combination John says *My mother is happy*.
- *TEACHER*: it is the second most frequently used pictogram by John after the pair *MY-MOTHER*. To make the sentence *My mother is a teacher*, the first few months, John did not specify the verb *to be* and obtained mixed results from the compansion system. However, during the past couple of months, John is using the verb *to be* more. Nevertheless, due to its initial use, the trio *MY-MOTHER-TEACHER* still appears in the list of predicted pictograms.
- *LOVE*: It is the third most frequently used pictogram by John after the pair *MY-MOTHER*.
- *GO*: it is the most used verb by John the past two days.
- *SEE*: it is the second most frequently used verb by John the past two days.
- *SING*: it is the third most frequently used verb by John the past couple of days.
- *BE*: it is a common verb used by John. The pair *MOTHER-BE* is one of the most common combinations when John talks about his mother.

In this example, the selected pictogram after *MY* and *MOTHER* is *BE*; included in the prediction bar.

In the following section, we will end the description of the behavior of the algorithm for Example 1 and Example 3 after the third pictogram has been selected in the sentence bar.

6.2.2.4. 3 or more pictograms selected in the sentence bar

In this phase, the algorithm changes depending on the type of word of all the previously selected pictograms.

Example 1: Figure 26 shows the predicted pictograms when John has selected *I-GO-BEACH* as the first three elements of the sentence. In this case, as the location slot of the verb *GO* has already been filled, the weighted score considers the optional slots of the patterns of the verb, such as tool or transportation, company or time slots.

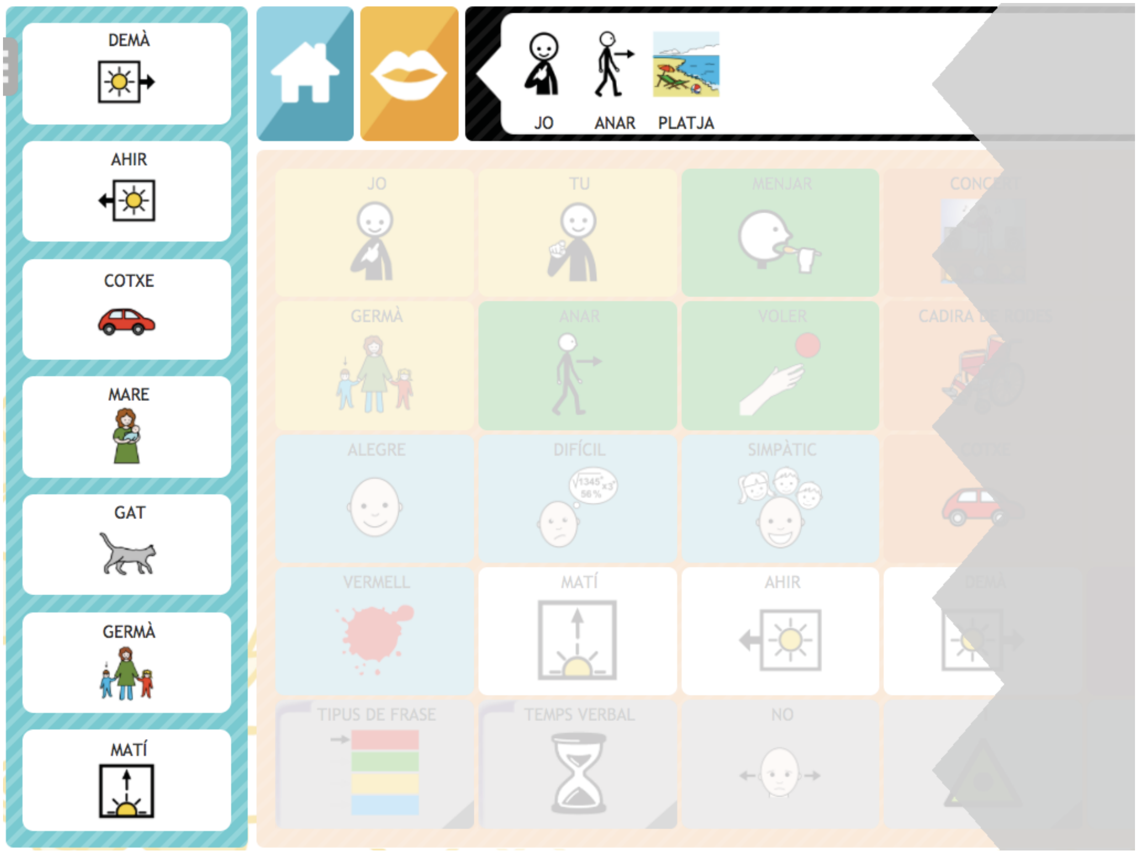


Figure 26: Predicted pictograms for Example 1 when there are 3 pictograms in the sentence bar

- *TOMORROW*: it is the most common pictogram used by John after the pair of pictograms *GO-BEACH*, as in *Tomorrow I'll go to the beach*.
- *YESTERDAY*: it is the second most frequently used pictogram by John after the pair *GO-BEACH*, as in *Yesterday I went to the beach*.
- *CAR*: it is the third most frequently used pictogram by John after the pair *GO-BEACH*, as in *We went to the beach by car*.

- *MOTHER*: it is the likeliest pictogram that may be used that can fit an optional slot according to the weighted score. In this case to make a sentence such as *I go to the beach with my mother*.
- *CAT*: it is the second likeliest pictogram that may be used that can fit an optional slot according to the weighted score. Here, even though *CAT* is one of the most used animate nouns lately, the prediction algorithm does not know that it is not common to go with cats to the beach.
- *BROTHER*: it is the third likeliest pictogram that may be used that can fit an optional slot according to the weighted score.
- *MORNING*: John usually goes to the beach in the morning with his family.

The final chose pictogram for this example is *YESTERDAY*, which was included in the prediction bar. Then, the resulting sentence is *Yesterday I went to the beach*.

Example 2: The construction process of the sentence already ended on the previous phase.

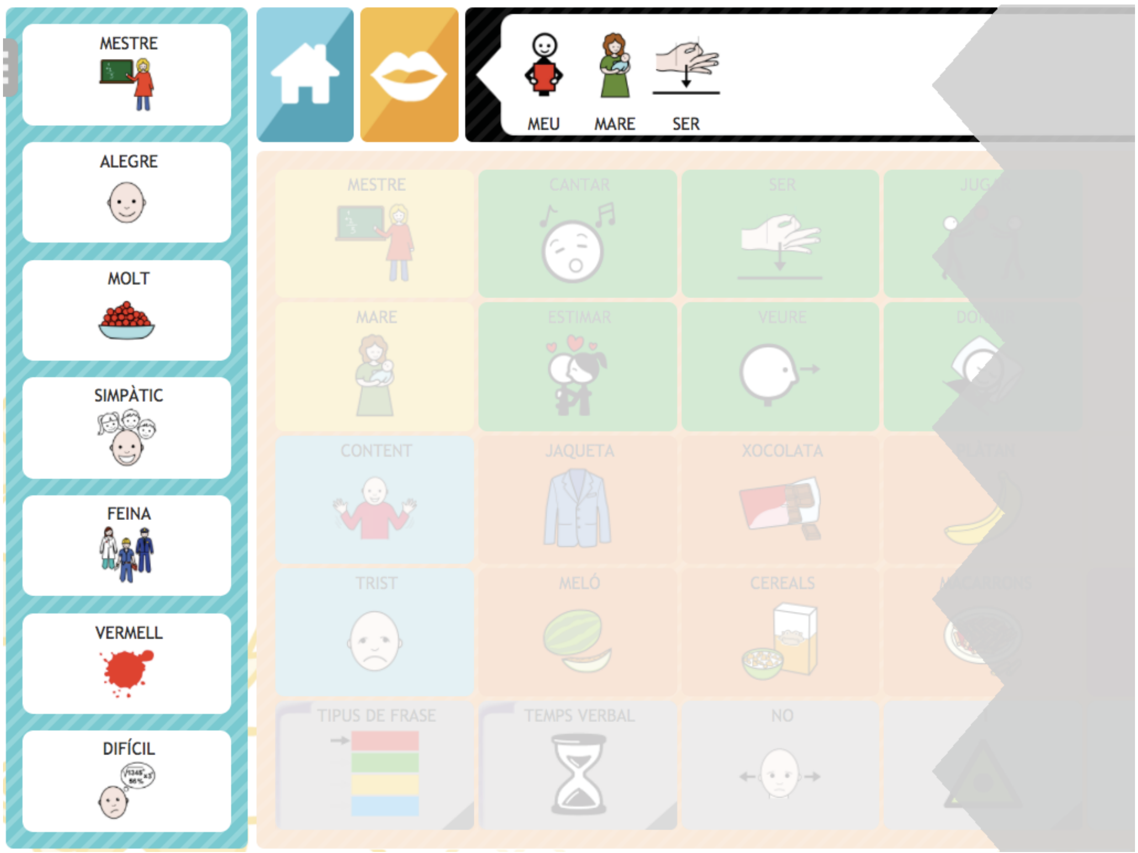


Figure 27: Predicted pictograms for Example 3 when there are 3 pictograms in the sentence bar

Example 3: Figure 27 shows the predicted pictograms when John has selected *MY-MOTHER-BE* as the first three elements of the sentence. In this case, notice that the patterns of the verb *to be* have several possibilities to fill the attribute slot (e.g. adjectives, profession class nouns, location class nouns, etc.).

- *TEACHER*: it is the most common pictogram used by John after the pair of pictograms *MOTHER-BE*.
- *CHEERFUL*: it is the second most frequently used pictogram by John after the pair of pictograms *MOTHER-BE*.
- *VERY*: it is the third most frequently used pictogram by John after the pair of pictograms *MOTHER-BE*, as in *My mother is very kind* or as in *My mother is very angry*.
- *NICE*: Having in mind that the attribute slot can be filled by several kinds of words, *NICE* is the likeliest pictogram that may be used that can fit the attribute slot according to the weighted score.
- *WORK*: it is the second likeliest pictogram that may be used that can fit the attribute slot according to the weighted score, as in *My mother is at work*.
- *RED*: it is John's favorite color. John uses it frequently with *BE* and it can fill perfectly well the attribute slot in some cases. However, in the current sentence it would not fit very well, although the prediction algorithm does not have the necessary contextual and semantic information to know it.
- *DIFFICULT*: John frequently uses it paired with *BE* to say that things are or are not difficult. It is a label that he likes to assign to the activities that he does. However, as with *RED*, in the current sentence it would not fit very well either.

Finally, on the third example, the final pictogram after *MY-MOTHER-BE* is *TEACHER*, resulting in *My mother is a teacher*.

CHAPTER 7

TESTS WITH PEOPLE WHO RELY ON AAC

This chapter describes the tests we conducted with people who were already familiar with pictogram-based augmentative and alternative communication (AAC) systems because they regularly used them to communicate. It begins by explaining the objectives of the tests, the methodology used and the profile of the people who participated. It then describes and quantifies the tests carried out. The next chapter presents the qualitative and quantitative results, followed by the conclusions.

Before conducting the tests with people who rely on AAC, we ensured that the compansion system for Catalan and Spanish was performing above our proposed performance threshold: 90% of sentences deemed "acceptable" and 80% of deemed "perfect". We set the threshold level based on Vaillant (1997), one of the few documented tests on compansion systems. With a vocabulary of 300 pictograms, Vaillant reported an 80.5% sentence acceptability rate in the technical tests, and unsatisfactory results in the tests with people who rely on AAC, due to their frustration with badly formulated sentences and the system's small vocabulary. Vaillant (1997) found that people who rely on AAC, who sometimes take more than 60 seconds to form a sentence, may give up on a system very quickly if it frequently forms sentences with errors. Based on those data, we set a sentence acceptability threshold of 90% for our system, i.e. errors in no more than one in every ten sentences allowed by the controlled grammar. The system performed well above the threshold level, with a level above 98%, as seen in the previous chapter.

7.1. Objective of the tests

The main objective of the tests with people who rely on AAC was to validate the research that we had conducted so far. Based on the PESICO template described by Schlosser, Koul and Costello (2006), we designed the necessary tests and questions that would enable us to assess the responses objectively and draw conclusions that would help AAC professionals to decide whether our system would be suitable for their clients. The PESICO template is used to formulate good questions using evidence-based practices (EBP) for AAC. It is derived from the PICO template (Richardson et al. 1995), which is commonly used in EBP in the field of medicine. With the PESICO template, to formulate good research questions one must take into account the person (P – in our study, participants in the tests whose results can be extrapolated to people who will use the system in the future), their environment (E), the stakeholders (S – including the person), the type of intervention (I – i.e. the tests or the future implementation of the compansion system, the prediction system and the graphical user interface), comparison (C) with the current situation, and the expected outcome (O). To validate our research, we created the following research questions, each with a series of sub-questions, which we later incorporated into the evaluation sheets for each test session.

1. **Does the compansion system meet the needs of people who rely on AAC?**
   By answering this question, we wanted to analyze to what extent the compansion system's restrictions in terms of sentence structures were suitable to users' needs. If possible, we wanted to address any restrictions and, if appropriate, eliminate them during the tests so that we could test the improvements. We also wanted to examine whether the system's default vocabulary was large enough and consider whether we ought to expand it so that users would not be left feeling frustrated so often.

2. **Does the compansion system reduce the time that people who rely on AAC need to form sentences or does it increase it? And does it increase their communication ratio?**

We wanted to see whether features of the system, such as adding default subjects and verbs and applying sentence modifiers, might affect the time needed to form sentences. Even if the compansion system did not make it quicker to form sentences, we needed to see whether the communication ratio increased anyway thanks to the insertion of function words (articles, prepositions, etc.) into the resulting sentence.

3. **Can the compansion system improve the language skills of people who rely on AAC?**

   We wanted to evaluate whether the compansion system improved the communication capacities of participants, since recent research (Ganz et al., 2014; Kagohara et al., 2013) has shown that many AAC technology applications are effective at teaching various communication skills to people with autism spectrum disorder (ASD). We therefore wanted to see whether this was the case for the participants in our study, who did not have ASD, but had varying degrees of cerebral palsy and intellectual disabilities. In particular, we wanted to see whether the compansion system enhanced participants' communication skills or helped them to structure their sentences better. In other words, we wanted to see whether the system encouraged users to communicate in a way that was more akin to natural language.

4. **Does the compansion system improve the communication experience of people who rely on AAC? Is it easy to use?**

   To answer these questions, we wanted to record the reactions and feedback of participants to obtain quantitative and qualitative data on whether they preferred to communicate using natural language, rather than using the telegraphic language obtained via pictograms. We also wanted to hear the views of the communication partners. Another aspect to evaluate was whether the compansion system was user-friendly, i.e. whether it had a steep learning curve. We wanted the compansion system to genuinely benefit the people who use it, rather than benefiting the technology. Light and McNaughton (2013) encapsulated this idea by saying that the design of AAC technologies should "minimize learning demands and maximize communication power for individuals with complex communication needs".

5. **Can the compansion system be used for purposes other than communication?**
   To find out whether the compansion system could be used for purposes other than that for which it was originally designed, we looked at feedback from participants, speech therapists and other workers at the center who were familiar with AAC. In addition to testing the prediction that the system could be useful in speech therapy sessions, in line with the third research question, we also wanted to see whether it could be used for other functions related to people who rely on AAC.

6. **Does the graphical user interface (GUI) that has been developed meet the requirements? Does the interface's prediction system reduce the time needed to form sentences?**
   Finally, it was very important to evaluate whether the GUI had the necessary accessibility functions to make full use of the other systems. If necessary, the tests could be used to debug the interface and make any necessary changes. In the interface, we also wanted to see whether using the pictogram prediction system in conjunction with the compansion system further reduced the time needed to form sentences.

Before the start of the tests, we understood that to draw firm conclusions for questions three and five, we needed more time than was available to us. Nevertheless, our intention was also to generate questions for future research topics and to get an idea of what the answers would be.

7.2. Methodology

For the tests, we planned weekly sessions for three months, i.e. 12 sessions per participant. When conducting tests with people who rely on AAC, it is strongly recommended to establish a routine (Hamilton & Kingsbury, 2006). The user profile section describes the profiles of the four participants and explains that we struggled to find more participants with the type of profile we needed.

We wanted to make the tests a collaborative process similar to the one described in Hamilton and Kingsbury (2006) with the participants, the center and the staff at the center all involved as much as possible. By collaborating with them, we would not only answer the research questions, but we would also focus on evaluating the proposed system. Each session therefore incorporated changes based on participants' observations in previous sessions, allowing them to see that their contribution was having real, tangible results, which made them more committed to the exercise and improved the systems being implemented.

Another important aspect we had to consider when designing the tests was the communication partners. As mentioned in Section 1.1 of the Introduction, communication partners play a vital role in AAC (Augé & Escoin, 2003). Therefore, we made the range of communication partners in the tests as similar as we could to the range of partners they would communicate with in the real world. The section on user profiles gives specific details on the communication partners.

7.2.1. Description of the sessions

We needed to test and evaluate the different systems in isolation, so we began by using the compansion system we had developed in the computer application Plaphoons. This free, open-source software is available in several languages (Catalan, Spanish, Galician, Basque, English, Portuguese and Arabic) and is used in many centers around Spain, especially by people who have just begun using AAC, but also by people who have more experience with AAC tools. Since our participants were already familiar with Plaphoons, it provided a benchmark for comparisons.

The main contribution of this thesis is the compansion system, so the first eight sessions focused entirely on testing that system in isolation, hence why we conducted the tests using Plaphoons. In the final four sessions, we tested the pictogram prediction system and the GUI that we had designed. These tests still used the compansion system – which was no longer new to the participants – so that we could compare the communication rate when the compansion system was used alone with the rate when the two systems (compansion and prediction) were used together.

The tests with the companion system alone were incremental. At first, the tests used participants' usual vocabulary, but at each session, as deemed fit, we introduced specific new features of the compansion system, such as tense modifiers, sentence modifiers, and gender and number modifiers, as well as new vocabulary. People who use pictogram-based AAC need time to learn how to use new communication features, so we could not introduce all the features in the first session. We introduced new features according to the language skills of each participant.

In order to evaluate objectively any improvements that the expansion system can make to pictogram-based communication, in addition to extracting quantitative indicators from the sentences created by participants in the tests, we also obtained qualitative evaluations from participants, their communication partners and the speech therapists at the center. Most professionals in the field conduct a quantitative assessment before making decisions related to AAC (Dietz et al., 2012). Several studies (Hustad et al., 2008; McNaughton et al., 2003) and lists of tools for AAC researchers (Creswell, 2002; Hedge & Pomaville, 2008; Hill & Harkawik, 2011) have shown that by combining qualitative and quantitative indices, we can build a more comprehensive picture of participants' experiences and listen to the contributions of professionals and other people who come into contact with the participants.

7.2.2. Quantitative indicators

We used a series of indicators to quantify some of the AAC parameters. The following section defines those indicators and defines the criteria we used to measure some of the parameters needed to calculate those indicators.

7.2.2.1. Communication rate

The communication rate (CR) is one of the main indicators used to evaluate communication speed in AAC and in other fields. It is defined as the number of words (W) produced in one minute and is expressed as words per minute (wpm).

$$CR = \frac{W}{M}$$

The communication rate can be used to compare the performance of a single user over a period of time or in different circumstances. In our study, we compared the communication rate of the test participants when using the compansion and prediction systems with their rates when using neither of these systems.

We calculated the communication rate for each of the sentences formed during the tests by timing how long it took each participant to form a sentence and counting the number of words in that sentence. The CR is not effective for comparing the performances of different users, however, since the CR depends on the selection rate.

7.2.2.2. Selection rate

The selection rate (SR) is the number of bits per second that the participant selects in the interface. It is calculated as follows:

$$SR = \frac{log_2(IE) * C}{S}$$

Where *IE* refers to the number of elements that can be selected in the interface and *C* is the number of clicks or selections made. If an interface has 32 pictograms, for instance, and a single selection takes place, the number of bits selected is 5. This value divided by the number of seconds (*S*) it takes to make the selections is the selection rate.

7.2.2.3. Rate index

"Different individuals have different selection rates. Also, different AAC systems have different numbers of 'keys' from which selections may be made. Selection rate influences communication rate. Therefore, compensation for selection rate differences is necessary for the comparison of communication rates to be clinically useful. The solution to this problem is the use of what is being defined as the rate index." (Hill & Romich, 2002)

As explained in the excerpt above, the rate index (*RI*), measured in words per bit (*wpb*), was created to compensate for differences in the communication rates among different users. Using the *RI*, we could compare different users using similar or different systems, or compare a single user in different circumstances. The rate index is expressed using the following formula:

$$RI = \frac{CR}{(SR/60)}$$

7.2.2.4. Word-click rate

The word-click rate (*WCR*) is a similar indicator to the rate index, but it measures the number of words formed from each click. This indicator is slightly more intuitive and better reflects the efforts of the participants. In the tests, our four participants all had an average number of bits of between 4.17 (participant A) and 5.36 (participant D) on their pictogram panels, so we believe that it was useful to compare the number of words per click (wpc) by each participant, since this indicator compensates for differences in participants' selection rates, too. The *WCR* is calculated as follows:

$$WCR = \frac{CR}{C/(S/60)} = \frac{W/(S/60)}{C/(S/60)} = \frac{W}{C}$$

7.2.2.5. Number of clicks

This indicator measures the number of switch presses, screen touches, or mouse clicks used to form each sentence. To compare participants using a switch with those using a touchscreen, we calculated two click values: finger clicks and switch clicks. For the participants using a switch, we compared the number of clicks they needed to form the same sentences with a switch and using a touchscreen.

7.2.3. Model for comparing results with and without the compansion system

By comparing the process of producing the same sentences with and without the compansion system, we can obtain an important indicator of whether the system creates an extra burden to the users or not. This is important since our system has features that are not straightforward telegraphic communication. As seen in the description of the compansion system, some modifiers can be used in order to solve ambiguities or in order to build specific sentences. These modifiers may carry an extra burden in the time needed to plan the sentence or in the time needed to input them, but, at the same time, some of them may save the input of other words. For instance, if a user wants to talk about computers in general, as in "*Computers are useful*", the plural modifier will be needed, otherwise the system will produce "*The computer is useful*". On the other hand, if a user wants to say "*Please, could you give me an apple?*", this sentence can be achieved with an input of *GIVE-APPLE* and the imperative modifier, thus saving the input of the subject *you*, the receiver *I* and the expression *please*. To sum up, we wanted to see whether, overall, the rates, the time and the number of clicks needed to form a sentence with the compansion system were higher or lower than without it, bearing in mind that the semantic content was the same in both cases.

Figure 28 compares how sentences are formed with and without the compansion system. The left side shows how a sentence is formed using pictograms with the compansion system (P_{wcomp}). Inputting these pictograms requires a certain number of clicks (C_{wcomp}) and a certain amount of time (S_{wcomp}). These pictograms pass through the compansion system, which generates a sentence with a certain number of words (W_{wcomp}).

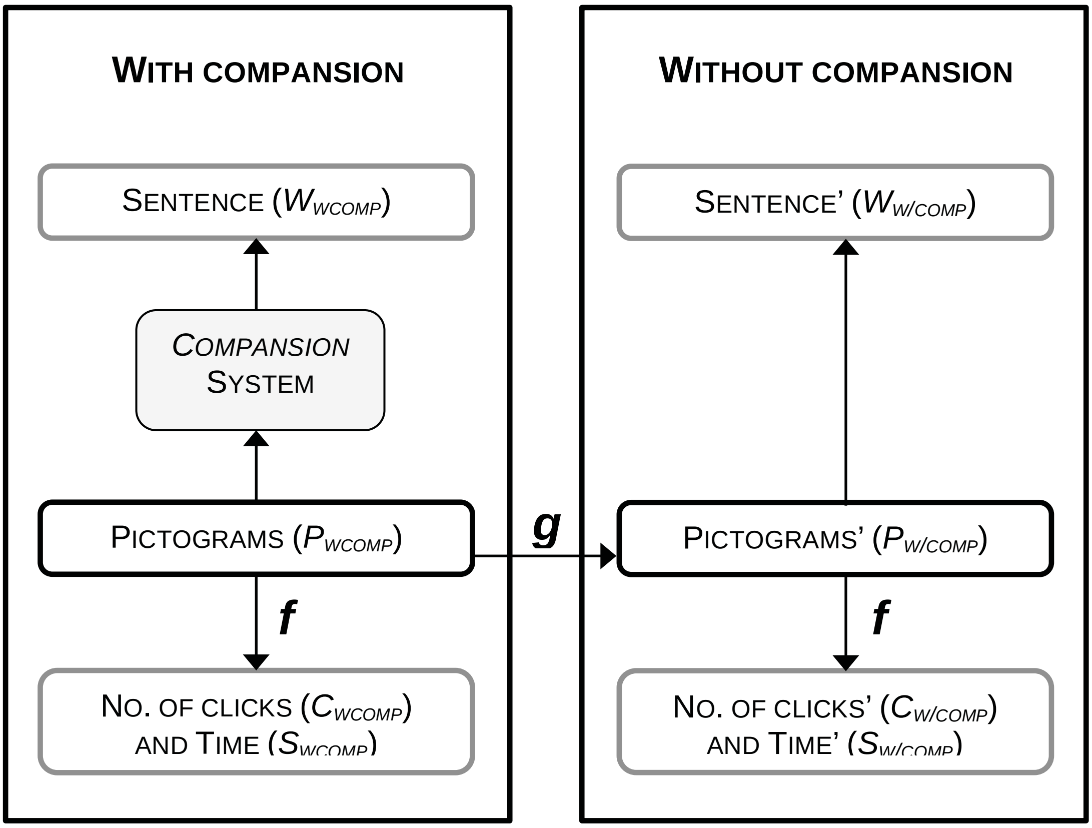


Figure 28: Flow and functions for comparing test results with and without the *compansion* system

The right side shows how a sentence is formed using pictograms without the compansion system ($P_{w/comp}$). Inputting these pictograms requires a certain number of clicks ($C_{w/comp}$) and a certain amount of time ($S_{w/comp}$.). In this case, the pictograms do not pass through any system, and they directly become a sentence with a certain number of words ($W_{w/comp}$). Therefore, $P_{w/comp} = W_{w/comp}$.

To compare these two ways of forming sentences, ideally we would have needed to form each sentence in the tests twice, once with and once without the compansion system. AAC has a much slower communication rate than natural speech, especially when selection is not done directly (Horstmann Koester & Arthanat, 2017), as was the case for three of our four participants, who used scanning methods to form each sentence. Those who use AAC need a long time to form each sentence, and because of their physical and intellectual disabilities, it is also strenuous for them. Forming each sentence twice would have required twice as much time and twice as much effort, and participants would probably have become disinterested and mentally fatigued. Also, we would have generated only half as much data, or perhaps less, which would have affected our results. Worse still, the tests may have failed had participants become disinterested or mentally fatigued. We therefore rejected repeating sentences from the outset.

Instead, we chose to compare results as shown in the above figure, where the left side refers to experiments conducted in sessions with participants and the right side refers to sentences formed without the compansion system, modeled using the function *g* described below.

7.2.3.1. Description of the function *g* in the figure

Function *g*, depending on the number of pictograms (including modifiers) inputted with the compansion system (P_{comp}) and the context for a given sentence, gives the number of pictograms that would have been necessary to produce the same sentence without it ($P_{\text{n\_comp}}$).

We modeled the function *g* according to two cases: where the sentence does not use modifiers or other features of the compansion system, and where the sentence does use these modifiers or other features.

In the first case, where the sentences do not use modifiers to help form sentences, the compansion system is transparent to the user, and the only thing that changes is the resulting sentence. In these sentences, the function *g* of the figure is the identity function, therefore $P_{wcomp} = P_{w/comp}$.

Concerning the number of clicks (C) and the amount of time (S) needed to build the sentences, in the device's interface, the modifiers of the compansion system did not affect the placing of the rest of the pictograms as they were in separate panels. The only modification needed was an extra row in some of the user's panels with links to these extra panels of modifiers (panels for sentence-type modifers and for verb tense modifiers). In our tests, this extra row did not imply extra clicks, as the scanning method used by our participants that used scanning was automatic and they did not have to press in order for the higlighted area to advance. As for extra time, as explained later in Section 7.2.3.3, in sentences where modifiers were used, there was no extra planning time observed.

All in all, this also suggests that in sentence where $P_{comp} = P_{n\_comp}$ the number of clicks and the amount of time needed to build a sentence with and without the compansion system were the same ($C_{\text{comp}} = C_{\text{n\_comp}}$ and $S_{\text{comp}} = S_{\text{n\_comp}}$). In the tests, 79% of the sentences were of this kind. Therefore, function *g* was only used in the remaining 21% of the sentences.

In the second case, where the sentences use modifiers or other features of the compansion system (e.g. subject or verb reductions), P_{wcomp} and $P_{w/comp}$ are not identical. To go from P_{wcomp} to $P_{w/comp}$, we did the following:

- **For verb subjects:**
  It was assumed that the default subject for all sentences was the first person. Where a different subject was used, we did not count it if the context made it clear and it did not need to be inputted. In other words, for pictograms that act as subjects, in most cases, we counted the pictograms with and without the compansion system equally (i.e. $P_{w/comp} = P_{wcomp}$) to avoid boosting the rates with the compansion system. This decision actually reduced the compansion system rates a little, since the system for Catalan and Spanish combines the subjects "jo" (I) and "tu" (you) with the verb to form a single word in the form of a conjugated verb. For example, when converted into natural language, "jo-menjar" (I-eat) in Catalan and "yo-comer" (I-eat) in Spanish become a single word: "menjo" in Catalan and "como" Spanish. In such cases, the communication rate is lower with the compansion system.

The only cases where $P_{w/comp}$ and P_{wcomp} are not identical are the following:

Where the compansion system inputs a default subject other than "jo" (I) – which occurs by default with the sentence type, imperative and question modifiers – and in the sentence without the compansion system the subject is not implicit, then $P_{w/comp} = P_{wcomp} + 1$.

Where the participant has specified a subject to override the system's default behavior and in the sentence without the compansion system it is not necessary to input the subject because it is implied by the context, then $P_{w/comp} = P_{wcomp} – 1$.

In both cases, application of the function *g* is not automatic, since it depends on context, so we performed the analysis manually.

- **For modifiers of gender and number and for the conjunction "and" ("i" in Catalan, "y" in Spanish):**

  These modifiers are not normally used in pictogram-based AAC, so even where they were used with the compansion system, they were not counted for the sentences without the system. In these cases, $P_{w/comp} = P_{wcomp} – 1$ for each of these modifiers present in the sentence.

- **For tense modifiers:**

  Sentences with tense modifiers are those where $P_{w/comp} = P_{wcomp}$, since pictogram-based AAC also gives users with better language skills the option to specify verb tenses such as "past" and "future" with a pictogram. The difference is that the resulting sentences use unconjugated verbs accompanied by the "past" and "future" pictograms. For instance, if the user wants to say that he will go to the swimming pool and inputs the pictograms "anar-piscina-futur" (go-pool-future), the sentence will come out as "anar piscina futur" ("go pool future"). In this sentence, $P_{w/comp} = 3$ and $W_{w/comp} = 3$.

- **For sentence-type modifiers:**

  For each sentence-type modifier in the compansion system, we used the following criteria:

  Desire modifier: We assumed that to form sentences without the compansion system that have the same semantic content, with no ambiguity, it was necessary to input the verb "voler" (want). Therefore, in these cases, $P_{w/comp} = P_{wcomp}$.

Permission modifier: We assumed that to form sentences without the compansion system that have the same semantic content, with no ambiguity, it was necessary to input the verb "poder" (can). Therefore, $P_{w/comp} = P_{wcomp}$.

Negation modifier: This means inputting the pictogram for the Catalan and Spanish equivalent of the negative particle "not", therefore $P_{w/comp} = P_{wcomp}$.

Conditional modifier: This means inputting the pictogram for the Catalan and Spanish equivalent of the conditional conjunction "if", therefore $P_{w/comp} = P_{wcomp}$.

Imperative, question, answer, exclamation and declaration modifiers: We assumed that, without the compansion system, adding a pictogram is not necessary because each of these modifiers is implied by the context. Therefore, in these cases, $P_{w/comp} = P_{wcomp} - 1$.

- **For verbs that the compansion system automatically inputs:**
  For default verbs that the system automatically inputs because they are associated with certain nouns, we considered that in sentences formed without the compansion system, it was necessary to add those verbs to create the same semantic meaning, therefore $P_{w/comp} = P_{wcomp} + 1$. However, for the default verbs "ser" and "estar" (both meaning "to be" in both Catalan and Spanish), if they appeared alongside an adjective we decided that, as copulative verbs, they did not carry any semantic weight, therefore $P_{w/comp} = P_{wcomp}$.

- **For expressions that the compansion system automatically inputs:**
  Currently, the only expression that the system ever inputs automatically is "si us plau", the Catalan equivalent of the English adverb "please". The system adds the expression to sentences that use the "request" sentence-type modifier and to sentences that are affirmative commands. For these sentences, we considered that to obtain the same meaning without the compansion system, it was necessary to input the pictogram for "si us plau", therefore $P_{w/comp} = P_{wcomp} + 1$.

- **For receiver case frames automatically added by the compansion system:**
  We implemented similar criteria to those used for default subjects. If the compansion system automatically inputs a receiver, and the sentence without the compansion system only makes sense if the receiver must be specified, then $P_{w/comp} = P_{wcomp} + 1$. Where the participant has specified a receiver to override the system's default behavior and in the sentence without the compansion system it is not necessary to input the receiver because it is implied by the context, then $P_{w/comp} = P_{wcomp} - 1$. In all other cases, $P_{w/comp} = P_{wcomp}$.

The numerical values of P_{wcomp} and $P_{w/comp}$ in all the sentences formed by the participants during the tests are listed in Digital Appendix W.

7.2.3.2. Description of the function *f* in the graph

In addition to the function *g*, Figure 28 also has the function *f*, which is shared by both the side in which the compansion system is used and the side in which it is not. Based on the pictograms and their distribution in the GUI, this function counts the number of clicks needed to generate the sentence.

This function counts not only the clicks needed to select the pictograms, but also all the clicks required to read the sentence, delete the sentence previously formed, and, if necessary, return to the main user panel (2-3 finger clicks or 6-9 switch presses). This decision may hinder comparisons between our system and future studies if the designers of the latter decided not to count these preparation clicks needed before the next sentence, but it does not affect comparisons between different participants in this study, since we used the same criteria for each participant. Other settings could have made these clicks unnecessary, but we decided to use our participants' usual settings for the tests.

7.2.3.3. Method for calculating the time required to form sentences without the compansion system

The *g* function converts P_{comp} to P_{n_comp}, while the *f* function calculates C_{comp} and C_{n_comp}. To calculate the the time required to form sentences without the compansion system (S_{n_comp}), we used a linear conversion from S_{comp} based on the number of clicks (C_{comp} and C_{n_comp}) required to access each pictogram. In the device’s interface, modifiers did not affect the placing of the rest of the pictograms as they were in separate panels, so their placement did not influence the conversion. In order to make sure that the use of the special features of the compansion system did not have an effect either on the planning time and the execution time of the users, we calculated the number of clicks per minute: *(a)* When users built the 79% of sentences that did not use any of the features of the compansion system (6.02*cpm*); *(b)* In the 21% of sentences that did use these features (6.61*cpm*). The values (in parentheses) show that the pace to build sentences using the special features of the compansion system was a bit higher. This suggests that these features did not slow the planning or the execution time, thus the linear conversion seems a valid option.

Therefore, to calculate the time required to form the sentences without the compansion system (Sn_comp), we used the following expression:

$$S_{n\_comp} = \frac{S_{comp} * C_{n\_comp}}{C_{comp}}$$

7.2.4. Session evaluation sheets

We designed an evaluation sheet so that we could store data from the sessions and record participants' evaluations of those sessions. Appendix C shows a blank evaluation sheet. The sheets dealt with the following aspects:

1. **General data about the session**: session length, whether the session ended early, and if so, the reason why. We believed it was important to obtain this information, since the state of health and well-being of people with severe disabilities varies considerably. It is important to bear this in mind when interpreting isolated results from certain sessions where a participants' state of health or well-being was not so good. By doing so, we could detect outlying data generated in those sessions or take participants' state of health into account in qualitative assessments.

2. **End-of-session evaluation**: numerical indicators of the participant's and the speech therapist's satisfaction at the end of the session, emotions that the speech therapist detects in the participant (happiness, surprise,

frustration, excitement, etc.), comparison with the previous session, and other comments on how the session went.

3. **Subsequent evaluation of the sentences generated during the session**: average number of pictograms used, average time taken by the participant to form a sentence, the most frequently used features of the compansion system (subject omission, verb omission, tenses, sentence-type modifiers, etc.), comparison with the previous session (such as changes to sentence structures) and other comments.

4. **Session sentences**: pictograms used and the final sentence formed, the participant's quantitative numerical satisfaction with the resulting sentence, and the time it took the participant to form the sentence. When the pictogram prediction system was used during the final sessions, we also recorded the pictograms that participants selected directly from the prediction bar. Finally, for each sentence, in addition to the information collected in the evaluation sheet, we also had access to the information stored in the application's database: all the pictograms and modifiers used in each sentence, the parse tree, the end result produced by the generator, and the statistics used by the pictogram prediction system at any given moment.

7.2.5. Sentence types

We designed three methodologies to record the sentences during the sessions. We introduced the three communication situations gradually over the course of the sessions, switching back and forth between them as per the needs and the pace of each participant:

1. **Prepared sentences (Fx)**: Participants were given a series of sentences formed from the vocabulary on their panel. These sentences had several aims:
    - To establish a benchmark for the amount of time participants took to form a sentence, from which we calculated a benchmark communication rate.

- To allow participants to see natural-language versions of sentences that they regularly produced using telegraphic language so that we could observe their reactions and ask them to evaluate the natural-language sentences.
- To gradually introduce specific features of the compansion system, such as gender and number modifiers, verb tenses, sentence modifiers (each adapted to the language skills of each participant), and special features of the system such as subject omission, pictogram reordering and the inputting of default verbs. Because we introduced these features in a controlled manner, we could learn about how they were used and what the expansion system could achieve.
- To gradually incorporate new vocabulary and other pictogram types, such as question words, onto the panels.
- To repeat sentences produced in previous sessions to remind participants of them and allow them to build on the knowledge they had acquired of the system.
- To compare how participants reacted when sentences were expanded correctly and incorrectly. We achieved this by adjusting the system in two or three sessions so that it would form incorrect sentences (such as by choosing the wrong subject, preposition or article, etc.).

2. **Guided conversations (Fr)**: The guided conversations used questions that we had prepared in advance. These questions had the following aims:
    - To simulate situations from participants' daily lives.
    - To allow participants to build on their ability using the compansion system, which they had been introduced to with the prepared sentences.
    - To keep repeating questions we had posed during the guided conversations in previous sessions to see whether participants had improved the structure of their answers or the way they produced their answers.
    - To make it easier to hold a conversation with certain participants who struggled to start conversations in many sessions, either because

their communication tools were not adapted to their linguistic limitations caused by their intellectual disabilities or severe physical disabilities, or simply because they were not used to starting up conversations.
- To record more sentences for subsequent analysis.

3. **Spontaneous conversations (Fr)**: The spontaneous conversations gave participants the freedom to express themselves and to reproduce fragments of conversations, either with people they met only recently, with conversation partners whom they did not know, or with conversation partners whom they already knew. The aims of these conversations were as follows:
   - To encourage participants to start conversations themselves and to ask their partners questions, especially those with the most severe intellectual disabilities, since they tended not to actively initiate conversations.
   - To keep a record of the resulting questions posed by participants.
   - To foster communication and conversation among different participants, which is rare in such settings.
   - To make the sessions more enjoyable.

Finally, it is important to mention that for the guided and spontaneous conversations, we took into account the observation made by Reichle et al. (2016) that certain types of responses are easier to give using certain modes of communication. We therefore needed to avoid asking yes–no questions, since participants would find it easier to respond to such questions using a gesture rather than pictograms, thus preventing us from obtaining any information with which to evaluate the compansion system. For obtaining feedback, however, we did need to use mainly yes–no questions.

7.2.6. Participant profiles

AAC test environments use few participants. In the publication guide for the first-quartile indexed journal *Augmentative and Alternative Communication*, for instance, the "Aims & Scope" section says that the journal publishes "both group and single case research designs."[37]

In our specific case, we needed to find participants who were familiar with pictogram-based AAC. Ideally, we were looking for participants who had been using these systems for a long time and had a stable linguistic profile. By using these kinds of participants, we would know that any developments they made during the trials would be thanks to our system, rather than thanks to the additional experience they were acquiring. Sessions were also scheduled during their weekly speech language pathology sessions, so the results were not biased due to extra practice. Finally, we were also ideally looking for participants who were familiar with the Plaphoons software, and we wanted a range of cognitive profiles that would reflect the range of people who might use our compansion system.

To conduct the trials, we needed to find a center for people who rely on AAC that was enthusiastic about participating in the project and had a speech therapist who was also enthusiastic. After analyzing several centers for people who rely on AAC, we decided to conduct the tests at Prodis[38] (Prodiscapacitats Fundació Privada Terrassenca), a non-profit organization in the Catalan city of Terrassa that provides comprehensive assistance and support to adults with intellectual disabilities, mental illness and cerebral palsy. We conducted the tests with four adult volunteers who met all our requirements. All four had at least 30 years' experience using pictogram-based communication, were familiar with Plaphoons and had different degrees of cerebral palsy and intellectual disabilities. Three of the participants (A, B and C) communicate in Catalan, while the fourth (D) does so in Spanish, but also understands Catalan.

Table 25 shows the various user profiles in more detail, based on data obtained from the center's speech therapist, who helped with the tests and has more than 15 years' experience in AAC. Before starting the tests, we met with the speech therapist to find out as much as possible about the profiles of the participants and

---

[37] Augmentative and Alternative Communication Journal Guidelines (http://www.tandfonline.com/action/journalInformation?show=aimsScope&journalCode=IAAC20)

[38] Prodis Terrassa (http://prodis.cat/ca/)

to draw up the methodology explained above. We wanted to know which vocabulary and panels they normally used when communicatiing so that we could incorporate them into our system. Table 25 also shows the number of panels that the participants used and their vocabulary size. Digital Appendix Y shows some of the panels of the participants in the tests. It might seem surprising that Participant A, with the most severe intellectual disabilities, initially had a broader vocabulary than participants B and C, but that was because Participant A was in charge of announcing the lunchtime menus at the center, which meant that her 184 pictograms included 43 food items that were not on the panels of participants B and C.

Participant D, who had far more vocabulary than the other three, had his words grouped by grammatical or semantic category (people, places, food items, verbs, adjectives, etc.). Participants A, B and C, on the other hand, had their vocabulary grouped primarily based on locations, i.e. vocabulary used in the workshop, in the home, etc. Although these three users did have some words grouped by categories (e.g. friends, supervisors, menus), none of their words were grouped by grammatical categories such as verbs and adjectives. The vocabulary distributions were consistent with each participant's language skills.

In addition to the participants and the speech therapists, each session was also attended by three or four communication partners, each of whom had worked with AAC and with each participant for different amounts of time: either not at all, for a few months, for 3-4 years, or for more than 15 years. These data are summarized for the five communication partners in the table below.

Communication partner	**I_1**	**I_2**	**I_3**	**I_4**	**I_5**
Years of experience with AAC	18 years	24 years	5 years	2 months	0 months
Years since meeting participants	15 years	8 years	4 years	0 months	0 months

Table 24: Communication partner data

	Participant A	Participant B	Participant C	Participant D
Physical disability	Cerebral palsy Spastic quadriplegia	Cerebral palsy Spastic quadriplegia	Cerebral palsy Dystonic quadriplegia	Cerebral palsy Spastic quadriplegia
Intellectual disability	Severe	Moderate	Moderate	Mild or none
Reading/writing skills and Language produced	No. Unstructured sentences using 1 or 2 pictograms Usually to answer questions.	No. Unstructured sentences using 1 or 2 pictograms Often uses sounds or gestures rather than pictograms.	No. Sentences with little structure, using 1-3 pictograms, in response to questions or to explain events that have occurred.	No. Complex, structured sentences using 3-8 pictograms for all kinds of communication situations.
Age and Experience with pictogram-based AAC	45 years old. Since elementary school.	44 years old. Since elementary school.	43 years old. Since elementary school.	47 years old. Since elementary school.
Panels/pictograms used[39]	15 / 184	13 / 143	13 / 133	34 / 791
Av. picts. per panel	12 (4 to 25) + 2 menus	11 (3 to 25) + 3 menus	10 (3 to 25) + 3 menus	23 (6 to 40) + 4 menus
Vocabulary distribution	By locations and semantic categories	By locations and semantic categories	By locations and semantic categories	By semantic and grammatical categories
Computer access	Direct selection with the hand	Using a head switch, with auto-scanning at 5500ms	Using a head switch, with auto-scanning at 5500ms	Using a knee-operated switch, with auto-scanning at 4500ms
Personality traits	Highly sociable. Depending on her mood, communication can be difficult.	Very sociable. Struggles to recognize some symbols on his panels.	Highly sociable. Likes to ask about others, despite having no question words on panels.	Very sociable. Likes to communicate, share ideas and suggest activities.

Table 25: Summary of the profiles of those who participated in the tests

[39] Total number of pictograms on all panels used by the participants before beginning the tests, excluding repetitions and the "delete" and "read aloud" pictograms.

7.3. Tests conducted

This section describes the content of the tests and the data generated.

In total we conducted 40 tests (11 with participants C and D, and 9 with participants A and B, who missed two sessions for medical reasons) across 11 sessions. There are therefore 40 evaluation sheets from the 11 sessions. The sessions were conducted with three tablets running the Windows operating system. The participants formed 333 sentences in total, of which 238 were prepared sentences (Fx) and 95 were unprepared (Fr). Table 26 shows a summary of the total number of sessions and sentences formed. Table 27 shows the distribution of the 40 tests across the different sessions and among the four participants. It also shows a summary of the number of sentences formed in each session and by each participant. Values marked with an asterisk (*) in this second table refer to sessions that participants were unable to attend or that were shortened due to time constraints or a participant feeling unwell.

	Compansion system alone	Prediction and compansion systems	Total
Tests conducted	27	13	**40**
Evaluation forms	27	13	**40**
Sentences formed	252	81	**333**
Prepared sentences	182	56	**238**
Unprepared sentences	70	25	**95**
With ≤ 2 pictograms	111	16	**127**
With ≥ 3 pictograms	141	65	**206**
Incorrect sentences	10	1	**11**

Table 26: Summary of the total number of sessions and sentences formed by participants.

7.3.1. Procedure for the test sessions

Each session, except those that ended early, lasted between 30 and 90 minutes per participant. The sessions combined the various methodologies described in

Section 7.2.5 and all the data were recorded for every sentence that was formed. At the end of each session, we entered the participants' qualitative evaluations on sections 2 and 3 of the forms (see Section 7.2.4) and noted down improvements we could make to the systems based on the observations made during the session by us, the speech therapist, the other communication partners and the participants. The Y Digital Appendix contains a computerized transcript of all the forms filled in during all the sessions.

Session no.	Participant A	Participant B	Participant C	Participant D	Total
1	6	8	3*	16	33
2	8	-*	9	11	28
3	8	10	9	11	38
4	5	9	9	15	38
5	7	9	10	11	37
6	-*	11	11	9	31
7	7	13	11	7	38
8	-*	-*	7	7	14
9	9	7	5*	7	28
10	6	7	5*	7	25
11	6	5	4*	8	23
Total	**62**	**79**	**83**	**109**	**333**

Table 27: Number of sentences formed per participant and per session

At subsequent meetings during the week, we would work with the speech therapist to prepare the guided sentences and conversations for the next sessions. We would take the specific capabilities of each participant into account when determining aspects such as sentence complexity, new modifiers, new vocabulary, new panels, and the redistribution of existing panels. Table 28 compares these aspects for each participant from the first to the final session. Note that in the first session, no specific features of the compansion system had yet been added, so none of the participants had any pictograms for modifiers, tenses or sentence types at that point, hence the 0 values in the "Start" columns in the table.

	Participant A		Participant B		Participant C		Participant D	
	Start	End	Start	End	Start	End	Start	End
Panels available	15	16	13	15	13	15	34	38
Pictograms (pcs)	184	201	143	191	133	186	791	857
Modifier pcs	0	1	0	1	0	1	0	6
Tense pcs	0	0	0	3	0	3	0	5
Sentence-type pcs	0	2	0	2	0	2	0	9
Prediction pcs	0	4	0	5	0	7	0	7

Table 28: The number of panels and pictograms that each participant had at the start and end of the tests

Finally, as explained above, different sentence types were introduced gradually as the sessions progressed. Table 29 shows an overview of the different sessions.

Sessions	Sentence types	Software used	System evaluated
1 to 2	- Prepared sentences (for learning how to use the system)	Plaphoons & Compansion system	Compansion system
3 to 4	- Mainly prepared sentences - Guided conversations - First spontaneous conversations with the participant with the least severe intellectual disabilities (Participant D)	Plaphoons & Compansion system	Compansion system
5 to 8	- A larger number of guided conversations - Some prepared sentences to introduce new aspects of the system and remind participants of ones they have already seen - More spontaneous conversations with Participant D and the first attempts with participants B and C (moderate intellectual disabilities)	Plaphoons & Compansion system	Compansion system
8 to 11	- Guided conversations - Spontaneous conversations recently introduced to the other participants; hardly any with Participant A, who had	Jocomunico: compansion system, prediction system and our GUI	Compansion system, prediction system and

	the most severe intellectual disabilities - A few prepared sentences (a few more with the participants with the most severe intellectual disabilities to build on the concepts they had already seen)	GUI

Table 29: Summary of the test sessions and the systems evaluated

7.3.2. Vocabulary

As expected, for participants A, B and C we only had to add to the compansion system words related to the people they spent time with (family members, friends, and supervisors at the center) and a few hobbies and interests of theirs that were not part of the system's core vocabulary. For Participant D, on the other hand, we had to add many more new words to reproduce his normal panel, especially nouns, adjectives and verbs. We decided to add all these new words to the system's core vocabulary so that future users could use them. Table 30 shows the total number of pictograms used at some point during the tests.

Type	Initial core vocabulary	Core vocabulary during the tests
Verbs	88	147
Nouns (including pronouns)	571	704
Adjectives	95	129
Adverbs	21	27
Set expressions	39	59
Question words	11	12
Modifiers (possessives, quantifiers, etc.)	15	19
Sentence modifiers (sentence type, tense, etc.)	16	17
Total pictograms	**856**	**1114**

Table 30: Core vocabulary words defined before the tests and used during the tests

7.3.2.1. Polysemic verbs

Some of the verbs introduced had several meanings but were represented by a single pictogram, so we decided to create two pictograms to remove ambiguities that the compansion system would have been unable to solve in any other way. For instance, based on some of the sentences that Participant D had tried to form using the Spanish verb "salir", we decided that we needed to create two different pictograms: one for when, in natural language, "salir" is used with the preposition "a" ("salir a" is used to say where one is going), and one for when it is used with the preposition "de" ("salir de" is used to say that one is leaving somewhere). Since the two prepositions are used with places, the compansion system had no semantic way of distinguishing between the two usages when the same pictogram was used. We believe this is how we ought to deal with other verbs that have different meanings, when different prepositions are used. This solution is in line with the pictogram-based communication approach of having several pictograms to represent the multiple meanings of a single word.

7.3.2.2. Specific pictograms

We introduced three new pictograms for the tests: one to indicate a "missing pictogram", one to mean "similar to", and one to mean "a combination of". When a user selects the "missing pictogram" option, the speech therapist can add the missing pictogram to the system. The pictograms meaning "similar to" and "a combination of" can be used to express something using a similar pictogram or a combination of two pictograms, such as "similar to an orange lemon" (i.e. a lemon that is orange in color) if the panels contain a pictogram for the color orange but not for the fruit, or "a combination of tiger and lion" if there is no pictogram for a leopard. Only Participant D used these pictograms during the tests.

7.3.3. Language-related restrictions and the system's behavior

During the tests, we found only one language-related restriction that we decided we needed to deal with, because it was frustrating our participants. Initially, our system allowed the copulative conjunction "and" to be used only to connect two nouns or two adjectives. During the tests, however, both Participant C, with moderate disabilities, and Participant D, with no intellectual disabilities or only very mild intellectual disabilities, wanted to form sentences connecting three or more items, such as "chicken" and "pizza" and "ice cream" for the sentence "I eat chicken and pizza and ice cream." We therefore decided to remove this restriction, since doing so did not make sentences more ambiguous and did not affect the system's success rate.

Another change we made based on users' requests – not so much regarding restrictions but regarding the compansion system – was to introduce the Catalan phrase "si us plau", meaning "please", at the end of imperative sentences to make them more polite by default. For example, when the user inputs "anar-comprar" (go-buy) followed by the "imperative" pictogram, it generates the sentence "Ves a comprar, si us plau" (Please could you go shopping?) rather than "Ves a comprar" ("Go shopping"), which sounds rude. Initially we had decided not to do this, since users could always add the pictogram for "please" manually, but since they used it most of the time, it made sense to program the system to use it by default.

7.3.4. GUI

The GUI was the part of the system that underwent the most changes following the problems detected during the final test sessions. In each session, the participants and the speech therapist made observations that helped us to make the necessary changes to meet the needs of people who rely on pictogram-based AAC.

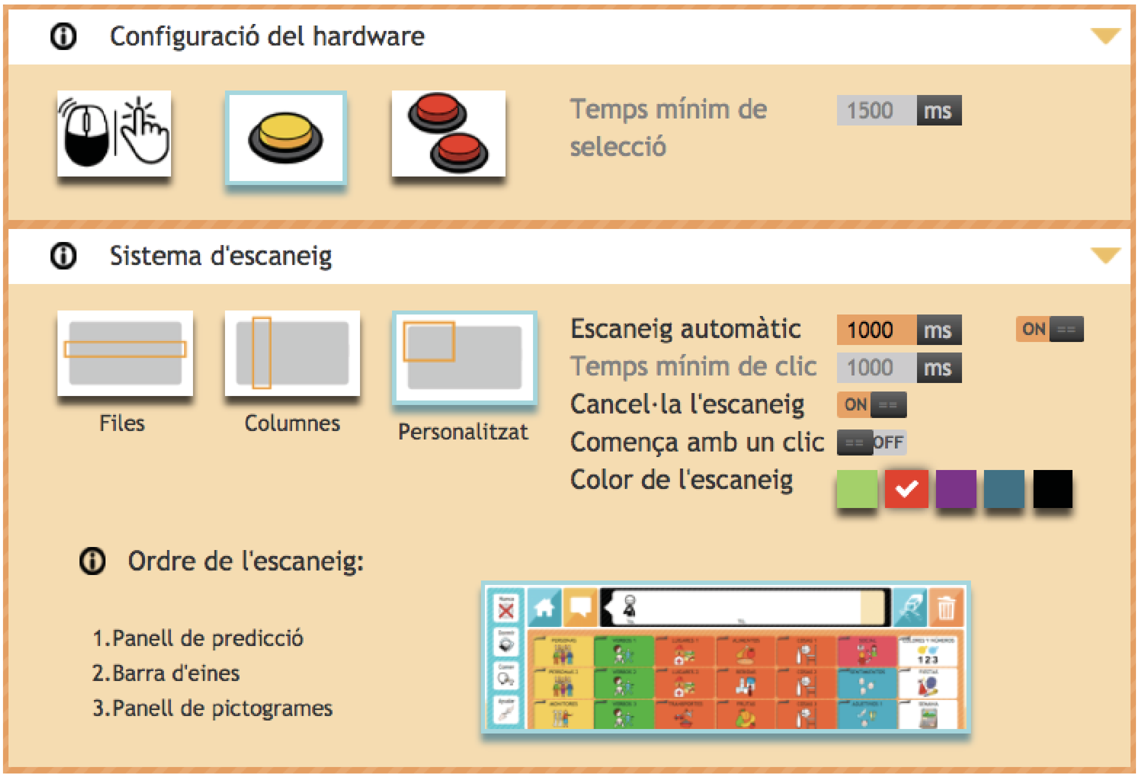


Figure 29: Screenshot of the auto-scan options

We made the GUI more user-friendly by changing the size and color of certain parts of the interface and introducing new settings that allowed participants to customize how different parts of the panels were grouped together. We created a customized scan-based selection method for users of switches, including an option to cancel scans if the wrong row or column is selected and an option for auto-scanning to begin only once the user has made a switch click, rather than automatically (see Figure 29). Finally, in the interface's pictogram prediction system, we altered the column where the predictive pictograms appeared so that it could display customized images rather than default images (see Figure 30). We also expanded the column on low-resolution screens so that the participants could better distinguish between different pictograms.

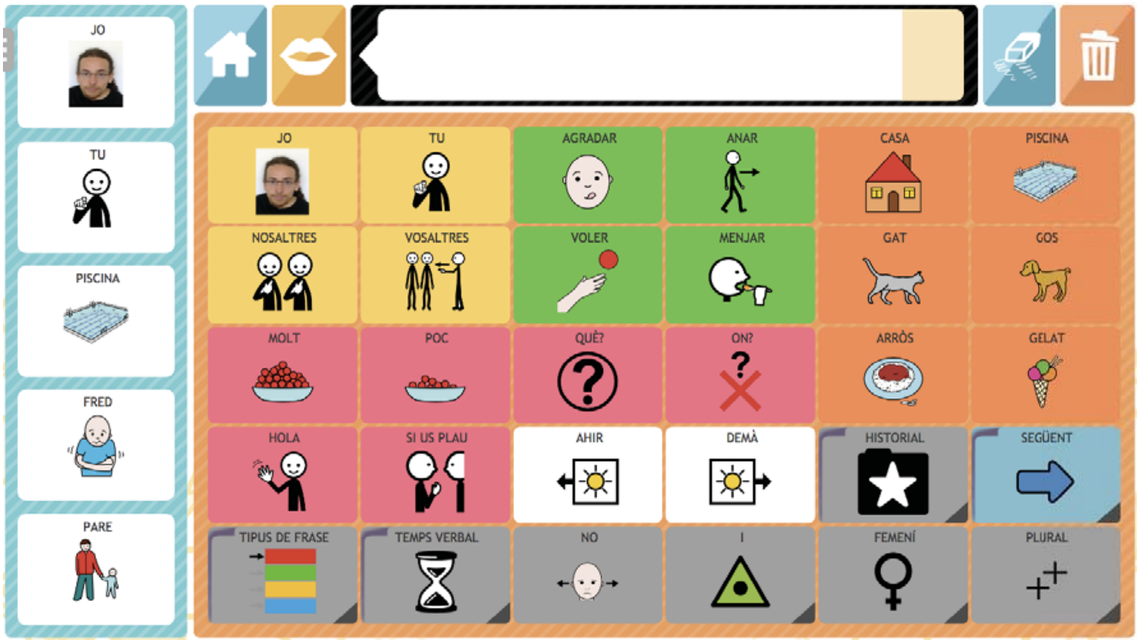


Figure 30: A prediction column (left) with a customized pictogram (top) within the interface

CHAPTER 8

RESULTS OF THE TESTS

This chapter discusses the qualitative and quantitative results of the tests conducted with persons who rely on AAC and the conclusions we drew. The results are presented in two main sections, the first looking at those generated by the compansion system and the second looking at those generated by the GUI with the prediction system. Both sections contain several subsections presenting the quantitative results of the indicators and the qualitative results.

8.1. The compansion system

Table 31 presents a summary of the overall averages of all participants for the various quantitative indicators measured in the tests (remind that non-compansion results were calculated using the model described in the previous chapter). Meanwhile, Digital Appendix W shows the values used in the calculations recorded in the sessions for all the sentences and for all the participants.

Indicator	With compansion (wcomp)	Without compansion (w/comp)	Difference (%)
Pictograms (P)	3.04 (*p*)	3.08 (*p*)	-1.27%
Resulting words (W)	4.03 (w)	3.08 (w)	30.93%
Time per sentence (S)	211 (*s*)	228 (*s*)	-7.53%
Clicks required (C)	8.04 (*c*)	8.66 (*c*)	-7.21%
Communication rate (CR)	1.15 (*wpm*)	0.81 (*wpm*)	41.59%
Word-click rate (WCR)	0.50 (*wpc*)	0.35 (*wpc*)	41.59%
Rate index (RI)	0.114 (*wpb*)	0.081 (*wpb*)	41.59%

Table 31: Overall averages for the 333 sentences created by participants during the tests

8.1.1. General quantitative analysis

Table 31 compares the average overall performances with and without the compansion system. The comparisons were made using the model described in Chapter 7 Section 7.2.3. The table shows that the number of pictograms required to form sentences fell by 1.27%, which led to a 7.53% reduction in the time required to form sentences and a 7.21% reduction in the number of clicks needed. The main conclusion we can draw from these data is that using the compansion system does not make users take longer to form sentences. In other words, the compansion system does not slow down communication. The additional time users of the system need to insert verb tenses, feminine gender modifiers, plural modifiers, copulative conjunctions, etc., is more than offset by the time saved inserting subjects, verbs and other default insertions made by the system. This conclusion is supported by the fact that this was the case for all participants, for both the prepared and the unprepared sentences, despite the wide range of language skills among those participants. All the details are shown in Digital Appendix W.

Overall, the communication rate, rate index and word-click rate all improved by 41.59% when using the compansion system. This value suggests the amount of words that would be saved when comparing it to a similar system where non-content words, such as articles and prepositions, had to be input. Nevertheless, the improvement on communication rate is language dependent; therefore these results are only valid for Catalan and Spanish. The improvement was the same for all three rates because the formulas used to calculate them depend on the number of words (W) and the number of clicks (C) required, and are independent from the selection time (S). The following formula calculates the percentage improvement in the communication rate (CR):

$$\%CRateDiff\ (\%) = \left(\frac{CR_{wexp}}{CR_{w/exp}} - 1\right) * 100 = \left(\frac{\frac{W_{wexp}}{S_{wexp}/60}}{\frac{W_{w/exp}}{S_{w/exp}/60}} - 1\right) * 100 =$$

$$= \left(\frac{W_{wexp} * S_{w/exp}}{W_{w/exp} * S_{wexp}} - 1\right) * 100 = \left(\frac{W_{wexp} * \left(\frac{C_{w/exp} * S_{wexp}}{C_{wexp}}\right)}{W_{w/exp} * S_{wexp}} - 1\right) * 100 =$$

$$= \left(\frac{W_{wexp} * C_{w/exp}}{W_{w/exp} * C_{wexp}} - 1 \right) * 100$$

Figure 31 shows that the improvement was similar for all four participants (39% for participants B and D, 43% for Participant A and 44% for Participant C), which suggests that the companssion system brings a consistent degree of improvement.

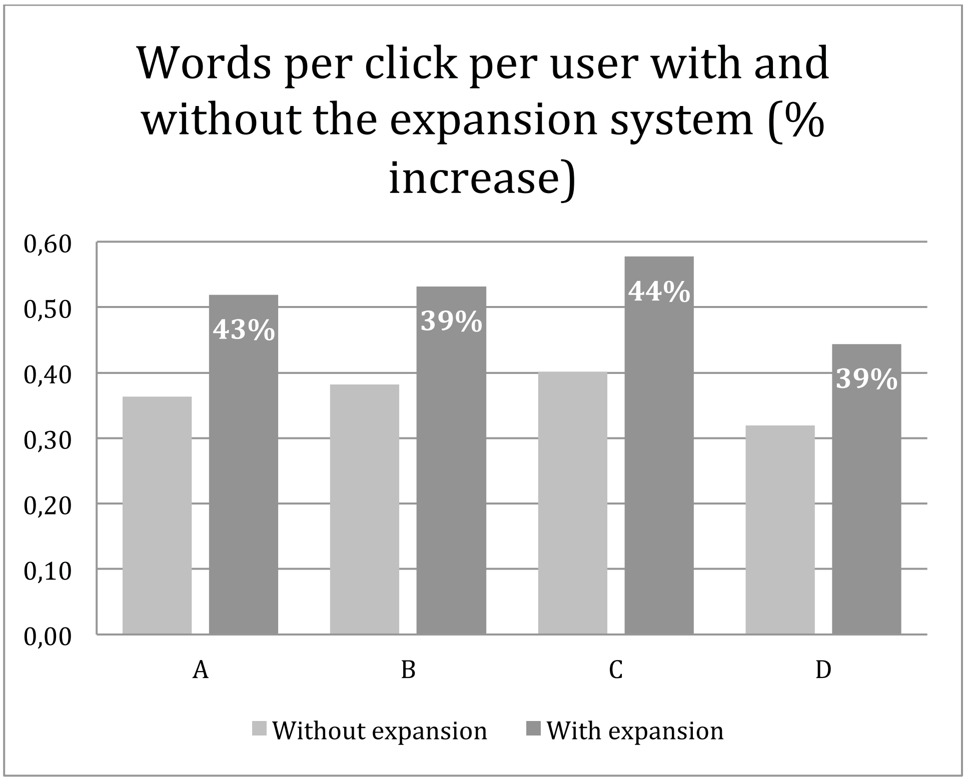


Figure 31: Words per click per user with and without the companssion system (% increase)

8.1.2. Quantitative analysis based on the number of pictograms

Although the improvement in the communication rate with the companssion system was stable, we wanted to investigate the system's performance further. One aspect we observed was that the percentage increase in the communication rate varied based on the number of pictograms in the sentences. It was immediately clear that as the sessions became more complex, with more pictograms per sentence, the communication rate improvements became less pronounced. To obtain more tangible information, we analyzed the system's performance by separating the sentences into two sets: one in which sentences had only one or two pictograms, and one in which they had three or more. Table 32 shows the key data from this comparison.

Indicators	**≤ 2 pictograms (with compansion)**	**≥ 3 pictograms (with compansion)**
Number of sentences	127	206
Participant rating (out of 5)	4.50	4.66
Pictograms (P)	1.65	3.90
Resulting words (W)	2.69	4.86
% improvement of wcomp vs. w/comp rates	59.51%	36.37%

Table 32: Average performances of the compansion system by number of pictograms per sentence

These figures prove our initial observation that the rate improvement provided by the compansion system is greater in sentences with only one or two pictograms (+59.51%) than in sentences with three or more pictograms (+36.37%).

8.1.2.1. Rating by number of pictograms

This greater improvement in shorter sentences does not correlate directly with participants' ratings of the resulting sentences. In fact, the opposite is true: participants seem to prefer the longer sentences, probably because they make the difference between telegraphic language and natural language more apparent. The difference in ratings, however, is only slightly significant: 4.50 out of 5 for sentences with one or two pictograms vs. 4.66 out of 5 for longer sentences.

In view of these results, we wanted to further study the performance of these two indicators based on sentence length.

8.1.2.2. Variation in rate improvement by number of pictograms

Figure 32 shows how the magnitude of improvement varied according to the number of pictograms. It confirms that the improvement was less pronounced in sentences with three or more pictograms than in those with only one or two pictograms. This is perhaps because in sentences with three or more pictograms in which one of the pictograms is a subject, such as "jo" (I) or "tu" (you), and another is a tense modifier, the number of words in Catalan and Spanish drops dramatically if function words like articles or prepositions are not needed. For instance, when a person communicating in Catalan inputs the pictograms "jo-menjar-pollastre-futur"

(I-eat-chicken-future), the system will produce a sentence that is only two words long: "menjaré pollastre" (I will eat chicken).

Nevertheless, the figure shows an upturn in the percentage improvement once sentences have five or more pictograms. This is probably because longer sentences that have more nouns, verbs and nouns modifiers also have more articles before those nouns and noun modifiers and more prepositions at the start of each semantic case frame for the verb patterns.

Although Figure 32 does show some cases where the rate becomes lower, such as in the sentences discussed above, the average increase is always above 30%.

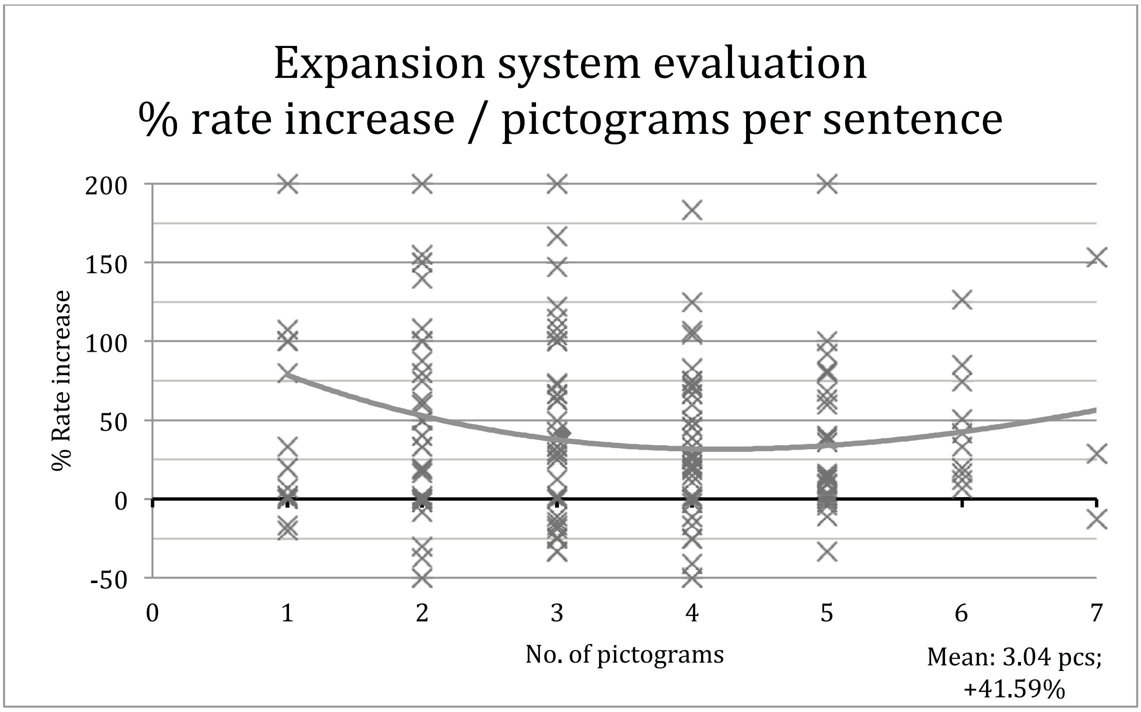


Figure 32: % rate increase vs. number of pictograms per sentence

8.1.2.3. Ratings and sentence lengths by participant language skills

Figure 33 also shows the relationships between participants' language skills and two other indicators: sentence length, and participants' average ratings of those sentences. The *x* axis shows the four participants (A, B, C and D), who are placed in order according to their language skills, making the *x* axis like a continuous variable. The results indicate a positive correlation in both cases.

They also imply that users' ratings with the compansion system could be an indicator of their language skills. Sentence length can also be an indicator of the language skills of people who rely on AAC. It seems obvious to conclude that those with better language skills get more satisfaction out of communicating in natural language.

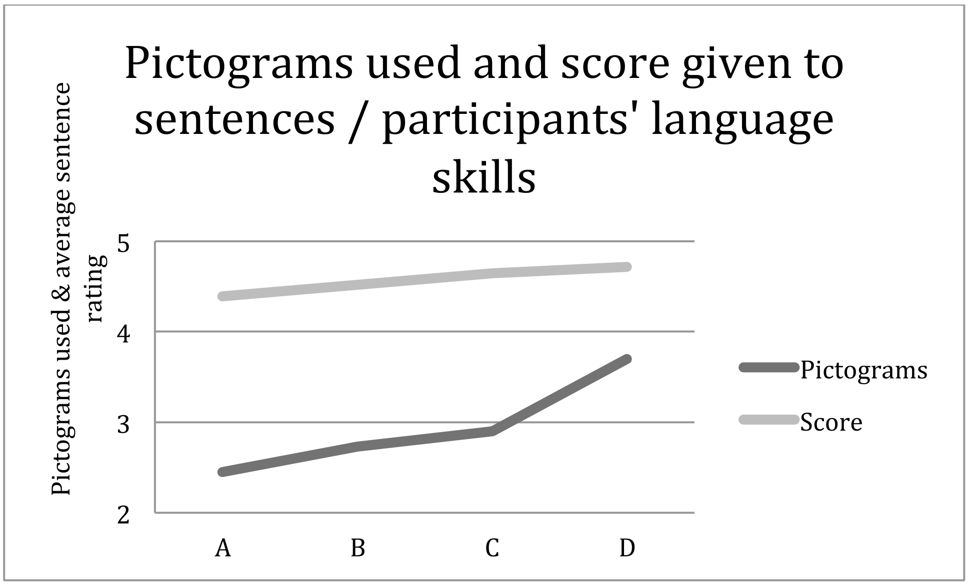


Figure 33: no. of pictograms used and rating per sentence by user language skills

8.1.3. Quantitative analysis by sentence type

In addition to the results by sentence length, we also analyzed the compansion system's performance in prepared sentences (Fx) and unprepared sentences (Fr) produced by the participants in the guided and unguided conversations. As shown in Table 33, during the tests, participants' unprepared sentences had a much greater communication rate improvement (up 55.34%) than their prepared sentences (up 36.67%), even though there was only a slight difference in the average number of pictograms used (2.91 for the unprepared sentences vs. 3.09 for the prepared sentences).

This is because in the prepared sentences we used more subjects to create sentences that were better structured, and those subjects were eliminated in the resulting sentences, since they are not required in Catalan and Spanish, which indicate the subject of verbs through verb conjugations. With the unprepared sentences, however, participants made better use of the features of the compansion system and omitted words that the system added automatically, especially subjects, but also certain default verbs, mainly in sentences containing the copulative verbs "ser" and "estar" (both meaning "to be" in both Catalan and Spanish).

Indicators	Prepared sentences (Fx)	Unprepared sentences (Fr)
Number of sentences	238	95
Participant rating (out of 5)	4.55	4.69
Pictograms (P)	3.09	2.91
Resulting words (W)	4.00	4.09
% difference between the wcomp and w/comp rates	36.67%	55.34%

Table 33: Average performances of the compansion system for prepared and unprepared sentences

8.1.4. Qualitative analysis of changes in participants' language skills

Before discussing these changes in participants' language skills, it is important to remember that they had all been using pictograms-based AAC systems for more than 12 years and had stable language skills. Below, quantitative and qualitative results based on participants' language skills are presented. All qualitative evaluations were made with the help of the main speech therapist at the center. For the quantitative values, during the tests we incorporated different numbers of words and modifiers (tense modifiers, sentence type modifiers, gender and number modifiers, etc.) into participants' panels with varying success, as outlined below. Table 28 of Section 7.3.1, summarizes these figures.

8.1.4.1. Participant A:

At the start of the tests, because of how she normally used her system, Participant A had difficulties separating the end of one sentence from the start of the next. With the prepared sentences, she worked on making the system read her sentences, using the “read aloud” button, after completing them, and gradually, this became part of her routine. Nevertheless, even during the final sessions she would still make mistakes every so often. For her, the new routine was probably more a result of her memorizing a routine, rather than understanding why she was doing it. When introducing new pictograms during the tests, we incorporated a few verbs, the question words "què" (what) and "on" (where), the conjunction "i" (and), and the "negation" and "desire" modifiers. Of these pictograms, the only ones she became accustomed to using in unguided sentences were the verb "estimar" (love), which she enjoyed using frequently, the conjunction "i" (and), such as when saying "verdura i peix" (vegetables and fish), and the "desire" sentence type, for saying things like "vull coliflor" (I would like some cauliflower). She used the "desire" modifier in several sentences that she formed by herself, which surprised the speech therapist, since the participant would not normally say what things she wanted without being prompted. We would have needed to have conducted the tests for longer to see whether she had become accustomed to this new way of expressing herself and whether she would continue to use it in the long term. But it was worth taking note of this small improvement. There was no significant difference in sentence structures between the unprepared sentences at the start and at the end of the tests. In fact, the compansion system features that the participant used most instinctively were the omission of subjects and the inclusion of default verbs. We believe that when speech therapists work with participants with the least developed language skills, they ought to configure the compansion system to associate default verbs with more nouns from the entire vocabulary set.

8.1.4.2. Participant B:

Like with Participant A, we introduced several new verbs, the question words "què" (what) and "on" (where), the conjunction "i" (and), and the "negative" and "desire" modifiers, but we also introduced three tenses ("past", "present" and "future"), and the adverbs of time for "yesterday" and "tomorrow". Participant B not only became accustomed to using the conjunction "i" (and) and sentences expressing "desire", but also the "negation" modifier, verb tenses (as explained in more detail at the end of this section) and question words. When asking questions, he struggled to distinguish between the Catalan question words "què" (what) and "on" (where), but he did ask some complex questions during the conversation, such as "on-vosaltres-anar-cap de setmana" (where-you [plural]-go-weekend), which became "On aneu el cap de setmana vosaltres?" (Where are you going this weekend?). He had some difficulties using new verbs, though we believe this was not because his language skills were limited, but because he struggled to recognize the pictograms, even some of those he had been using for many years. Although he used new vocabulary that we had introduced, his sentence structures were no better than before: he made no lasting improvements to the order of his sentences, and he did not use more verbs or subjects. He did, however, improve his use of different sentence types, verb tenses, the conjunction "i" (and) and questions, which has encouraged the speech therapist to continue working on those aspects so that he can consolidate his progress during his regular speech therapy sessions.

8.1.4.3. Participant C:

Participant C made the most progress of the four. He had limited language skills at the start, similar to those of Participant B but more advanced than those of Participant A. We introduced him to the same features that we had introduced to Participant B, and he assimilated all of them. He was particularly keen and skilled (by the end of the tests) at using question words, despite never having used them in the past. He was also more successful than Participant B at using verb tenses, having formed several sentences, without guidance, that used the "past" and "future" tense modifiers, even when responding to questions that he would previously have answered using only a single pictogram. Indeed, it was when he was answering questions that he showed the most progress in terms of sentence structures. Having previously answered questions using just a single pictogram, during the tests he began answering questions by specifying the subject, the verb (and sometimes even the tense, if necessary), and the occasional adverbial adjunct or complement, such as a location. One of his most complex responses was "jo-anar-bar-i-casa-futur" (I-go-bar-and-home-future), which became "Aniré al bar i a casa" (I will go to the bar and go home). However, we do not know whether he made these improvements by memorizing a routine rather than fully understanding why he was doing it. It was probably a combination of the two. We would need to conduct longer tests to draw conclusions that are more definitive, but the results were very positive nonetheless. We believe that hearing the sentences in natural language proved a source of motivation. Like with Participant B, the speech therapist will continue to use the system with Participant C in his speech therapy sessions.

8.1.4.4. Participant D:

The tests with the compansion system were also a watershed moment for Participant D in terms of the language he can use to express himself. Although he could already form complex sentences with his existing system, and could even express certain prepositions, by the end of the tests he was able to use an even wider range of sentence structures. During the tests, Participant D's panels incorporated all the items added to those of Participant C, sometimes with greater complexity. Participant D's panel, for instance, had pictograms for three past-tense verbs: one labeled "distant past" for the imperfect tense, used in Spanish to refer to a continuing or repeated event; one labeled "past" to refer to the Spanish language's periphrastic tense, which is normally used to refer to events that happened before the start of the current day; and one labeled "immediate past" to express the perfect tense, usually used in Castillian Spanish to refer events that occurred earlier the same day. By the final sessions, he was using all three tenses correctly. In addition to using the conjunction "y" (and), he also learned how to use modifiers to label nouns and adjectives as feminine and/or plural, and how to use the "missing pictogram", "similar to" and "a combination of" pictograms described in Section 7.3.2.2. In addition to the "desire" and "negation" modifiers, the participant also used the "permission", "imperative", "question", "exclamation" and "conditional" modifiers in unprepared sentences. He used new vocabulary that we had introduced, too, such as adverbs of time like "nunca" (meaning "never" or "ever", according to the context), adverbs of place, subordinate conjunctions like "però" (meaning "but", and requiring a separate sentence in our compansion system: "Estic malalt. Però aniré a la festa" ["I'm unwell. But I'll go to the party"]), and words to tell the time to the nearest hour or quarter of an hour. Furthermore, participant D was the only one that faced the need of having on his panels a verb with several meanings or with similar sentence structures split into two pictograms, as explained in Section 7.3.2.1. Despite the added complexity, he was able to use them properly. The linguistic complexity he was able to achieve is illustrated by some of the sentences he produced: "¿Ir al concierto te gusta?" ("Do you like going to concerts?"); "¿Puedo escuchar la música muy fuerte, por favor?" ("Please can I listen to the music really loudly?"); "Quiero que Brenda me dé cava." ("I want Brenda to give me cava"). The greatest improvement to his sentence structures came with his replies to questions. Like Participant C, having previously given very short replies to questions, he began using subjects, verbs, adverbial adjuncts, complements, and in some cases, tenses. The speech therapist will continue using the system in her sessions with Participant D.

8.1.4.5. Observation on the use of tenses by participants B and C

With participants B and C, we observed that although we introduced the concepts of "past" and "future", as explained above, they often mixed up the "past" tense modifier and the adverb "ahir" (yesterday), which would produce the same verb forms and similar structures. They also had the same difficulty distinguishing between the "future" tense modifier and the adverb "demà" (tomorrow) for the same reason. These difficulties seem to be consistent with research that suggests that people with moderate to severe intellectual disabilities often struggle with stimulus generalization (Horner & Albin, 1988; Joseph & Konrad, 2009; Westling & Fox, 2009; Johnston et al., 2012).

8.1.4.6. An indicator of how well users assimilated new concepts

Our final linguistic observation of our participants is that one indicator of how well they assimilated the new system could be the difference between the number of pictograms they used in prepared sentences (Fx) and the number they used in unprepared sentences (Fr) as a percentage of the latter. This percentage might indicate how well participants assimilated concepts based on their complexity in terms of the number of pictograms. More importantly, it might allow us to compare results among people who took part in similar tests. Table 34 gives these percentage values for each user. The value for Participant C is much lower than the one for Participant B, which confirms the finding in the qualitative evaluations that Participant C had become much better accustomed to using the new concepts introduced. Another value that stands out is the negative one for Participant D, suggesting that we could have introduced more new concepts to him.

Participant	Pictograms in Fx	Pictograms in Fr	Diff. as % of Fx
A	2.55	2.08	22.60%
B	3.02	1.90	58.95%
C	3.06	2.37	29.11%
D	3.59	3.86	-6.99%

Table 34: Comparison of the number of pictograms used by each user in prepared and unprepared sentences

8.1.5. Qualitative analysis of the communication experience

The qualitative assessment of the compansion system by the center's main speech therapist indicates that the system produced much better communication than telegraphic language, which is of poor quality compared with natural language. All communication partners involved pointed that communication felt more natural and that sentences were easier to understand, mainly due to conjugated verbs and words in the correct order within the sentence. Communication also improved for the participants using AAC. A highlight for us was the initial joy, happiness and surprise that they expressed when they heard sentences that they had formed themselves in natural language for the first time. Over the course of the sessions, these outbursts of joy gradually gave way to other emotions, such as motivation, expectation and satisfaction.

The quantitative values also support these qualitative observations. Participants' average ratings for the sessions were 4.67 out of 5 and their communication partners' ratings were 4.68 out of 5, where 3 represented the same level of satisfaction as when producing telegraphic sentences with the previous system. The highest average rating by a participant was 4.90 by Participant D, and the lowest was 4.34 by Participant A. The highest average rating by communication partners was 4.90 for the sessions with Participant D, and the lowest was 4.42 for the sessions with Participant A.

Another interesting aspect concerning participant ratings, as shown earlier in Figure 33, is that satisfaction ratings were higher the lower the participant's intellectual disabilities were. These findings seem to suggest that the compansion system brings most satisfaction to those who have more advanced language skills. Still, even the participants with the most intellectual disabilities gave average ratings well above 3 out of 5: Participant A's was 4.39 and Participant B's was 4.52.

8.2. The graphical user interface (GUI)

This section focuses on the impact of including the GUI in the tests, and more specifically, the prediction system that was part of the interface. It is important to mention that at each session the prediction system did not take into account previous sentences that the participant had formed, since many of those sentences were prepared, so there was no point in the system learning from them. The prediction system therefore used only the features that rely on the syntactic and semantic data stored in the compansion system, and not the features that learn from usage by the user.

The new interface and the prediction system were introduced in session 8 for participants B, C and D and session 10 for Participant A. Table 35 summarizes the averages for the main indicators, comparing the values obtained when using both the prediction and the compansion system together with the values obtained without using either of the two systems. If we compare these values with those in Table 32, we see that the combination of the prediction and compansion systems reduced the time needed by 15.54%, compared with a 7.53% time reduction using the compansion system alone. The combination of the two systems also reduced the number of clicks needed by 15.25%, compared with a 7.21% reduction when using the compansion system alone. These figures show that the prediction system, which had a 23% hit rate, reduces the number of clicks required to form a sentence.

Indicator	**With prediction (wpred & wcomp)**	**Without prediction (w/pred & w/comp)**	**Difference (%)**
Pictograms (P)	3.70 (*p*)	3.73 (*p*)	-0.80%
Resulting words (W)	4.58 (*w*)	3.73 (*w*)	22.79%
Time per sentence (S)	326 (*s*)	386 (*s*)	-15.54%
Clicks required (C)	9.28 (*c*)	10.95 (*c*)	-15.25%
Communication rate (CR)	0.84 (*wpm*)	0.58 (*wpm*)	45.67%
Word-click rate (WCR)	0.49 (*wpc*)	0.34 (*wpc*)	45.67%
Rate index (RI)	0.113 (*wpb*)	0.077 (*wpb*)	45.67%
Hit rate	23.00%	-	-

Table 35: Averages for the 81 sentences formed using the prediction and compansion systems together vs. averages for the sentences formed without the systems

8.2.1. Hit rate

We observed that the prediction system's 23% hit rate was thanks essentially to its ability to predict the subjects "jo" (I) and "tu" (you) at the start of sentences. The hit rate varied from one participant to another, with Participant A having the lowest rate, at 19%, and Participant D having the highest rate, at 25%. Participant D produced the best structured sentences and took best advantage of the prediction system, which uses sentence structure data programmed into the compansion system. The difference, however, between his hit rate and that of the other participants was not particularly high.

8.2.2. Communication rate

Looking again at the values in Table 35, there are two main reasons why the number of pictograms per sentence was higher when the two systems were used together (3.70 pictograms) than when the compansion system was used on its own (3.04 pictograms): first, the prediction system encouraged participants to include the subject "jo" (I), rather than omit it; and second, the prediction system was added in the final test sessions, and sentence lengths had been gradually increasing since the beginning of the tests.

Longer sentences ought not to have led to a lower communication rate, yet the rate fell from 1.15 wpm to 0.84 wpm when the prediction system was introduced.

Perhaps the lower rate was not because of the system itself, but because participants were less confident when using the new interface. Or perhaps they were less confident because the prediction system creates a separate, dynamic column that changes based on the pictograms that are selected. Or perhaps the participants needed time to see which pictograms were visible in the prediction column at any given moment

Due to a combination of these factors, perhaps the participants needed time to adapt to the prediction system and the new interface. We investigated this by looking at how each participant's communication rate evolved from session to session (Figure 34).

The figure shows that Participant D's communication rate was fairly constant throughout the first seven sessions, but the communication rates of the other participants were more variable because their state of health varied due to their disabilities, which affected their selection rates. Nevertheless, the figure shows a drop in the rates of participants B, C and D following the introduction of the GUI with the prediction system in session 8. For Participant D, this may be partly a result of us changing the distribution of his verb panels, since the speech therapist suggested that we group the verbs together by the three Spanish verb endings (-ar, -er and -ir).

Despite the initial fall, the communication rates of participants C and D recovered in the final two sessions thanks to an improved selection rate, with more clicks per minute in the later sessions, when they were using the prediction system. Figure 34 shows that in the final two sessions, the values for users A, C and D returned to the levels recorded in the sessions without the prediction system. This suggests that the prediction system and the new interface require a small adjustment period, unlike the compansion system, which did not. None of the rates in Participant A's sessions with the prediction system were all that different from her rates in the other sessions. Participant B's rates, on the other hand, did not return to normal after the final two sessions, suggesting that he needed a few more sessions to adapt to the prediction system and the new interface.

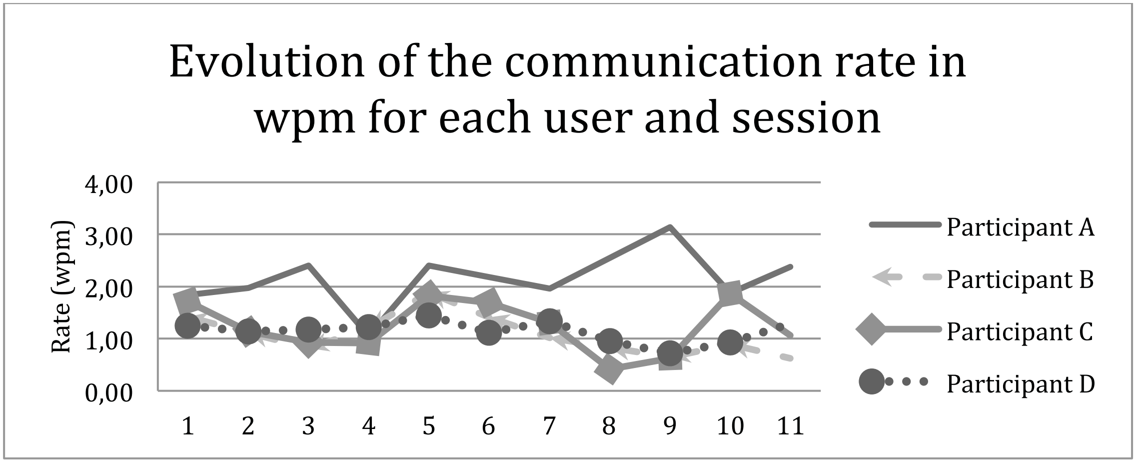


Figure 34: Evolution of the communication rate in *wpm* for each user and session

8.3. Discussion of the results

8.3.1. The compansion system

Before our tests, there was little literature on applied pictogram compansion systems, and even less on tests of such systems with people who rely on AAC. We were afraid that the lack of other similar experiments and studies might be because they were not likely to produce fruitful results, either because the people who relied on AAC were unable to adapt or for other reasons, but our results proved our fears to be unfounded. The lack of studies prior to ours was probably because nobody had created a reliable enough system adapted to the needs of people who rely on

AAC with a large enough vocabulary to prevent users from becoming frustrated, as they did in Vaillant (1997).

Our compansion system fulfilled these requirements, and the test results show that it brings many benefits to pictogram-based communication:

First of all, it enriches the telegraphic language produced by pictograms by converting it to natural language. The participants' joyful reactions when hearing full sentences and their average rating of 4.59 out of 5 for such sentences show that they were satisfied with the conversions to natural language. The speech therapists at the center where the tests took place were also delighted with the experience, so much so that they will continue to use it with the participants and will incorporate it into speech therapy sessions with other people who rely on AAC. Another positive indicator of this benefit is the fact that, just four months after the system became publicly available, at least 800 centers and people were already using it.

Second, those who use the system improve their communication rate, according to the used model, by well over a third (41.59% in our tests).

Third, although the system requires additional information to form certain sentence types, neither the communication time nor the number of clicks required increases. On the contrary, in our tests the communication time fell by 7.53% and the number of clicks by 7.21%.

Fourth, the system is adapted to people with a range of language skills. There is no learning curve to start using it, as those with severe intellectual disabilities can continue to use the methodology that they use with their current system without using the modifiers of the compansion system, which are optional, while those with more advanced language skills can create sentences that are more complex with the help of these modifiers. Participant D had a high linguistic potential, so he was ideal for testing the system's restrictions and features.

Based on the results observed in Section 8.1.4, we believe the system could improve people's language skills through speech therapy sessions. Specifically, we believe that those who use the system could improve how they structure their sentences, but we would need to hold sessions over a longer period of time to verify this hypothesis. One possible future line of study worth investigating is whether producing sentences in natural language allows people who rely on AAC to learn new structures and incorporate new vocabulary, since we believe that much of the vocabulary that we introduced could probably have been introduced in other pictogram-based systems that do not expand language. Our observations and those made by the speech therapist seem to support this idea, since in their regular speech therapy sessions, the participants – whose language skills had been stable for many years – became extra motivated by the compansion system. Even on days when they felt unwell and would not normally have attended their sessions, they came anyway, especially Participant C. We therefore believe that the additional learning was not the result of an intrinsic feature of the compansion system, but was psychological, as a result of the participants' enthusiasm at being able to communicate using natural language, producing sentences that were tangibly different to those they used to produce.

In addition to aiding communication and allowing users to learn new language skills, as observed during the tests, the compansion system has one other benefit: it provides indications of the language skills of people with speech disorders. Although some other tools are better suited to identifying language skills, such as the Boston Naming Test (Kaplan et al., 1983), which is widely used in clinical evaluations, our system revealed some useful data during the tests. The main speech therapist at the center, who monitored the participants and conducted follow-up work, is of the same opinion. In addition to being a communication tool, when used in speech therapy sessions the compansion system can improve users' language skills. We believe the compansion system could also be used to evaluate those language skills based on the following observations:

- In the prepared sentences, when we modified the system to produce poorly structured sentences, the overall ratings fell from an average of 4.59 to 1.73 out of 5, and participants made their shock and dismay plain to see. We began designing sentences to reproduce major errors committed when forming sentences, such as the wrong subject or direct object being selected, and minor errors that changed the meaning of sentences only slightly. Participant A, with the most severe intellectual disabilities, reacted negatively only when the system picked the wrong subject, or when it picked the wrong adverbial adjunct or complement if that led to substantial changes to the word order or meaning of the sentences. The participants with moderate intellectual disabilities (B and C) detected the same mistakes as Participant A, but also detected wrong prepositions, except where two prepositions had a similar pronunciation, such as the Catalan prepositions "amb" (with) and "a" (meaning "to", "in", "on" or "at", depending on the context). However, they did not spot changes to prepositions, articles, and or verb tenses where the mistakes were similar to the correct versions. Finally, Participant D, who had no or very mild intellectual disabilities, spotted all the errors in the sentences that were generated. This shows that participants were able to recognize – some better than others – whether the resulting sentence was formed as they intended.

- As we saw in Section 8.1.4.5, subtle errors in tense usage, such as those observed in the tests with participants B and C, may be indicators of a user's level of intellectual disabilities and language skills.

- We believe that another indicator of users' language skills might be the number of sentence modifiers ("desire", "request", "imperative", "exclamation", "conditional", "question", etc.) that they are capable of using, such as whether they can use the conjunction "i" (and), and whether they can use feminine and plural modifiers with nouns that have various forms, such as the Catalan word for "cat", which has masculine singular ("gat"), feminine singular ("gata"), masculine plural ("gats") and feminine plural ("gates") forms. As explained in the results section, with Participant D, who has the most advanced language skills and has no or very mild intellectual disabilities, we introduced and tested these options successfully both in prepared and unprepared sentences. With the other participants, however, we were not able to introduce so many modifiers. Nevertheless, as explained earlier, Participant C assimilated more concepts than Participant B, which probably indicates that his language skills were better than the speech therapist thought before the start of the tests. As we have already said, these observations could probably be made with other pictogram-based systems that do not expand language, but we believe that it is easier to make such observations when the resulting sentences are in natural language.

- Finally, as seen in Section 8.1.2.3 of the results, another potential indicator of users' language skills are their ratings of the sentences that they produce using the compansion system.

The system may even provide indicators of which people who use pictograms might be capable of reading and writing, and might even encourage such people to take that step. From our study, we believe Participant D clearly has such potential, and at the sessions he told the speech therapist that he wanted to learn how to read and write.

8.3.2. The graphical user interface (GUI)

The GUI, incorporating the prediction system, also achieved good results, though since we tested it only in a few sessions, the data sample is small. The prediction system performed better than the compansion system alone, with a 23% hit rate, a 15.54% reduction in the average time needed to form a sentence and a 15.25% reduction in the number of clicks made.

By predicting the subjects of verbs, the prediction system also helped users to start their sentences in a more structured way. Unlike the compansion system, the prediction system did have a small learning curve, as described in the results section, but we would need to conduct more tests over a longer period of time to generate more results and confirm our current findings. Further tests would also allow us to test the part of the system that learns from the user.

Regarding the rest of the GUI that we developed, by the final two sessions, the communication rate of three of the four participants had returned to the values seen when those participants were working with their usual system, as shown in Figure 34. This, along with the feedback we received from the participants and the staff at the center, leads us to believe that, overall, the GUI complies with the usability requirements for these types of users.

CHAPTER 9

CONCLUSIONS

After nearly five years of work, more counting the masters' dissertation where this work comes from, we are happy to reach this final chapter, not only because it signifies the completion of the doctoral dissertation, but also because our research obtained positive results and because it was applied to an application that is currently helping people who rely on AAC.

Having said that, in this final chapter we will first summarize the work and the conclusions extracted during this thesis and then we will discuss future lines of research.

9.1. Conclusions

The first part of this doctoral dissertation (Chapters 4 and 5) focuses on a compansion system that expands the telegraphic language that comes from the use of pictograms in AAC into natural language. As we have seen in the State of the Art chapter, before this thesis began, there was no reported research of a similar system successfully tested with persons who rely on AAC on a daily basis and with ranging degrees of linguistic competence. Thus, there was a need of doing research on a reliable enough system that could later be tested with end-users and that could show or dispel the potential that outputs in natural language from pictogram-based AAC promised: an overall improvement to the communication experience both for people who rely on AAC and their communication partners, the potential to become a tool that supports literacy or language rehabilitation and the ability to increase the rate of communication without meaning an extra effort by the users, among others.

The compansion system that we researched transforms telegraphic language into syntactically and semantically well-formed sentences, both in Catalan and Spanish. As telegraphic language can sometimes have several interpretations due to lack of non-content words, which can be important to determine the function of each word, in order to narrow the options and in order to allow users to build more complex sentences, we decided to add optional word, tense and sentence modifiers to telegraphic communication.

In turn, the core vocabulary used by the compansion system is based on the CACE vocabulary. It is a vocabulary selection specially designed for basic pictogram-based communication. In total, our system has got 1100 vocabulary items that can also be expanded by each person, adapting it to his or her specific needs.

The system, in order to transform the input pictograms into a natural language sentence, has two main components: a parser and a generator. The parser takes the input pictograms and, mainly using semantic information, decides the semantic roles of each of them. In order to do it, it uses a controlled grammar dependency parsing method that does not accept every possible construction in Catalan and Spanish, but only a restricted set of it. The aim of the controlled grammas is also to reduce the amount of ambiguity that the parser faces in terms of possible sentence structures and possible relations between words. The algorithm of the parser is nearly language independent. In fact, the algorithm is designed to support languages with different relative word order in the structures of the sentences.

In order to decide which pictogram fits each semantic role or frame, the algorithm of the parser only uses syntactic information (mainly word order) if the semantic information annotated for each word and for each verb pattern is not enough to solve an ambiguity. For each possible semantic frame that a pattern can accept, an associated word class that better fits it is assigned. A weighting algorithm is later used to decide how well a pictogram fits each frame.

Finally, from the resulting parse tree, the generator constructs the natural language sentence. The generator is formed by several independent modules that progressively expand the parsed input step by step. First, there is a slot-ordering module that reorders the components of the sentence according to the sentence type. Then, a word order module, which orders the words within the slots and which adds prepositions and makes sure that words agree in gender and number with each other. Afterwards, the articles' module, which attaches articles to nouns, if necessary, followed by the verb conjugation module. Eventually, there is a cleaning module that deals with the final details and gets the sentence ready to output.

Before conducting the tests with people who rely on AAC, we wanted to ensure that the compansion system for Catalan and Spanish was performing above a proposed performance threshold: 90% of sentences deemed "acceptable" and 80% of deemed "perfect". In order to do so, we conducted technical tests on the system with the help of three independent annotators. In these tests, 99.66% of the generated sentences by the compansion system, allowed by the controlled grammar, deemed “acceptable”. The most common differences that made annotators deem the sentences that they intended to build “acceptable” (~10% of the sentences) instead of “perfect” (~90%) were minor errors that did not affect the understanding of the sentence. In their assessments, the interrater agreement was over 85%. All in all, results proved that the system was reliable and it allowed to continue to the next steps of the doctoral dissertation with a solid foundation as its base.

Apart from these tests in Catalan and Spanish, we also conducted tests to explore the possibility of using a machine translation system (i.e. Google Translate) as the generator of our system in order to be able to add new languages to it in an easier way. However, these tests conducted in English had mixed results and were not as successful as the previous ones. This suggests that in order to add new languages it's better to code specific generators for each language or that there should be further post-processing of the translated output in order to address the main issues that arise from the use of a non-specific system such Google Translate (e.g. proper detection and translation of idioms and frozen expressions).

The second part of the thesis (Chapter 6) is the development of a user interface, which was necessary before conducting the tests with end-users. The user interface had to take into account all the necessary accessibility requirements in order for the participants of the tests to be able to use the compansion system. Some of these requirements included fully customizable pictogram boards and vocabulary and scanning methods to select pictograms, modifiers and interface functions.

While developing the interface, we also researched on a pictogram prediction system that could take advantage of all the encoded data of the compansion system in order to further reduce the amount of effort that users need to build sentences. A secondary goal of the system was to help people who rely on pictogram-based AAC better structure sentences. Apart from using the semantic data encoded in the verb patterns, the algorithm also learns of the use by the users by means of n-grams and also takes into account contextual time, in other words, the time of the day and days of the week when sentences are built. The algorithm mixes these three features and reserves slots for each of them in the final allotted number of predicted pictograms.

The final part of the doctoral dissertation (Chapters 7 and 8) are the tests conducted with four adult volunteers who had at least 30 years' experience using pictogram-based communication and had different degrees of cerebral palsy and intellectual disabilities. We did a total of 40 sessions where participants reproduced both prepared sentences, in order to learn the characteristics of the compansion system, and unprepared sentences. Their initial reactions after hearing their expanded sentences, as well as their long-term assessment of the system, show that they were really satisfied with the conversions to natural language. The system not only pleased the participants, but also their communication partners and the speech language pathologists that, after the end of the tests, decided that they would incorporate it into speech therapy sessions with other people who rely on AAC.

In terms of quantitative results, tests were also promising. First of all, according to the used model, the communication rate increased by well over a third. Furthermore, both the communication time and the number of clicks fell by ~7%. These figures are linked to the observation that the use of the compansion system did not have a learning curve, as those with severe intellectual disabilities can continue to use it as they use their current system without the specific modifiers of the compansion system, while those with more advanced language skills can create sentences that are more complex with the help of these modifiers.

Moreover, qualitative results from the tests, supported by the main speech language pathologist that surveyed them, suggest that the system can be a tool to improve the language skills of the users, as all participants, in different degrees, showed improvement in terms of either vocabulary, used modifiers or better structuring of certain types of sentences (e.g. questions and answers), and also a tool that can support in the assessment of a person's linguistic competence. Nevertheless, both statements would need further specific and longer tests in order to be proven conclusions.

Regarding the user interface and the prediction system, the user interface complied with the accessibility needs required by the participants. As for the prediction system, as it was not included from the beginning of the tests, more data is needed in order to extract conclusions. The only observation worth noting is that by predicting the subjects of verbs, the prediction system, seemed to help users to start their sentences in a more structured way.

To sum up, the results of the tests proved that the compansion system can be useful for people who rely on pictogram-based AAC with a wide range of linguistic competence. The overall communication experience definitely improves without causing an extra burden to the user and there are signs that it can also be of help in terms of acquiring new linguistic skills, both due to the motivation factor of hearing the sentences in natural language and due to obtaining better outputs the more the input pictograms are structured resembling the desired output.

To conclude this section, we would like to mention that the final application, result of all the conducted research, is named Jocomunico (www.jocomunico.com). Since its presentation in December 2016, more than 1000 users, hospitals, special education schools and centers for people who rely on AAC have started using it. We think that this great acceptance within the community supports the research findings and encourages future work.

9.2. Future work

In this section we will discuss lines of future work that have appeared throughout the chapters of this thesis.

As we have just seen, it has been proven that the compansion system developed has the potential to suit a large range of people who rely on pictogram-based AAC and that its controlled grammar adapts to their communication needs. Nevertheless, it may be the case that for users with near to intact linguistic competence, tests in daily live situations for a longer period of time might arise the issue that the controlled grammar is too restrictive in some cases. Therefore, a future line of work is to conduct tests in a real-life environment for a prolonged period of time to detect if the available structures of sentences that the system allows should be expanded (e.g. better support for coordination between sentences, for causal or disjunctive subordination, etc.) without trading much accuracy, so that it is an improvement and not a downgrade.

Another line of future research concerning new sets of tests, is to focus on longer tests to confirm whether the system can improve people’s language skills or not, whether it can help to assess the linguistic competence of users and also on tests with participants with autism spectrum disorders and other persons with complex communication needs to see if findings can be transferred to other groups of people who rely on AAC. It would also be interesting to research if similar results, as the ones obtained, can be obtained using telegraphic plus systems in commercial software, where tense and number modifiers can be applied in order to produce similar outputs.

Moreover, further tests with the prediction system are needed in order to extract solid conclusions from it. Further tests would also allow to test the part of the prediction system that learns from the user, which we were unable to try.

To conclude, as one of the main disadvantages of the compansion system is that the generator is language dependent, despite the mixed results obtained using the Google Translate approach, it would be interesting to further investigate the line of using existing machine translation engines in order to create generators for new languages. If this line of work did not succeed, research on creating generators for new languages, such as English, would be a line of future work in itself.

Furthermore, also in terms of making the system more easily scalable, as, at the moment, adding new vocabulary takes time and requires linguistic knowledge, it would be interesting to investigate new ways to make this process more user friendly and less time consuming maybe by finding similarities between the already available vocabulary and, by means of natural language processing and machine learning techniques, making reusable patterns from it.

Finally, as all the sentences built by the more than 1000 online users of Jocomunico are stored in a centralized database, another future line of work is to do data mining on this corpus in order to extract information on the use of the application and conclusions that can help us further improve the compansion system.

Alm, N., Iwabuchi, M., Andreasen, P. N. & Nakamura, K. (2002), A multi-lingual augmentative communication system. In *ERCIM Workshop on User Interfaces for All*, pp. 398-408. Springer, Berlin, Heidelberg.

Augé, C. and Escoin, J. (2003), Chapter 7: Tecnologías de ayuda y sistemas aumentativos y alternativos de comunicación en personas con discapacidad motora. *In Alcantud, F. (2003), Tecnologías de ayuda en personas con trastornos de comunicación*, Nau Llibres.

Blackstone, S. (2007), AAC Research and Development: Having an Impact, *Augmentative Communication News (ACN)*, Vol. 19, Num. 3, pp. 1-4. Retrieved at https://goo.gl/YAb0ll

Blackstone, S. (2009), AAC Communication and Technology. *Augmentative Communication News (ACN)*, Vol. 21, Num. 3, pp. 1-8.

Bonafonte, A., Aguilar, L., Esquerra, I., Oller, S. and Moreno, A. (2009), Recent work on the FESTCAT Database Speech Synthesis. *I Joint SIG-IL/ Microsoft Workshop on Speech and Language Technologies for Iberian Languages*, 2009.

Brennan R. L. & Prediger D. J. (1981), Coefficient kappa: some uses, misuses, and alternatives. *Educ Psychol Meas*. 1981; 41: 687–699.

Cabello-Luque F. & Bertola-López E. (2012), Símbolos pictográficos de ARASAAC: ¿son adecuados? *En Navarro, J; Fernández, Mª.Tª; Soto, F.J. y Tortosa F. (Coords.) (2012) Respuestas flexibles en contextos educativos diversos*. Murcia: Consejería de Educación, Formación y Empleo.

Cardey, S. & Greenfield, P. (2008), Micro-systemic Linguistic Analysis and Software Engineering: a Synthesis. *In revue RML6, 2008, Actes du Colloque International en Traductologie et TAL*, Editions Dar El Gharb, Oran, pp. 5-25.

Copestake, A. (1997), Augmented and Alternative NLP Techniques for Augmentative and Alternative Communication, *In Proceedings of the ACL Workshop on Natural Language Processing for Communication Aids*, Madrid. Retrieved at http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.49.1040

Copestake, A. & Flickinger, D. (1998), Enriched Language Models for Flexible Generation in AAC Systems. *Technology and Persons with Disabilities Conference (CSUN-98)*. Los Angeles, CA.

Covington, M-A. (1990a), A Dependency Parser for Variable-word-order Languages, *Technical Report AI-1990-01*, University of Georgia.

Covington, M-A. (1990b), Parsing Discontinuous Constituents in Dependency Grammar. *Computational Linguistics* 16: pp. 234–236.

Covington, M-A. (1994), Discontinuous Dependency Parsing of Free and Fixed Word Order: Work in Progress, *Technical Report AI-1994-02*, University of Georgia.

Covington, M-A. (2001), A Fundamental Algorithm for Dependency Parsing. *Proceedings of the 39th Annual ACM Southeast Conference*, pp. 95–102. Retrieved at http://citeseerx.ist.psu.edu/viewdoc/similar?doi=10.1.1.136.7335

Creswell, J. W. (2002), Educational research: Planning, conducting, and evaluating quantitative research. *Upper Saddle River*, NJ: Pearson Education. Retrieved at https://goo.gl/PggL9z

Debusmann, R., Duchier, D. & Kruijff, G-J. (2004), Extensible Dependency Grammar: A New Methodology. *Proceedings of the Workshop on Recent Advances in Dependency Grammar*, pp. 78–85.

Demasco, P. W. & McCoy, K. F. (1992), Generating text from compressed input: An intelligent interface for people with severe motor impairments. *Communications of the ACM*, 35(5):68-78. doi: 10.1145/129875.129881

Detheridge, T. (2005). Símbolos y futuras tecnologías: el proyecto Widgit Rebus. *Primeras Noticias: Comunicación y Pedagogía*, (205), 46-49.

Dietz, A., Quach, W, Lund, S. & McKelvey, M. (2012), AAC assessment and clinical decision-making: The impact of experience. *Augmentative and Alternative Communication*, 28, 149–159. doi:10.3109/07434618.2012.704521

Duchier, D. & Debusmann, R. (2001), Topological Dependency Trees: A Constraint-Based Account of Linear Precedence. *Proceedings of the 39th Annual Meeting of the Association for Computational Linguistics (ACL)*, pp. 180–187.

Duchier, D. (2003), Configuration of Labeled Trees under Lexicalized Constraints and Principles. *Research on Language and Computation*, 1: pp. 307-336.

Eisner, J. M. (2000), Bilexical Grammars and their Cubic-time Parsing Algorithms. In Bunt, H. and Nijholt, A. (eds), *Advances in Probabilistic and Other Parsing Technologies*, Kluwer, pp. 29–62.

Encarnação P., Adams K., Cook A., Nunes da Ponte M., Caiado A., Leite T. & Londral A. (2014), Cultural issues in implementing an integrated augmentative communication and manipulation assistive technology for academic activities. *In Proceedings of the 16th Biennal Conference of ISAAC: Discover Communication.* Canada. ISBN: 978-0-9881189-2-8.

Escoin, J. (2006), Tecnología informática en comunicación aumentativa. *Minusval*, Diciembre 2006, Madrid, pp. 23-25.

Fillmore, C. J. (1976), Frame Semantics and the Nature of Language. *Annals of the New York Academy of Sciences*, 280: 20–32. doi:10.1111/j.1749-6632.1976.tb25467.x

Gaifman, H. (1965), Dependency Systems and Phrase-structure Systems. *Information and Control* 8: 304–337.

Ganz, J. B., Boles, M. B., Goodwyn, F. D. & Flores, M. M. (2014), Efficacy of handheld electronic visual supports to enhance vocabulary in children with ASD. *Focus on Autism and Other Developmental Disabilities*, 29, 3–12. doi:10.1177/1088357613504991

Garay-Vitoria, N. & Abascal, J., (2004), A Comparison of Prediction Techniques to Enhance the Communication Rate. *Lecture Notes in Computer Science*. Springer, Berlin, Vol. 3196/2004, pp. 400-417.

Garay-Vitoria, N. & Abascal, J. (2006), Text prediction systems: a survey. *Universal Access in the Information Society*, *4*(3), 188-203.

García, P., Lleida, E., Castán, D., Marcos, J. M. & Romero, D. (2015), Context-Aware Communicator for All. *In International Conference on Universal Access in Human-Computer Interaction*, pp. 426-437. Springer International Publishing.

Garcia, L. F., de Oliveira, L. C. & de Matos, D. M. (2016), Evaluating pictogram prediction in a location-aware augmentative and alternative communication system. *Assistive Technology*, *28*(2), 83-92.

Golinker, L. (2009), Making and Remaking the Case for AAC. *Augmentative Communication News*, Dec 2009, Vol. 21, Num. 4, pp. 10-13.

González Rus, G., & Liébana, L (2013), Nuevos materiales digitales para aulas y centros específicos. *Revista Didáctica, Innovación y Multimedia*, núm. 26. Retrieved at http://dim.pangea.org/revista26

Grove, N., & Walker, M. (1990). The Makaton Vocabulary: Using manual signs and graphic symbols to develop interpersonal communication. *Augmentative and Alternative Communication*, 6(1), 15-28.

Hamilton, H. & Kingsbury, E. (2006), Participatory Design of AAC Systems: Practical Strategies & Positive Outcomes. *CCC-SLP*. 2006. Retrieved at https://goo.gl/rOZXLz

Hays, D-G. (1964), Dependency Theory: A Formalism and some Observations. *Language* 40: pp. 511–525.

Hedge, M. N. & Pomaville, F. (2008), Assessment of communication disorders in children. *Section 6: Nonverbal and Minimally Verbal Children’s Assessment Protocols*. San Diego: Plural Publishing Inc. ISBN: 9781597567848.

Higginbotham, J. (2007), Access to AAC: Present, Past, and Future. *Augmentative Alternative Communication*. Vol. 23, No. 3, pp. 243-257.

Hill, K. & Romich, B. (2002), A Rate Index for Augmentative and Alternative Communication. *International Journal of Speech Technology*, 5, 57-64. doi: 10.1023/A:1013638916623

Hill, K. & Harkawik, R. (2011), Tools and Resources to Conduct Comprehensive AAC Assessment. *ICAN Talk Clinics of the AAC Institute*. 2011. University of Pittsburgh. Retrieved at https://goo.gl/iXxm4G

Horner, R. H. & Albin, R. W. (1988), Research on general-case procedures for learners with severe disabilities. *Education and Treatment of Children*, 11(4), 375–388. Retrieved at https://www.jstor.org/stable/42899085

Horstmann Koester, H. & Arthanat, S. (2017), Text entry rate of access interfaces used by people with physical disabilities: A systematic review. *Assistive Technology*. doi: 10.1080/10400435.2017.1291544

Hustad, K. C., Keppner, K., Schanz, A. & Berg, A. (2008), Augmentative and alternative communication for preschool children: Intervention goals and use of technology. *Seminars in Speech & Language*, 29, 83–91.

Islas, L., González, V. M., & Mejía, M. (2013), Developing a Mobile Application for Language Disabled Children with User Centered Design. *In International Conference on Human-Computer Interaction*, pp. 236-240. Springer, Berlin, Heidelberg.

Järvinen, T. & Tapanainen, P. (1998), Towards an Implementable Dependency Grammar. *In Kahane, S. and Polguère, A. (eds), Proceedings of the Workshop on Processing of Dependency-Based Grammars*, pp. 1–10.

Johnston, S., Reichle, J., Feeley, K. & Jones, E. (2012), AAC strategies for individuals with moderate to severe disabilities. *Baltimore: Paul H. Brookes*. ISBN: 9781598572063.

Joseph, L. M. & Konrad, M. (2009), Teaching students with intellectual or developmental disabilities to write: A review of the literature. *Research in Developmental Disabilities*, 30 (1), 1–19.

Kagohara, D. M., van der Meer, L., Ramdoss, S., O'Reilly, M. F., Lancioni, G. E., Davis, T. N. & Sigafoos, J. (2013), Using iPods® and iPads® in teaching programs for individuals with developmental disabilities: A systematic review. *Research in Developmental Disabilities*, 34, 147–156. doi:10.1016/j.ridd.2012.07.027

Kaplan, E., Goodglass, H. & Weintraub, S. (1983), The Boston naming test. 2nd. *Philadelphia: Lea & Febiger*.

Karberis, G. & Kouroupetroglou, G. (2002), Transforming Spontaneous Telegraphic Language to Well-Formed Greek Sentences for Alternative and Augmentative Communication. *Lecture Notes in Computer Science*, Vol. 2308/2002, *Methods and Applications of Artificial Intelligence*, pp. 745-756. doi:10.1007/3-540-46014-4_15

Karlsson, F. (1990), Constraint Grammar as a Framework for Parsing Running Text. *In Proceedings of the 13th Conference on Computational linguistics: the 13th conference*, 168-173, Helsinki, Finland. doi:10.3115/991146. 991176

Kent, R. (Eds). (2007), Communicative Disorders Review, *Plural Publishing Inc*, San Diego.

Kim, G., Park, J., Han, M., Park, S. & Ha, S. (2009), Context-aware communication support system with pictographic cards. *In Proceedings of the 11th International Conference on Human-Computer Interaction with Mobile Devices and Services*, p. 86. ACM.

Koehn, P. (2009), Statistical machine translation. *Cambridge University Press*.

Kraat, A. (1990), Augmentative and Alternative Communication: Does it Have a Future in Aphasia Rehabilitation?, *Aphasiology*, Vol. 4, Issue 4, pp. 321-328.

Larranz, C. (2006), La comunicación aumentativa, *Minusval*, Diciembre 2006, Madrid, pp. 17-19.

Lee, E. S., Hwang, E. J., Hur, T. S., Woo, Y. S. & Min, H. K. (2003), A study on the predicate prediction using symbols in augmentative and alternative communication system. In *International Conference on Mobile Human-Computer Interaction*, pp. 466-470. Springer, Berlin, Heidelberg.

Lee E., Nam D., Kim Y., Hur T., Woo Y. & Min H. (2006), Verb Prediction and the Use of Particles for Generating Sentences in AAC System. *AI 2006: Advances in Artificial Intelligence. AI 2006. Lecture Notes in Computer Science*, vol 4304. doi: 10.1007/11941439_111

Lesher G., Moulton B., Higginbotham J. (1998), Techniques for Augmenting Scanning Communication. *Augmentative Alternative Communication*, Vol. 14, pp. 81–101.

Light, J. & McNaughton, D, (2013). Putting people first: Re-thinking the role of technology in augmentative and alternative communication intervention. *Augmentative and Alternative Communication*, 29(4), 299-309. doi:10.3109/07434618.2013.848935

Luque, F. C. & López, E. B. (2012), Símbolos pictográficos de ARASAAC: ¿Son adecuados?. *J. Navarro, M.T. Fernández, F.J. Soto, F. Tortosa (Coords)*.

Maruyama, H. (1990), Structural Disambiguation with Constraint Propagation. *Proceedings of the 28th Meeting of the Association for Computational Linguistics (ACL)*, Pittsburgh, PA, pp. 31–38.

Mathisen, B., Arthur-Kelly, M., Kidd, J., & Nissen, C. (2009). Using MINSPEAK: A case study of a preschool child with complex communication needs. *Disability and Rehabilitation: Assistive Technology*, 4(5), 376-383.

Maia, M. & Cruz, M. (2010), Software Plaphoons na comunicação de indivíduo com paralisia cerebral. *Saber & Educar*, (15).

McLeod, L. (2011), Game changer. *Perspectives on augmentative and alternative communication*, 20(1), 17-18.

McNaughton D., Light J. & Gulla S. (2003), Opening Up a 'whole new world': Employer and co-worker perspectives on working with individuals who use augmentative and alternative Communication. *Augmentative and Alternative Communication*, 19, 235–253. doi: 10.1080/07434610310001595669

McNaughton, D. & Light, J. (2013), The iPad and mobile technology revolution: Benefits and challenges for individuals who require augmentative and alternative communication. *Augmentative and Alternative Communication*, 29, 107–116. doi:10.3109/07434618.2013. 784930.

McNaughton, D. & Light, J. (2015), What we write about when we write about AAC: The past 30 years of research and future directions. *Augmentative and Alternative Communication*, 31, 261–270. doi:10.3109/07434618.2015. 1099736

Menzel, W. & Schröder, I. (1998), Decision procedures for dependency parsing using graded constraints. *In Kahane, S. and Polguère, A. (eds), Proceedings of the Workshop on Processing of Dependency-Based Grammars*, pp. 78–87.

Nivre, J. (2005), Dependency Grammar and Dependency Parsing. *Technical Report MSI report 05133*, Växjö University: School of Mathematics and Systems Engineering.

Pennington, C-A., & McCoy, K-F. (1998), Providing Intelligent Language Feedback for Augmentative Communication Users, *in Assistive Technology and Artificial Intelligence, Lecture Notes in Computer Science*, vol. 1458, Springer Verlag. doi:10.1007/BFb0055970

Peña-Casanova, J. (1991), Normalidad, semiología y patología neuropsicológicas. Programa Integrado de Exploración Neuropsicológica. *Test Barcelona*. Barcelona: Masson, 1991.

Randolph, J. J. (2005), Free-Marginal Multirater Kappa (multirater K [free]): An Alternative to Fleiss' Fixed-Marginal Multirater Kappa. *Online submission*.

Reichle, J. and Drager, K. (2010), Examining Issues of Aided Communication Display and Navigational Strategies for Young Children with Developmental Disabilities. *Journal of Developmental and Physical Disabilities*, 2010, Vol. 22, Num. 3, pp. 289-311.

Reichle, J., Ganz, J., Drager, K. & Parker-McGowan, Q. (2016), Augmentative and Alternative Communication Applications for Persons with ASD and Complex

Communication Needs. *In Prelinguistic and Minimally Verbal Communicators on the Autism Spectrum*, pp. 179-213. Springer Singapore. doi:10.1007/978-981-10-0713-2

Richardson, W., Wilson, M., Nishikawa, J. & Hayward, R. (1995), The well-built clinical question: A key to evidence-based decisions. *ACP Journal Club*, 123, A12–A13.

Schlosser R.W., Koul R. & Costello J. (2006), Asking well-built questions for evidence-based practice in augmentative and alternative communication. *J Commun Disord 2007*, 40, 225–38. doi:10.1016/j.jcomdis.2006.06.008

Sennott, S. & Bowker, A. (2009), Autism, aac, and proloquo2go. *Perspectives on Augmentative and Alternative Communication*, 18(4), 137-145.

Shane, H. C., Abramson, J. S., Corley, K., Fadie, H., Flynn, S., Laubscher, E. & Sorce, J. (2015), Enhancing communication for individuals with autism: A guide to the visual immersion system. *Baltimore: Brookes Publishing*.

Sleator, D. & Temperley, D. (1993), Parsing English with a Link Grammar. *Third International Workshop on Parsing Technologies (IWPT)*, pp. 277–292.

Soro, E., Basil, C., Rosell, C. & Boix, J. (2007), Vocabulari CACE-UTAC. *Grup d'Investigació per l'Estudi del Llenguatge i la CAA*, Universitat de Barcelona. Retrieved at https:// goo.gl/eFTPzy

Taylor, A., Arnott, J. & Alm, N. (2001), Visualisation to Assist Non-speaking Users of Augmentative Communication Systems. *In Proceedings of the International Conference of the IEEE on Information Visualisation*, London, UK, pp. 251–256.

Vaillant, P. (1997), A Semantics-Based Communication System for Dysphasic Subjects. *Lecture Notes in Computer Science, Vol. 1211, Artificial Intelligence in Medicine*, (pp. 381-392). doi:10.1007/BFb0029471

Waller, A. & Jack, K. (2002), A predictive Blissymbolic to English translation system. *Proceedings of the fifth international ACM conference on Assistive technologies*, Edinburgh, Scotland. doi: 10.1145/638249.638283

Warrens, M. J. (2010), Inequalities between multi-rater kappas. *Advances in data analysis and classification*, 4(4), 271-286.

Westling, D. & Fox, L. (2009), Teaching students with severe disabilities (4th ed.). *Upper Saddle River*, NJ: Merrill Prentice Hall.

Wiegand, K. & Patel, R. (2014), RSVP-iconCHAT: A Single-Switch, Icon-Based AAC Interface. *Journal on Technology & Persons with Disabilities, 29th Annual International Technology and Persons with Disabilities Conference*, San Diego. Retrieved at http://hdl.handle.net/10211.3/133375

Yovetich, W., & Young, T. (1988). The effects of representativeness and concreteness on the ‘‘guessability’’ of Blissymbols. *Augmentative and Alternative Communication*, 4, 35–39.

Zygo Industries Inc (2004). Dialect with Speaking Dynamically Pro, Boardmaker, and SAPI speech synthesis. *Portland, OR: Zygo Industries, Inc*.

Appendix A

Apostrophication rules of singular definite articles in Catalan

As a general rule, the definite articles "el" and "la" are reduced to "l'" when the word that follows begins with a vowel or a silent "h": "l'avi", "l'invent", "l'euga", "l'ull", l'hivern", etc.

This apostrophication rule presents the following exceptions, in addition to the common apostrophication rules of the definite articles and the preposition "de" exposed later on:

1. Before a feminine word beginning with an unstressed "i" or "u" (preceded or not by a silent "h"), the full form of the article "la" is mantained: "la idea", "la il·lusió", "la hipòtsi", "la història", "la unitat", "la universitat", "la humitat", etc.
2. In front of the words "una" (when it refers to an hour), "ira" and "host", the full form is used: "la una", "la ira", "la host".

Remarks: In front of Roman and Arabic numerals, the article adopts the reduced form only if when written in alphabetic characters the article is also apostrophied: "l'1 de gener" (= "l'u de gener), l'XI (= "l'onze), et**c.**

Common apostrophication rules of the definite articles and the preposition "de":

1. When the following word begins with a non vocalic "i" or "u", the full form of the article, both masculine and feminine, is maintained: "el iaio", "el ioga", "el iogurt", "el hiatus", "el hioide", "el uigur", "el uombat", "la iarda", "la ionosfera", "la hialita", "la hiena", etc. Exception: "l'ió".

2. Definite articles are not apostrophied in front of borrowed words or foreign proper nouns that start with a voiced "h": "el hippy", etc.

3. Definite articles are never apostrophied before the name of the letters of the alphabet: "la a", "la ene", "la hac", "la ele", etc.

4. The common rules for apostrophication are also applied to words in quotation marks or italics. The only exception is when italics are used for metalinguistic purposes.

5. Preceding numerical symbols or abbreviations, which start with a vocalic sound, definite articles are also apostrophied: "l'1 de maig", "l'XI Congrés", "l'ap. 4", etc.

6. General rules are also applied to acronyms, which are read like a word: "l'ONU", "l'IVA", "l'URSS", "la UNESCO", "la UEFA", etc.

7. Definite articles are apostrophied before acronyms, which are spelled and which start with a vocalic sound: "l'EMT", "l'AVL", "l'ONG", "l'FM", "l'LSD", etc.

8. In front of borrowed or foreign words that begin with a liquid "s", the masculine definite article accepts apostrophication or lack of it ("el speaker" or "l'speaker"), while the feminine definite always presents its full form ("la Scala"). Nevertheless, the most coherent solution is to never apostrophy in any case: "el speaker", "la Scala", etc.

Appendix B

100 target sentences for the technical tests of the Companion System

#	Catalan	Spanish	English
1	He menjat molt.	He comido mucho.	I ate a lot.
2	Vull anar al lavabo.	Quiero ir al baño.	I want to go to the bathroom.
3	Quina hora és?	¿Qué hora es?	What time is it?
4	Anirem al restaurant.	Iremos al restaurante.	We will go to the restaurant.
5	El vestit és nou.	El vestido es nuevo.	The dress is new.
6	Teniu tomàquets?	¿Tenéis tomates?	Have you got tomatoes?
7	He caigut.	Me he caído.	I fell.
8	El meu gos és molt graciós.	Mi perro es muy gracioso.	My dog is very funny.
9	Puc jugar a pilota, si us plau?	¿Puedo jugar a pelota, por favor?	Can I play ball, please?
10	Està marejat.	Está mareado.	He is dizzy.
11	On és la meva nina?	¿Dónde está mi muñeca?	Where is my doll?
12	No tinc la cadira de rodes.	No tengo la silla de ruedas.	I don't have the wheelchair.
13	M’agrada la neu.	Me gusta la nieve.	I like snow.
14	Vull dormir.	Quiero dormir.	I want to sleep.
15	Els macarrons són molt bons.	Los macarrones están muy buenos.	Macarroni are very good.
16	A la tarda necessitaré el medicament.	Por la tarde necesitaré el medicamento.	I will need the medicine in the afternoon.

17	Dona'm una forquilla, si us plau.	Dame un tenedor, por favor.	Give me the fork, please.
18	El monitor no és a l'escola.	El monitor no está en la escuela.	The monitor is not at the school.
19	Vull una poma.	Quiero una manzana.	I want an apple.
20	La piscina és molt guai.	La piscina es muy guay.	The swimming pool is very cool.
21	Tot és molt divertit.	Todo es muy divertido.	Everything is very fun.
22	Vindré tard.	Vendré tarde.	I will come late.
23	No vull més verdura.	No quiero más verdura.	I don't want any more vegetables.
24	Quantes croquetes vols?	¿Cuántas croquetas quieres?	How many croquettes do you want?
25	Tinc fred.	Tengo frío.	I'm cold.
26	Vols que anem a casa?	¿Quieres que vayamos a casa?	Do you want us to go home?
27	Espera'm al bar.	Espérame en el bar.	Wait for me at the bar.
28	Vaig a la biblioteca.	Voy a la biblioteca.	I go to the library.
29	Amb qui vas?	¿Con quién vas?	Who are you going with?
30	Ahir, vaig comprar unes sabates blaves.	Ayer compré unos zapatos azules.	Yesterday, I bought a pair of blue shoes.
31	Un bitxo molt raro s'ha amagat darrere la porta.	Un bicho muy raro se ha escondido detrás de la puerta.	A very weird bug hid behind the door.
32	És un pal.	Es un palo.	It's a pain.
33	No vinguis.	No vengas.	Don't come.
34	Tinc molta por.	Tengo mucho miedo.	I'm really afraid.
35	Vine.	Ven.	Come.
36	Quan vindreu a casa?	¿Cuándo vendréis a casa?	When are you coming home?
37	Vindrem demà.	Vendremos mañana.	We will come tomorrow.
38	Menjo un gelat de xocolata.	Como un helado de chocolate.	I eat a chocolate ice-cream.

39	M'agrada llegir llibres de por.	Me gusta leer libros de miedo.	I like reading terror books.
40	Avui no agafis l'abric.	Hoy no cojas el abrigo.	Today, do not grab the coat.
41	A l'estiu ens banyarem a la platja.	En verano nos bañaremos en la playa.	In summer we will bath in the sea.
42	Ahir vaig anar al metge.	Ayer fui al médico.	Yesterday I went to the doctor.
43	L'home del jersei negre és dolent.	El hombre del jersey negro es malo.	The man on the black jersey is mean.
44	No m'agrada el pastís de poma.	No me gusta el pastel de manzana.	I don't like apple pie.
45	Dimecres anirem a cantar.	El miércoles iremos a cantar.	We will sing on Wednesday.
46	Hem begut un got de llet calenta.	Hemos bebido un vaso de leche caliente.	We drank a glass of warm milk.
47	Què fa la mare?	¿Qué hace mi madre?	What is mother doing?
48	La infermera és molt alegre.	La enfermera es muy alegre.	The nurse is very cheerful.
49	Vull classificar-me per les olimpíades.	Quiero clasificarme para las olimpíadas.	I want to classify for the Olympics.
50	Un iogurt, si us plau.	Un yogurt, por favor.	A yoghurt, please.
51	Parla molt?	¿Habla mucho?	Does he/she talk a lot?
52	La meva germana llegeix malament.	Mi hermana lee mal.	My sister reads poorly.
53	M'he classificat primer.	Me he clasificado primero.	I got first place.
54	Ho sento.	Lo siento.	I'm sorry.
55	Hola!	¡Hola!	Hello!
56	El sol és taronja i vermell a la tarda.	El sol es naranja y rojo por la tarde.	The sun is orange and red in the afternoon.
57	Un got d'aigua, si us plau.	Un vaso de agua, por favor.	A glass of water, please.
58	Estic bé.	Estoy bien.	I'm fine.
59	Felicitats!	¡Felicidades!	Congratulations!

60	Ajuda'm a baixar les escales, si us plau.	Ayúdame a bajar las escaleras, por favor.	Help me go down the stairs, please.
61	Estàvem molt contents.	Estábamos muy contentos.	We were very happy.
62	Espera un minut.	Espera un minuto.	Wait a minute.
63	No vull que et barallis.	No quiero que pelees.	I don't want you to fight.
64	El tren anava molt lent.	El tren iba muy lento.	The train was very slow.
65	Això serà una sorpresa.	Esto será una sorpresa.	This will be a surprise.
66	Tinc mal al peu.	Tengo dolor en el pie.	My foot is in pain.
67	Les cinc.	Las cinco.	Five o'clock.
68	T'estimo.	Te quiero.	I love you.
69	El cuiner i la cuinera fan pastissos.	El cocinero y la cocinera hacen pasteles.	The cooks make cakes.
70	Tu ets molt més alt.	Tú eres mucho más alto.	You are much taller.
71	Per què?	¿Por qué?	Why?
72	A qui ho explico?	¿A quién lo cuento?	Who do I tell?
73	Dijous serem a Europa.	El jueves estaremos en Europa.	We will be in Europe on Thursday.
74	Com ho sabrem?	¿Cómo lo sabremos?	How will we know?
75	La mare està amb el pare.	Mi madre está con mi padre.	My mother is with my father.
76	No.	No.	No.
77	No ho sé.	No lo sé.	I don't know.
78	Canvia la cadira de color marró.	Cambia la silla de color marrón.	Change the brown chair.
79	Ja està.	Ya está.	I'm done. / It's done.
80	Mala sort.	Mala suerte.	Too bad.
81	Espero que vinguis demà.	Espero que vengas mañana.	I hope that you'll come tomorrow.
82	Amaga't.	Escóndete.	Hide yourself.

83	Celebren el seu aniversari al parc.	Celebran su aniversario en el parque.	They celebrate his birthday at the park.
84	Compra això a la farmàcia.	Compra esto en la farmacia.	Buy this at the pharmacy.
85	Vam beure cafè al casament.	Bebimos café en la boda.	We drank coffee at the wedding.
86	Cinc persones estaven incòmodes dins el taxi.	Cinco personas estaban incómodas dentro del taxi.	Five people were uncomfortable inside the taxi.
87	La meva gossa necessita una dutxa.	Mi perro necesita una ducha.	My dog needs a shower.
88	Horrible.	Horrible.	Horrible.
89	Bon any.	Feliz año.	Happy new year.
90	La meva amiga està trista.	Mi amiga está triste.	My friend is sad.
91	He aparcat el cotxe a la plaça.	He aparcado el coche en la plaza.	I parked the car at the square.
92	Cent euros.	Cien euros.	A hundred euros.
93	La meva germana descansa sobre el llit.	Mi hermana descansa sobre la cama.	My sisters rests on the bed.
94	Explica'm un conte.	Cuéntame un cuento.	Tell me a story.
95	Després mirarem una pel·lícula al cine.	Después miraremos una película en el cine.	Afterwards, we will watch a movie at the cinema.
96	Jugareu més tard.	Jugaréis más tarde.	You will play later.
97	El bebè caminarà aviat.	El bebé andará pronto.	The baby will walk soon.
98	També vull una bufanda lila i llarga.	También quiero una bufanda lila y larga.	I also want a long and purple scarf.
99	Bona nit.	Buenas noches.	Good night.
100	Adéu.	Adiós.	Bye bye.

Appendix C

Data Sheet Evaluation for Test Sessions of the Companion System

1. General data

Participant's name	**Date**	**# session**	**Duration**

Session ended early?	**Reason:**
Yes ☐ / No ☐	

2. Evaluation at the end of the session

Speech language pathologist's rating	**Participant's rating**
0 1 2 3 4 5	0 1 2 3 4 5

Perceived emotions (joy, surprise, frustration, excitement, etc.):

Compared to the previous session?	**Which aspects?**
Better ☐ Worse ☐ Similar ☐	

Comments:

3. Later evaluation of the constructed sentences

<table>
<tr><td>Mean pictograms</td><td colspan="2">Mean time</td></tr>
<tr><td></td><td colspan="2"></td></tr>
<tr><td colspan="3">Characteristics of the compansion system used (subject reduction, verb reduction, verb tenses, type of sentence modifiers, etc.):</td></tr>
<tr><td>Compared to the previous session?</td><td colspan="2" rowspan="2">Which aspects (more structured sentences, etc.)?</td></tr>
<tr><td>Better ☐
Similar ☐
Worse ☐</td></tr>
<tr><td colspan="3">Comments:</td></tr>
</table>

4. Sentences done during the session

Participant's name	Participant Id	# session

#	Id	Generated sentence	Rating	Time
1			0 1 2 3 4 5	
2			0 1 2 3 4 5	
3			0 1 2 3 4 5	
4			0 1 2 3 4 5	
5			0 1 2 3 4 5	
6			0 1 2 3 4 5	
7			0 1 2 3 4 5	
8			0 1 2 3 4 5	
9			0 1 2 3 4 5	
10			0 1 2 3 4 5	
11			0 1 2 3 4 5	
12			0 1 2 3 4 5	
13			0 1 2 3 4 5	
14			0 1 2 3 4 5	