Université Fédérale
Toulouse Midi-Pyrénées

THÈSE

AIRBUS

En vue de l'obtention du
DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

Délivré par l'Université Toulouse 2 - Jean Jaurès

Présentée et soutenue par

Nataly JAHCHAN

Le 15 février 2019

To What Extent Does Text Simplification Improve Human Comprehension? Cognitive Evaluations for the Optimization of the Airbus Cockpit Controlled Language for Future Aircraft

---

Ecole doctorale : **CLESCO - Comportement, Langage, Education, Socialisation, Cognition**

Spécialité : **Sciences du langage**

Unité de recherche :
CLLE - Unité Cognition, Langues, Langage, Ergonomie

Thèse dirigée par
Anne CONDAMINES et Hélène GIRAUDO

Jury
M. Christopher GLEDHILL, Rapporteur
M. Mathieu VALETTE, Rapporteur
Mme Stéphanie CAILLIES, Examinatrice
M. André TRICOT, Examinateur
Mme Anne CONDAMINES, Directrice de thèse
Mme Hélène GIRAUDO, Co-directrice de thèse
Mme Emmanuelle CANNESSON, Responsable technique (AIRBUS)

Thèse

En vue de l'obtention du

Doctorat de l'université de Toulouse

Délivré par

Université Toulouse - Jean Jaurès

■

Présentée et soutenue par

Nataly Jahchan

Le 15 février 2019

Titre

To What Extent Does Text Simplification Improve Human Comprehension? Cognitive Evaluations for the Optimization of the Airbus Cockpit Controlled Language for Future Aircraft

■

École doctorale et discipline ou spécialité

ED CLESCO: Science du langage

Unité de recherche

CLLE (UMR 5263, CNRS)

Directrices de Thèse

Anne Condamines – *Directrice* (Directrice de Recherche, CNRS)

Hélène Giraudo – *Co-Directrice* (Chargée de Recherche HDR, CNRS)

Emmanuelle Cannesson – *Responsable Technique* (AIRBUS), *invitée*

Jury

Christopher Gledhill, Professeur des Universités, (Université Paris Diderot), *rapporteur*

Mathieu Valette, Professeur des Universités, (INALCO), *rapporteur*

Stéphanie Caillies, Professeure des Universités, (Université Reims Champagne-Ardenne), *examinatrice*

André Tricot, Professeur des Universités, (ESPE, Université Toulouse Jean Jaurès), *examinateur*

When you can measure what you are speaking about, and express it in numbers, you know something about it; but when you cannot measure it, when you cannot express it in numbers, your knowledge is of a meager and unsatisfactory kind. It may be the beginning of knowledge, but you have scarcely, in your thoughts, advanced to the stage of science.

— Lord Kelvin, *Electrical Units of Measurement* (1883)

Abstract

THIS PhD was initiated by the Human Factors and Ergonomics in Design department at Airbus with the aim of optimizing an existing cockpit controlled language to integrate in future disruptive design. The need for clear and unambiguous communication is vital in safety critical domains, and the current controlled language was carefully constructed to avoid ambiguity and complexity, and is designed to help pilots operate and navigate the aircraft (with the help of cockpit screen interfaces) in normal and abnormal (in cases of emergency or failures) situations. In order to optimize the existing language, we set out to assess the appropriate levels of simplification that would achieve more accurate and faster comprehension with minimum pilot training. We first delved into the controlled language domain to form an overview of the existing controlled languages, their context, and rules. From this research we attempted to find solutions for optimization, but at the same time we strove to offer an original contribution to the field through this work.

In order to test and improve comprehension, perception, and use of controlled languages in the cockpits (and to offer new assessment techniques for evaluating other controlled languages), we conducted evaluations by taking advantage of new tools and research in the cognitive sciences and controlled languages domains to apply linguistic hypotheses concerning text simplification limits. More particularly: might a more natural syntax help pave the way for better pilot comprehension and faster reaction times?

Results show that the Airbus controlled language could benefit from more natural language structures to enhance pilot comprehension and reduce training times. This new more optimized language fits effectively into the future disruptive cockpit concept and its more intuitive designs. We also show that there is a noticeable lack of controlled language evaluations in the field, as well as adequate methods of evaluating linguistic hypotheses using firm cognitive sciences methodology to satisfy ergonomic needs. We propose that in the future, all controlled language rules be systematically evaluated before being applied, especially in safety critical domains, to ensure that the prescriptive and proscriptive rules that make them are as efficient as possible, and that they truly reduce ambiguity and improve human comprehension and performance.

Résumé

CETTE thèse a pour origine la volonté d'Airbus d'améliorer la langue contrôlée utilisée dans les cockpits de ses futurs avions. Une communication claire et non ambiguë est essentielle dans les domaines où la sécurité est engagée. La langue contrôlée actuellement utilisée dans les cockpits Airbus, a été soigneusement élaborée pour éviter toute ambiguïté et complexité. Elle est conçue pour aider les pilotes dans leur tâche de pilotage en temps normal et dans des situations anormales (en cas d'urgence ou de défaillance). Afin d'optimiser la langue existante, nous avons entrepris d'évaluer les niveaux appropriés de simplification qui permettraient une compréhension plus précise et plus rapide, réduisant ainsi le temps de formation des pilotes. Nous avons tout d'abord exploré le domaine des langues contrôlées afin d'avoir un aperçu des langues contrôlées existantes, de leur contexte et de leurs règles. À partir de cette recherche, nous avons tenté de trouver des solutions d'optimisation, tout en nous efforçant d'apporter une contribution originale à ce domaine.

Afin de tester et d'améliorer la compréhension, la perception et l'utilisation de langues contrôlées dans les cockpits (et d'offrir de nouvelles techniques d'évaluation pour évaluer d'autres langues contrôlées), nous avons effectué des évaluations en tirant parti des nouveaux outils expérimentaux et des recherches en sciences cognitives ainsi que dans le domaine des langues contrôlées pour appliquer des hypothèses linguistiques concernant les limites de simplification/contrôle d'un texte. Plus particulièrement : Une syntaxe plus naturelle pourrait-elle optimiser la compréhension du pilote et réduire le temps de réaction ?

Les résultats montrent que la langue contrôlée Airbus pourrait tirer parti de structures de langue plus naturelles pour améliorer la compréhension des pilotes et réduire le temps de formation à l'utilisation de cette langue contrôlée. Cette nouvelle langue plus optimisée s'intègre parfaitement au futur concept de cockpit disruptif et à ses conceptions plus intuitives. Nous montrons également qu'il existe un manque notable d'évaluations de langues contrôlées sur le terrain, ainsi que de méthodes adéquates pour évaluer les hypothèses linguistiques à l'aide de méthodologies des sciences cognitives qui répondent aux besoins ergonomiques. Nous proposons qu'à l'avenir, les règles de langues contrôlées en général fassent l'objet des évaluations cognitives systématiques avant d'être appliquées, en particulier dans les domaines où la sécurité est engagée, afin de garantir que les règles prescriptives et proscriptives soient aussi efficaces que possible, et qu'elles réduisent significativement l'ambiguïté et améliorent la compréhension humaine et la performance.

Acknowledgments

Firstly, I would like to express my sincere gratitude to my thesis advisor Anne Condamines who helped shape my way of thought since day one, and gave this work a solid theoretical background from which it grew. She was always there to guide and encourage me throughout the PhD, and offer friendly advice when needed. Her availability and many reviews and corrections helped make this work into what it is.

I would also like to thank Hélène Giraudo, my co-advisor, who always provided me with the necessary tools and means to conduct my experiments in the most optimal manner, and who was supportive and encouraging throughout the accomplishment of the work.

I am also very grateful for Christopher Gledhill and Mathieu Valette for agreeing to review and judge my work in this festive season, and for Stéphanie Caillies and André Tricot for accepting to be part of the jury of my thesis defense.

A very special thanks to Airbus for allowing me to work on this very interesting subject that was a big source of motivation all along the ups and downs of my PhD. I want to thank all of my colleagues and ex-interns in the human factors department without whom it would have been difficult to survive on a day to day basis. For the many theoretical and philosophical debates during lunch and on coffee breaks that have molded the way I think, and made me rethink my beliefs on many an occasion!

In particular, I would like to thank the M01 gang: Kevin who is always the life of the party (and the debates!), les 2 Caro(s) who made a late entry into the department but who have made it with a bang, Lolo who always makes everyone smile, Amine my fellow PhD candidate and companion on countless coffee breaks, Christine whose good humor, honesty, and encouragement always put everyone in a better mood (and whose desserts must be submitted in professional competitions!), Luigi Laura and Marina who recently joined the team and added a much appreciated playful component, Laetitia Benoit and Sonja who first welcomed me as an intern and taught me a lot on life and things, Estelle a fellow linguist whose take on society I find intruiging, Marielle Thomas and Jérôme who we could always rely on for good jokes. And Florence who always made sure I'm on the right track.

A special mention to Kevin who was a tremendous help in the construction of the experimental protocols, who managed to make me enjoy and learn a lot about

statistical analysis (by always answering my (not-so-quick) quick questions), and with whom I enjoyed many deep scientific and philosophical discussions, which ranged from the evil effects of averages to the physical realities of a falling elevator.

I would like to thank my lab mates and fellow doctoral candidates who were always there when I needed it, to complain about our deadlines and our students, but also to have fun and forget our PhD woes. Thanks to Pierre-Vincent Paubel, our trusty CLLE lab engineer, who helped me put in place the experiments exactly how I wanted, even when it meant a lot more work for both of us!

I am also very grateful for all my friends and my family for having faith in me. For my loving parents for always supporting me (financially and otherwise) through thick and thin, for believing in me, and for arming me with enough resilience to make the best of things in life.

For my adorable fur ball Spirou who sat beside me through late nights and silently helped me get through it all.

For Maria, Ardas, and Maïlys (my Masters buddies who made my life in Paris much more enjoyable), Pedro and Bernie (who made my time in Chicago become one of my fondest memories), Moony (whose opinions I always value, and whose friendship means the world to me), and Marianne (whose brain I've always admired, and whom I owe a lot for the way I think, for helping me on many occasions throughout my education, and for helping me figure out who I am).

For Christelle, who only knew me during my PhD and who is definitely the biggest victim of it, who supported me through many sleepless nights and moments of doubt, and who continues to surprise me with her smart and interesting take on things. This PhD and I owe a lot to her.

And last but not least, my biggest thanks go to Emma, my advisor at Airbus, who has been consistently and stubbornly there for me throughout my many ups and down, from my first days as a lost intern until the very last day of this PhD. I want to thank her for her cheerful way of being (always making everyone around her happy) and persistently positive attitude that have been a great help throughout this PhD. For always believing in me and pushing me to find my scientific path. For spending many days and weekends reviewing my work and giving me valuable feedback. I am lucky to call her a friend and mentor.

Table of Contents

List of Figures

List of Tables

Acronyms

AECMA	Association Européenne des Constructeurs de Matériel Aérospatial
AECMA-SE	AECMA-Simplified English
AI	Artificial Intelligence
AILA	International Association for Applied Linguistic
AMTs	Aircraft Maintenance Technicians
ARI	Automated Readability Index
ASD-STE	Aerospace and Defence Industries Association of Europe-Simplified Technical English
CEFR	Common European Framework of Reference for Languages
CL	Controlled Language
CLCM	Controlled Language in Crisis Management
CNL	Controlled Natural Language
DRC	Dual-Route Cascaded model
EASA	European Aviation Safety Agency
ECAM	Electronic Centralized Aircraft Monitor
EEG	Electroencephalography
EICAS	Engine Indicating and Crew Alerting System
ENAC	Ecole Nationale de L'Aviation Civile
ERP	Event Related Potentials
FAA	Federal Aviation Administration
FMRi	Functional Magnetic Resonance Imaging
FRES	Flesch Reading-Ease Score
GOLD	Guidelines for Operational Language in Documents
HOCL	Human-Oriented Controlled Languages
ICAO	International Civil Aviation Organization
ILSAM	International Language of Service and Maintenance
INL	Instances of Natural Language
MCL	More Coded controlled Language
MEG	Magnetoencephalography
MFD	Multifunction Display
MNL	More Natural Controlled Language
MOCL	Machine-Oriented Controlled Languages
MRI	Magnetic Resonance Imaging

NL	Natural Language
NLP	Natural Language Processing
OIS	On Board Information System
PENS	Precision, Expressiveness, Naturalness, Simplicity
PET	Positron Emission Tomography
PFD	Primary Flight Display
SE	Simplified English
SLANG	Standard LANGuage
SMI	SensoMotoric Instruments
SMOG	Simple Measure of Gobbledygook
SPO	Single Pilot Operations
TC	Text Complexity
UI	User Interface
UX	User eXperience

I

INTRODUCTION: BACKGROUND AND APPROACH

1

Introduction

This PhD research was launched by the Human Factors and Ergonomics in Design department of Airbus Operations SAS in Toulouse, France in collaboration with CLLE (Cognition, Langues, Langage, Ergonomie) laboratory of Toulouse Jean Jaurès University. Based in Toulouse, an aerospace hub, CLLE laboratory has cultivated a knowledge base in the CNL domain and specialized corpora related to space and aviation.

The main goal is to optimize an existing Airbus Cockpit Controlled Language in order to integrate a new more improved one in future cockpit design. The current controlled language was carefully constructed to avoid ambiguity and complexity (as are all comprehension oriented controlled languages, (cf. Section 3.1.3.2) and is designed to help pilots operate and navigate the aircraft (with the help of cockpit screen interfaces) in normal and abnormal (in cases of emergency or failures) situations. Thus, the need for clear and unambiguous communication is vital in safety critical domains. This language and the rules that make it were put in place at a time when design flexibility was not an option (for example small screen sizes that restrict word and sentence length). As we are addressing a more flexible disruptive cockpit design for future aircraft, these limitations are no longer an issue, and the controlled language need not be so coded and compact, or follow very strict simplification rules. Therefore, in order to optimize the existing language, we set out to assess the appropriate levels of simplification that would achieve more accurate and faster comprehension with minimum pilot training.

To do so, we first delved into the controlled natural language[1] domain to form an overview of the existing controlled languages, their context, and rules. From this research we attempted to find solutions for optimization of the Airbus controlled language, but at the same time we strove to offer an original contribution to the field through this work.

In this sense, our work falls within the realm of Applied Linguistics. AILA[2] (International Association for Applied Linguistics) defines it *"as an interdisciplinary field of research and practice dealing with practical problems of language and communication that can be identified, analysed or solved by applying available theories, methods and results of Linguistics or by developing new theoretical and methodological frameworks in Linguistics to work on these problems."*

Contrary to misconceptions, the field of Applied Linguistics does not simply offer solutions for practical problems from available theories, but could also develop new theoretical frameworks and methodological tools from different fields and sources to deal with language and communication issues. Condamines & Narcy-Combes, 2015 propose the term Situated Science or "science située"*: « [...] situer la science c'est entrer dans une perspective où la recherche n'est plus appliquée à un projet, mais où elle est une partie de ce projet et où les deux se modifient réciproquement au fur et à mesure que le projet avance. »*

Therefore, we start by finding or creating new solutions to concrete real-life problems (in this case, high stakes industrial ones). Those solutions are not mere applications of linguistic knowhow, but also constitute a means to study and advance language functions and assessments in relatively underdeveloped, unknown fields (at least within global Linguistics circles) and where language plays a vital role in ensuring safety (since misinterpretation could lead to potentially catastrophic outcomes).

In order to find solutions, one must carefully investigate the problem in context. In this sense, we are more particularly dealing with Ergonomic Linguistics (Condamines, 2018) in which linguistic models/theories/hypotheses are used in specified work contexts (mainly in industry) to achieve precise goals efficiently. These hypotheses and propositions are derived from actual

1. "CNL", term interchangeable with "CL" or controlled language

2. AILA. [online] Available at: https://aila.info [Accessed 4 Dec. 2018].

language productions and should be evaluated using experimental techniques and acceptability tests.

In this thesis, we use psycholinguistic and cognitive psychology tools and evaluation techniques in order to confirm or deny linguistic hypotheses that are directly linked to a human centred industrial need.

1.1 Thesis Composition

This thesis is composed of 3 parts which are each composed of several chapters. The first part introduces the subject, primary need emanating from Airbus, and the background. We then discuss the literature in the controlled language domain, and describe the Airbus controlled language (the original corpus) and the way its rules were constructed. After that, we proceed to expose the different empirical evaluations that were done on some controlled languages, all of which allowed us to hone our approach and introduce the core question that will define the second part of the thesis: The evaluation and analysis part in which we introduce the 3 experiments (which use psycholinguistic & cognitive sciences techniques) that we put in place to find answers to the core question, and confirm or deny our hypotheses. And finally, the third part constitutes a global overview and discussion of the results. It offers a conclusion and a way forward to the thesis subject, and to the evaluation of controlled languages in general.

2

Background

2.1 Human Factors and Ergonomics at Airbus

Airbus is a commercial aircraft manufacturer, with Space and Defense as well as Helicopters Divisions. It is the largest aeronautics and space company in Europe and a worldwide leader. Airbus designs, manufactures and delivers aerospace products, services and solutions to customers on a global scale.

The Human Factors and Ergonomics in Design department at Airbus consists of a multidisciplinary team composed of linguists, cognitive psychologists, physiologists, and cognitive ergonomists. They organize equipment and functions' evaluations of procedures being designed, and follow-up on the process until the certification phase with aviation authorities such as the EASA (European Aviation Safety Agency) or the FAA (Federal Aviation Administration). They write up reports analysing different functions from a human factors perspective and offer conclusions about the operation, security, and ergonomic aspects of the function under study. These evaluations are done on different parts of the aircraft concerning the equipment in the cockpit, cabin, and maintenance[3].

3. The Human Factors and Ergonomics in Design department in which this PhD was carried out only deals with Cockpit design. Other Human Factors departments within Airbus deal with Cabin Design and Maintenance.

The human factors department also does research to find and implement new solutions and functionalities for the design of future Airbus aircraft. Moreover, they make recommendations for the conception of functions and then provide technical documents. The goal is to ensure that the end user is catered for all along the design process.

The Human factors field has a specific approach to engineering and ergonomics because ergonomists consider matters from a human point of view and take into consideration the multiple possible interactions between the individual and his/her physical and cognitive environment.

Attention to human-machine interactions is extremely important since poorly designed interfaces may result in dangerous situations and safety hazards. Human factors science has become mandatory in certain fields such as medical, transportation, and aviation industries. Issues of aviation safety highlight the importance of the human factors' role in the validation of equipment and functions that would be implemented on commercial aircraft. Technology has advanced in a rapid pace through the years, airplanes have become extremely safe machines, and fatal airliner accidents have been persistently decreasing.

According to Boeing's Aero quarterly magazine (QTR_02, 2007), *"in the early days of flight, approximately 80 percent of accidents were caused by the machine and 20 percent were caused by human error. Today that statistic has reversed. Approximately 80 percent of airplane accidents are due to human error (pilots, air traffic controllers, mechanics, etc.) and 20 percent are due to machine (equipment) failures."* (cf. Figure 1) Even though aviation accidents have been constantly decreasing in general (thanks to more advanced technology), the accidents that do occur nowadays have an 80% chance to be caused by human error, which is why there is a greater need to involve human factors and ergonomics specialists in the design process to mitigate these risks.

More concretely, the human factors team works closely with Airbus test pilots and flight engineers, often on the available flight simulators of the various aircraft, with the aim of testing flight scenarios on existing or newly introduced equipment and functions. Test pilots, who in most cases have had prior extensive careers working for airlines or for the army aboard fighter jets, stand to be experts on the pilot work environment and piloting needs, and are essential for certifying and testing Airbus planes for the general pilot population.

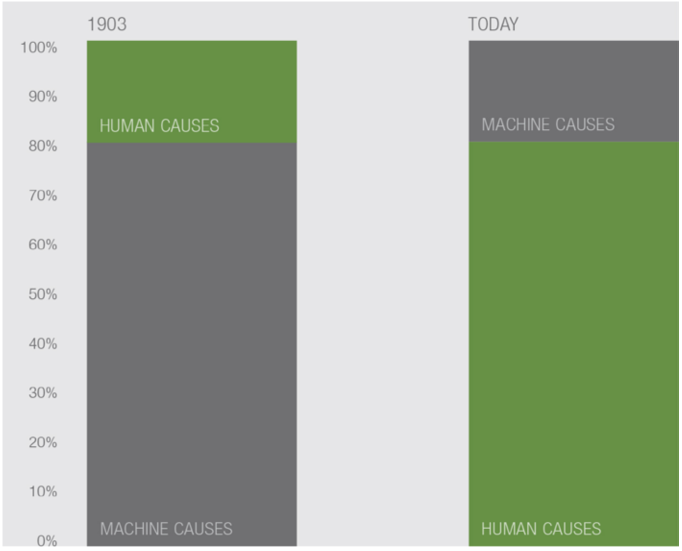


Figure 1. Causes of Aviation Accidents: 1903-2007

2.2 Linguistics in Human Factors and Ergonomics

Human Factors experts use science and general knowledge of human capacities and limitations – along with test pilots' experience – in order to determine the most efficient ways of designing the cockpit.

The physiologists make sure that the physical design of the cockpit is adequate, for example whether pilots are able to reach all the screens and all the buttons easily even during turbulence, or if pilots' field of view is not obstructed by a given instrument or by sun rays. The cognitive ergonomists use evaluation techniques to measure the adequacy of the proposed design on the human cognitive skills in different scenarios. Among other things, the cognitive psychologists measure the effects of fatigue or stress on decision making in cases of high workload.

As for linguists, they use traditional linguistic descriptive research (syntactic, semantic, pragmatic and terminological, etc. theories), cognitive psychology and ergonomics evaluation tools (statistical analyses, questionnaires, etc.) natural language processing (for virtual assistants' technology for instance), and psycholinguistic tools to develop linguistic corpora which are field specific (taking into consideration aviation limitations and specificities) and engineered to be human centered (taking into consideration human capabilities of comprehension and perception).

They have to deal with all the communication and operational issues that could arise in a cockpit environment. The modern glass cockpit (one which features electronic/digital flight instrument displays on screens rather than traditional mechanical gauges) is a somewhat complex work environment (one needs extensive training and expertise in order to operate in such an environment) in which linguistic information is abundant (cf. Figure 2).

Figure 2. Airbus A350 Glass Cockpit

The different monitors contain messages and tags destined to be read and understood by pilots. Some of those messages refer to buttons and levers in the cockpit which themselves have language tags. Additionally, pilots may hear aural alerts which are essentially warning messages that announce incoming dangers or ones that give them additional flight information.

Lastly, pilots flying and pilots monitoring (previously known as pilots and co-pilots) communicate with each other, with the plane itself, with ground air traffic controllers, with different flying aircraft, and with the rest of the crew.

Therefore, amid all this linguistically infused work environment, the linguist's work is first and foremost to leave no room for misinterpretation of intended meaning or ambiguities that could lead to potentially dangerous situations with very high stakes.

In order to do so, linguists have to construct a more restricted language (than natural language) in which syntax and lexicon are controlled so complexity and ambiguity which could lead to misinterpretation and miscommunication are reduced. They would also have to construct norms and regulations as to how this controlled language is being put to use in different contexts.

According to the International Civil Aviation Organization (ICAO[4]), between 1976 and 2000, more than 1,100 passengers and crew lost their lives in accidents where language issues played a contributory role (Mathews, 2004). Aviation accidents almost always have several contributory factors. Linguistic issues which play contributory roles in accidents are relatively poorly known or focused on in accident reports, which is why there is a need for linguists and aviation specialists to keep working together to make the use of language in the cockpit as intuitive as possible to avoid dangerous situations.

2.3 Controlled Languages at Airbus

Currently there are several controlled languages at Airbus that were put in place and tested to achieve unambiguous comprehension (avoid **ambiguity** (multiple interpretation), **inaccuracy** (inexact interpretation), **inconsistency** (non-standardized incoherent terminology), and **inadequacy** (incorrect term employment in a specific context)) in order to ensure the safety of the navigation, operational needs, and the adaptability of the human-computer interaction to different situations in the cockpit, cabin, and maintenance:

- Cockpit Controlled Language (also known as Airbus Warning Language) that is used for ECAM[5] PFD[6] MFD[7] monitors
- GOLD[8] that is used for OIS[9] for both flight communications and cabin communications

4. ICAO. [online] Available at: https://www.icao.int/Pages/default.aspx [Accessed 10 Dec. 2018].
5. ECAM: Electronic Centralized Aircraft Monitor: Monitors Aircraft Functions, details failures and provides sequential procedures for pilots to deal with them.
6. PFD: Primary Flight Display: Displays primary flight data such as altitude, airspeed, vertical speed etc.
7. MFD: Multifunction Display: Monitor that can display multiple pages such as navigation routes, maps and weather.
8. GOLD: Guidelines For Operational Language in Documents
9. OIS: On Board Information System

- ASD-STE[10] (Previously known as AECMA-SE[11] or AECMA for short) for aircraft ground maintenance

From the first flight of the Airbus A340 plane in 1991 to the introduction of the Airbus A380 in 2004, there was a significant process of simplification and standardization to include new cockpit controlled language rules (cf. Section 3.2). However, this controlled language has several limitations mostly due to:

- Small screen sizes (limited number of words and sentences)
- Highly codified nature (non-conforming to natural language syntax, highly abbreviated, typographically variable, color-coded and so on), thus it requires prior pilot training in order to achieve fluency
- "Family concept" and standardization in the Airbus fleet so no substantial changes in the interfaces would be made between two different generations of aircraft even if new technology allows it.

Figure 3 is an example of different messages found at different locations in one of the corpora at hand (this is not an *exact*[12] replica of an alarm).

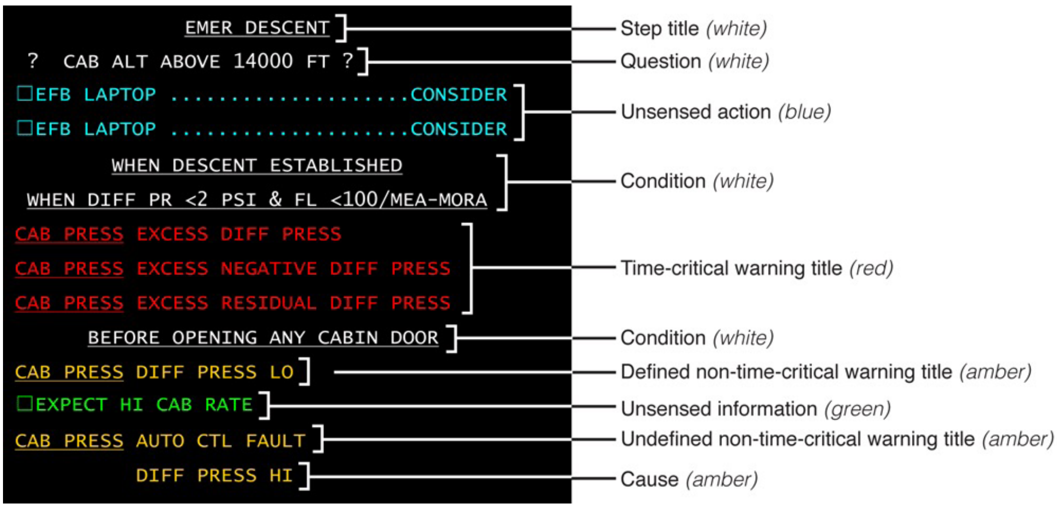


Figure 3. Example of Different Types of Messages in Cockpit CL

10. ASD-STE: Aerospace and Defence Industries Association of Europe-Simplified Technical English.
11. AECMA-SE: Association Européenne des Constructeurs de Matériel Aérospatial, Simplified English.
12. For confidentiality reasons, complete Airbus alarms cannot be released. The lines in Figure 3 are assembled from different alarms, and they are representative of the various types of information in the corpora.

2.4 Subject Introduction

However, as we are now addressing a disruptive cockpit design for future airplane generations, we might be looking at different flexibility margins: less limitations, bigger screen sizes, less coding etc.

Additionally, *"for over a decade, the international aviation community has been considering the concept of Single Pilot Operations (SPO) as a viable solution to the rising costs associated with commercial air transport. Recent advances in Communications, Navigation, Surveillance/Air Traffic Management and Avionics (CNS+A) technologies have allowed higher levels of automation, creating an opportunity for commercial airliners to transit to SPO."* (Lim et al. 2017). Therefore, there is an even stronger need for an optimized, intuitive, and easy to use interfaces with as little linguistic ambiguities as possible (by reducing as much as possible any form of misinterpretation).

Consequently, in order to test and optimize comprehension, perception, and use of controlled languages in the cockpits, we seek to conduct behavioral experiments by taking advantage of new tools and research in the cognitive sciences and controlled languages domains and apply linguistic hypotheses.

We will be targeting three main aspects[13]:

- Faster comprehension
- More accurate comprehension
- Limited training needs

13. Not in any particular order

3

Literature Review

3.1 Controlled Languages

The first step to optimizing the current controlled language and improving comprehension is to delve into the controlled language literature and define the spectrum of our research.

In "A Survey and Classification of Controlled Natural Languages", Kuhn (2014) gives a comprehensive overview of the Controlled Natural Language (CNL) (or for short Controlled Language (CL), terms used interchangeably in this thesis) domain by establishing common terminology and providing a starting point for interested researchers.

We will start by discussing this paper as it is wide-ranging and covers different aspects that form a comprehensive theoretical background to this research. Other research and commentary will also be included throughout the discussion when relevant.

Firstly, the author starts by enumerating the many different ways these constructed languages are referred to, such as, but is not limited to:

- Controlled
- Constrained
- Processable
- Simplified

- Technical
- Structured
- Basic

As the variety of attributes suggests, there is no general agreement on the characteristic properties of controlled languages, making CNL a very "fuzzy term". The author suggests that there are two main reasons for that:

1. *"CNL approaches emerged in different environments (industry, academia, and government), in different disciplines (computer science, philosophy, linguistics, and engineering), and over many decades (from the 1930s until today).*

   People from different backgrounds continue to use different names for the same kind of language".

2. "*Although CNLs seem to share important properties, they also exhibit a very wide variety:*
   - *Some are inherently ambiguous, others are as precise as formal logic*
   - *Everything can be expressed in some, only very little in others*
   - *Some look perfectly natural, others look more like programming languages*
   - *Some are defined by few grammar rules, others are so complex that no complete grammar exists*

   This variety makes it difficult to get a clear picture of the fundamental properties."

There are various definitions of controlled languages mostly depending on usage, such as

- Kittredge (2003): *"A controlled language (CL) is a restricted version of a natural language which has been engineered to meet a special purpose, most often that of writing technical documentation for non-native speakers of the document language.*

  A typical CL uses a well-defined subset of a language's grammar and lexicon, but adds the terminology needed in a technical domain."

- and Fuchs & Schwitter (1995): *"Controlled natural language is a subset of natural language that can be accurately and efficiently processed*

> *by a computer, but is expressive enough to allow natural usage by non-specialists."*

Kuhn (2014) argues that both of these definitions show a bias towards a particular type of CNL such as human oriented ones for comprehensibility (Kittredge, 2003) or machine-oriented ones for computer interpretation (Fuchs & Schwitter, 1995). But both definitions agree that a CNL is based on a certain natural language but is more restrictive, and they agree that CNLs are constructed languages (engineered, not naturally occurring linguistic phenomena). He points out that the term "subset" in its mathematical sense is too restrictive to cover a large part of CNLs because many of these languages exhibit small deviations from natural grammar and semantics and some make use of unnatural elements such as colors and parentheses to help comprehensibility.

Kuhn's (2014.) proposed definition is more inclusive to different kinds of what is generally referred to as controlled language:

"A language is called a controlled natural language if and only if it has all of the following four properties:

1. *It is **based** on exactly **one** natural language (its "base language").*
2. *The **most important difference** between it and its base language (but not necessarily the only one) is that it is **more restrictive** concerning lexicon, syntax, and/or semantics.*
3. *It **preserves** most of the natural **properties** of its base language, so that speakers of the base language can **intuitively and correctly understand texts** in the controlled natural language, at least to a substantial degree.*
4. *It is a **constructed language,** which means that it is explicitly and consciously defined, and **is not the product of an implicit and natural process** (even though it is based on a natural language that is the product of an implicit and natural process)."*

Kuhn (2014) also offers a short version of his definition, *"A controlled natural language is a **constructed language** that is based on* a *certain **natural language,** being more **restrictive** concerning lexicon, syntax, and/or semantics, while preserving most of its **natural properties.**"*

This definition (both in its long and short versions) is not as restrictive or targeted to one domain as previous given ones, but it encompasses virtually all languages identified as a form of controlled language, it excludes natural language (since it is not constructed), it excludes constructed languages that have several base languages such as Esperanto, and it excludes common formal languages because they lack intuitiveness and comprehensibility. However, KUHN (2014) does not invoke the notion of limiting ambiguity but hints to it by saying that the language is more restrictive concerning lexicon, syntax, and semantics.

We therefore propose the following diagram (cf. Figure 4) as a visual representation of KUHN's general short definition and SCHWITTER's (2010) definition (which does evoke ambiguity): *"CNLs are engineered subsets of natural languages whose grammar and vocabulary have been restricted in a systematic way in order to reduce both ambiguity and complexity of full natural languages"*

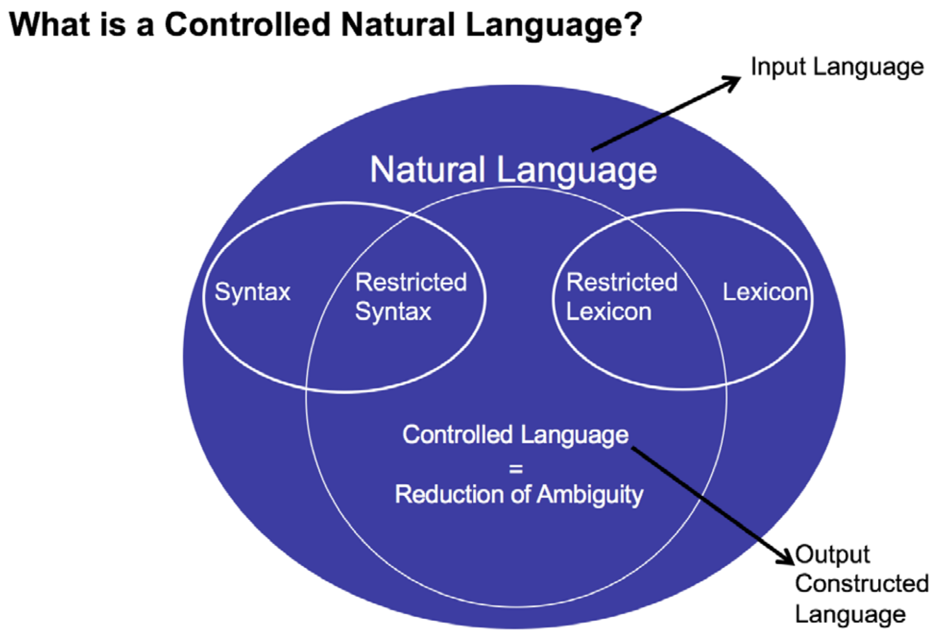


Figure 4. Visual Representation of Controlled Language Definition

We can observe a specific natural language that is the superset and input language from which a subset is consciously constructed. This subset is the controlled language, or the output constructed language, that has been restricted and reduced in syntax and lexicon to limit and manage ambiguity. As we can see, the restricted lexicon and syntax are part of the larger superset. Clearly the mathematical representation here, of the subset relation to the superset, does not accurately convey how most CNLs relate to their base language, as there can be small deviations from the base syntax and lexicon, as well as additions of semiotic and extra-linguistic elements.

3.1.1 Related terms

In this section, we would like to discuss some related terms that could be easily confused with (or that sometimes incorporate/are part of) controlled languages.

- **Natural language** and **Instances of Natural Language** (INL)

  Natural language could be (perhaps somewhat simplistically) defined as a non-constructed uncontrolled language whose definition is only obtained by contrast to constructed or controlled languages. In other words, the absence of complete control and deliberate construction is what makes up natural language.

  All standard/established languages are mere instances of depiction of natural language in the sense that language is alive and in constant evolution, and thus natural language is a virtually unattainable theoretical concept that is a product of an implicit process.

  Therefore, in our sense established languages like English, French, or Spanish (etc.) are what we would like to call mere **Instances of Natural Language** (**INL**), as they could be officially regulated as well as emerge naturally as a result of spontaneous regularities.

  Officially regulated languages such as French (by the "Académie Française") could only be classified as INLs because they have prescriptive rules, but not enough to be constructed from scratch (since language constantly evolves in a larger context and with diverse peoples and history) nor controlled enough to be a CNL (no limitations on vocabulary whatsoever as ambiguity is not a central concern, as it is with CNLs).

  Natural language's very existence is a paradox. It is an ever-changing infinite living machine that is hard to concretely describe, but it is also meant to be the be-all and end-all of officially regulated languages or INLs. It lies somewhere between the Saussurean Langage on the one hand, and Langue and Parole on the other hand (Saussure, 1995). It is not only the official spoken language or the individual speech act, but also the constantly changing universal system of fundamental communication. Valette (2006) summarizes Gustave Guillaume's postulates concerning language as a "pre-science" or "avant-science": *« (i) la langue est une théorisation naturelle de la pensée ; (ii) le langage est une avant-science.* » In this sense, what we call natural language is a theoretical continuum of thought and linguistic knowledge.

Two units originate from natural language; on one side deliberately **constructed languages**, and on the other, languages that emerge naturally, which we will now call **spontaneous languages**. One should bear in mind that, as mentioned before, natural language is the source of both of these processes (spontaneous processes and prescriptive and proscriptive ones (in cases of INLs)), which basically constitutes the chicken and egg paradigm.

- **Constructed languages** (or artificial/planned) are languages that are consciously manipulated and defined to suit a specific purpose. A CNL is an engineered language and therefore is a constructed language since it does not emerge spontaneously. A programming language is another example, and so is Esperanto (literally "one hopes"), a widely spoken constructed language based on several base languages designed in the late 19^{th} century by Zamenhof in order to create a universal language that is easy to use to foster peace and communication. The difference between a CNL and Esperanto is that a CNL must be based on one and only one base language, which is not the case for Esperanto.
- **Controlled vocabularies** are standardized collections of names and expressions, *"lists of controlled terms, synonym rings, taxonomies, and thesauri"* (NISO, 2005). They are a part of almost every established controlled natural language. It is a terminological exercise consisting of a list of allowed vocabulary to be used in conjunction with syntactical rules (that are developed separately) to form a coherent new restricted language, which aims to avoid lexical ambiguities such as instances of polysemy, homophony, or morphological ambiguities. Controlled vocabularies are closely related to the domain's expert population's day to day use of the language and technical environment.
- **Controlled syntactic rules** are sets of rules and norms that are inspired from the base language but are restricted to reduce complexity and avoid structural syntactic ambiguity.
- **Style guides** could contain instructions/hints/advice on how to use an existing natural language (an INL) or in some cases offer prescriptive

guidelines that restrict the initial language. If a style guide merely describes good practices that have emerged spontaneously then it is not a CNL, if, on the contrary, it describes a new more restricted language then it is considered a CNL. Although a lot of CNLs use examples instead of clear rules, which gives a special status to the given example, and makes it a prototype that needs to be emulated (ICAO phraseology for instance, or PLAIN Language Guidelines).

- **Spontaneous Languages** are languages which emerge implicitly and naturally
- **Sublanguages** are languages that **naturally occur** when *"a community of speakers (i.e., 'experts') shares some **specialized knowledge** about a restricted semantic domain [and] the experts communicate about the restricted domain in a recurrent situation, or set of highly similar situations"* (Kittredge, 2003). The notion of sublanguage was first introduced by Harris (1968) from a mathematical point of view, in order to explore the possibility of reaching the core meaning of a corpus by using linguistic transformations. As we can see, the difference between a CNL and a sublanguage is that CNLs are intentionally constructed and defined by a group of experts/linguists. Sublanguages emerge naturally from one base natural language.
- Bhatia (1993) defines **Textual Genre** as *"a recognizable communicative event characterized by a set of communicative purpose(s) identified and mutually understood by the members of the professional or academic community in which it regularly occurs"*. The terms *sublanguage* and *textual genre* are often used interchangeably. For example, Somers (1998) defines *sublanguage* as *"an identifiable genre or text-type in a given subject field, with a relatively or even absolutely closed set of syntactic structures and vocabulary"*.

  Biber (1988) also defined genre similarly to Bhatia (1993) in the sense that he also mentioned that it is predominantly based on use rather than form. *"Genre categories are determined on the basis of external criteria relating to the speaker's purpose and topic; they are assigned on the basis of use rather than on the basis of form."*

  Lee (2002) highlights the difference between external extra-linguistic criteria and the internal linguistic ones: *"A genre, in this view, is defined as a category assigned on the basis of **external** criteria such as intended audience, purpose, and activity type, that is, it refers to a conventional,* *culturally recognised grouping of texts based on properties other than lexical or grammatical (co-)occurrence features, which are, instead, the* ***internal*** *(linguistic) criteria forming the basis of text type categories."*

- **Fragments of language** denote *"a* ***collection of sentences*** *forming a naturally delineated subset of [a natural] language"* (Pratt-Hartmann & Third, 2006). They are identified rather than defined. As with a sublanguage's spontaneous process, they are closely related to the initial natural language and related terms. The purpose of fragments of language is to theoretically analyse them rather than put them to use in an official CNL aimed for a specific purpose. "*A CNL can be seen as a fragment of a language developed for the purpose of supporting some technical activity"* (Pratt-Hartmann, 2009).

  Sublanguages are also meant to be analysed and not put to use but are excellent starting points for future CNLs in a specific domain. The difference between sublanguages and fragments of language seems to be that the latter consists of a limited collection of entire sentences and phrases rather than a set of syntactical rules and targeted lexicon which could emerge in a sublanguage, and which could be used in a number of different ways and sentences. Fragments of language could be seen as a subset of a sublanguage.

- **Phraseology** denotes a *"set of expressions used by a particular person or group"* (Mifflin, 2000). Contrary to sublanguages and fragments of language, a phraseology is not a selection of sentences but a **selection of *phrases***. Phraseologies when spontaneous are used to alleviate the dense grammatical structure than full natural language sentences, and when constructed (could be considered CNLs in this case) they have the role of reducing ambiguity as well as alleviating complex syntax. Phraseologies could also be considered subsets of sublanguages because they draw on spontaneous regularities of a certain expert population in specific domains, as well as subsets of CNLs because they contain restrictive rules that govern the spontaneous predictabilities.

It is important to differentiate between these related terms since they allow us to better understand generalities about CNLs, how they emerge, and the theory behind them. We will now discuss linguistic norms and the processes of norming and normalization without which CNLs would not exist.

3.1.2 Norms, Norming, and Normalization

Condamines et al. (2017) note that spontaneous linguistic regularities occur at the workplace whenever humans are involved in a common task. *"These regularities also constitute norms since all the speakers involved in a common task have to use them in order to be accepted as a member of the speech community (Hymes, 1972). But unlike Controlled Natural Languages, these norms are not consciously prescribed."*

The fact that regularities appear spontaneously within speech communities leads us to conclude that commonality in language use (or norms) is not a function of an individualistic choice or style, but rather an effect of group interactions. Gledhill (2000) mentions that *"the regularity and pervasive nature of collocation appears to be incompatible with the intuition that an individual's use of language is inherently unique and creative, [...] mastering phraseology is one of the proofs of belonging to a discourse community; in particular, it is one of the proofs of belonging to a scientific community."* And interestingly, observing regularities, or in this case collocations, is what allows us to define discourse communities: *"Collocations appear to confirm the existence of a discourse community. Their very consistent nature suggests that collocations have a central role to play in discourse."* (Gledhill, 2000)

Ryan (2018) notes that *"natural languages have evolved spontaneously to deal with all situations of human communication. However, they can be deliberately controlled to varying extents, in different ways and for ranging purposes by additional prescriptive rules."* And since CNLs are based on this spontaneous evolution in INLs, they inevitably should/do carry those spontaneous regularities in their simplified form.

Furthermore, the French linguistic School of Rouen employs two terms "normaison" and "normalization" (Gaudin, 1993). "Normaison" or norming describes the spontaneous process of linguistic norm creation that occurs within a language activity in a community of experts, often at the workplace. Normalization is a conscious officially prescribed process of norm creation that is based on the spontaneous linguistic regularities that emerge in the norming process:

Elles [= les terminologies] connaissent des réalisations, notamment à l'oral, qui sont le lieu de créations ; ce sont les usages professionnels du laboratoire, de l'atelier, de l'usine qui les suscitent (cf. Candel, 1993). La genèse de ces formes a lieu *dans les pratiques langagières et leur stabilisation leur confère le statut de normes de discours permettant l'intercompréhension. Et c'est ici que l'on peut distinguer deux types de procès aboutissant à la construction d'une norme, la normalisation et la normaison : « En résumé, on peut dire que la normalisation, c'est le processus qui vise à la construction consciente d'une norme unifiée, et la normaison, le processus responsable de la logique même de tout système linguistique » (Guespin, 1993: 218). Qu'elle soit nationale ou internationale, politique ou technique, la normalisation émane toujours d'une institution qui fixe les termes recommandés ou obligatoires. En revanche, la normaison relève de ce que Teresa Cabré décrit comme « un processus au moyen duquel un système terminologique déterminé s'autorégule en accord avec ses utilisateurs »* (Gaudin, 2003)

Controlled languages must therefore systematically rely on existing sublanguages that emerge from spontaneous linguistic regularities within a speech community of domain experts. These controlled languages would be more likely used by the target users since they reflect the day to day domain-specific lingo. These regularities within the speech community also reflect a certain level of understanding between the writer and reader, or the linguist and the engineer. This level of entente may translate into a very dense communication system that is closed off to outside circles, and which could very well end up being unmanageable and extremely complex, and thus this communication would benefit from the process of normalization which would help manage the spontaneous complexity. *"It is assumed that author and reader share a common language and that when certain words or phrases are used, each understands what is meant [...]. Writer and reader, or speaker and hearer, are assumed to have the same level of expertise [...]. This expert to expert communication context is likely to be the one with the highest density of terms."* (Pearson, 1998)

Furthermore, communication among experts is subject to change, thus linguistic regularities are also likely to change, and since it is impossible to account for all the changes in a given situation, the use of language, even if controlled, may always generate risks of misunderstanding (Condamines, 2008; 2010). A good CNL should therefore define efficient linguistic norms based on the domain's spontaneous regularities while taking into account the fact that language contexts are in constant evolution.

Humbley (2001) states that a language should always consider the constant changes it undergoes and that any normalization must absolutely rely on the norming process and the existing social regularities in order to succeed.

« *La normaison est le processus qui conduit à ce qu'une langue et, pour ce qui nous occupe, les vocabulaires, sont en état d'équilibre et de renouvellement permanent du fait de la multitude des usages qui traversent la langue. Il s'agit d'un processus spontané et collectif. Spontané (ce que n'est pas la normalisation) et collectif, ce que la normalisation n'est pas non plus. Et à mon sens, [...] toute normalisation qui ne s'appuie pas sur la normaison, est une entreprise brutale et presque entièrement vouée à l'échec. Il convient en effet de s'appuyer sur l'existant pour décider, mais encore faut-il avoir un tableau de l'existant.* » (Depecker, 1996)

Lopez (2013) proposed a visual diagram (cf. Figure 5) that shows the expected influence of spontaneous regularities on prescribed norms as a function of time and language evolution. The circle of norming and normalization evolve in the same cyclic manner with the passing of time and language evolution.

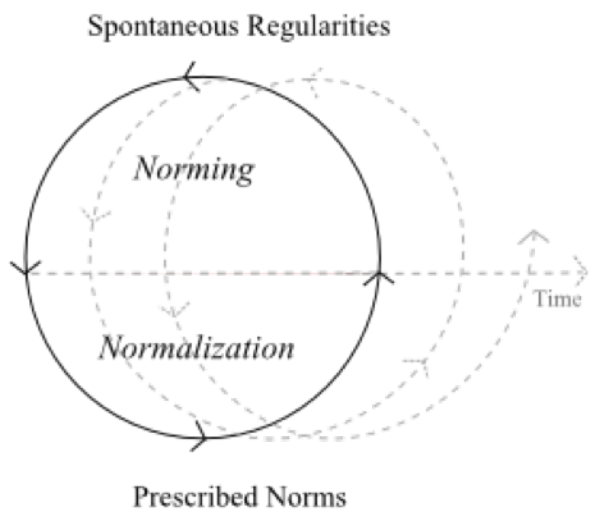


Figure 5. Expected Influence of Spontaneous Regularities on Prescribed Norms and Vice Versa (Condamines et al., 2017)

The notion of norming goes hand in hand with the notion of acceptability. Acceptability is defined in many ways by different authors, but a general definition is that of Schade and Schlag (2003) as *"the prospective judgment of measures to be introduced in the future"*.

A controlled language produced by the process of normalization and extracted from spontaneous regularities of the norming process must be acceptable enough to be used by the target audience. If the controlled language is not based on real life usage, then it would be difficult for users to accept it or easily make use of it:

La réalisation d'une politique linguistique nécessite des choix d'aménagement *linguistique qui ne peuvent rencontrer le succès qu'à la condition d'être compatibles avec les opinions des locuteurs et donc avec les sentiments et les pratiques linguistiques. C'est pourquoi la négociation terminologique, l'information et la consultation des acteurs concernés sont des facteurs favorables pour que les décisions soient suivies d'effets. Il faut pour cela que les conditions d'une adhésion sociale soient réunies et, donc, que les décisions prises reposent sur une description fine des pratiques et une consultation préalable des usagers. [...] En fait, on peut affirmer que la* ***normaison*** *devrait être la priorité des organismes de politique linguistique.* (Gaudin, 2005)

We would like to propose a diagram (cf. Figure 6) that summarizes all these related terms and the relationships that binds and connects them.

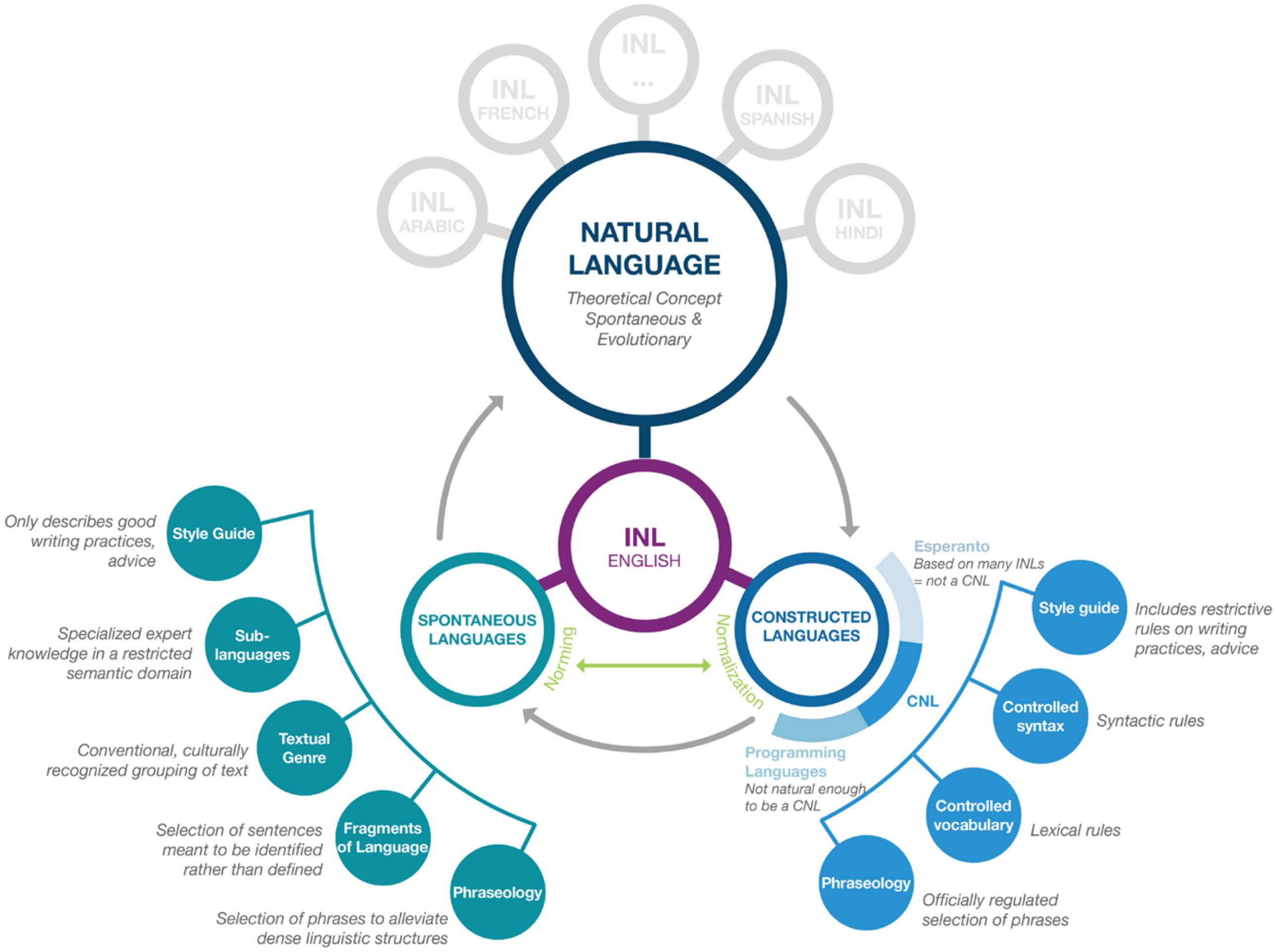


Figure 6. Relationship Between Spontaneous and Constructed Languages, and Related Terms

As we can see, natural language as a theoretical concept, spontaneous and evolutionary, is at the base of different established INLs such as English, French, etc. The diagram then makes the distinction between spontaneous and constructed languages. Spontaneous ones (in teal) are issued from the norming process (in light green) that occurs implicitly from linguistic norms and spontaneous regularities. Constructed ones (in blue) are explicitly regulated, often for specific purposes, which gives birth to the normalization process (in light green) that is inspired from the norming process of spontaneous languages. Therefore, as we can see the two processes are interconnected.

Additionally, as the cyclical diagram shows (upper part), spontaneous languages which are themselves issued from INLs (in this case English), which are in turn issued from natural language, circle back up to natural language, as the theoretical concept also evolves in time and usage of language. This language evolution trickles back down to the INLs and the constructed and spontaneous languages to form a never-ending cycle of language evolution in time.

On the left side of the diagram (in teal), we can see the different spontaneous languages or concepts evoked in this section with a brief description for each: Style guide – sublanguage – textual genre – fragments of language – phraseology.

On the right side of the diagram (in blue), we can see that constructed languages are not necessarily only CNLs. Languages like Esperanto are constructed languages but are based on several different INLs, therefore it is not considered a CNL even if it could have traditionally CNL goals like simplification and easier comprehension among humans. Programming languages and highly codified and formalistic languages are not natural enough to be considered a CNL since they are not understood intuitively by native speakers of the language.

The CNL fork of the constructed language shows the different types of concepts considered CNLs which were evoked in this section with a brief description for each: Style guide – Phraseology – controlled syntax – controlled vocabulary. The diagram does not give examples of specific CNLs, but what constitutes them, such as controlled vocabulary and syntax. It also shows style guide and phraseology in both spontaneous and constructed languages since it depends on whether the style guide's good writing practices are officially regulated or merely describing and giving advice; and it depends on whether the selection of phrases in phraseologies are officially regulated for specific purposes or not. It is clear that the two concepts are closely related: Officially regulated constructed languages are mostly always based on their spontaneous counterpart and the regularities that naturally occur in the latter.

3.1.3 Types and Properties

3.1.3.1 HOCL and MOCL

The next step into understanding controlled languages is looking into the different types and properties. Huijsen (1998) introduced the distinction between "human-oriented" controlled languages (HOCL) and "computer-oriented" or "machine-oriented" controlled languages (MOCL). This describes the function of the language and what it aims to resolve rather than a description of the language itself. Historically, the first controlled languages were human-oriented as they were designed to facilitate international communication and reflect a growing gravitation towards globalization. Computer-oriented languages appeared later on with the advent of new technologies and the increasing need for fast and automatic translation, and formal language notations.

3.1.3.2 CNL Types C, T, F

Schwitter (2002) categorized controlled languages in 3 types:

1. Type C: to improve communication or comprehension among people, especially speakers with different native languages. They are often used to reduce ambiguity and complexity in the base language.
2. Type T: to improve translation (manual, semi-automatic, or automatic translation).
3. Type F: to provide a natural and intuitive representation for formal notations.

3.1.3.3 Restrictive or General

Pool (2006) aimed to evaluate the costs, benefits, and feasibility of a controlled-language for semantic web that could be used by both humans and machines while avoiding structural and semantic ambiguities. In order to do that, he conducted evaluations in which he took a few well documented controlled languages that he split into restrictive and general projects:

- Restrictive languages which *"overtly or apparently aim for expressivity in a domain (e.g., truck repair) and/or genre (e.g., instructions) and do not specify how to extend this expressivity."*
- General languages aim to target multiple domains and genres. Adriaens and Schreurs (1992) associate this distinction to open or closed lexicon within languages which makes them restrictive or reusable by the general public.

For his evaluations, Pool (2006) chose ambiguous sentences aiming to discover limits in precision. He then "translated" sentences into their controlled language equivalent following instructions from the documentation. He concluded that CNLs *"that have been reported as successes have been mainly restrictive: designed for limited, intra-organization or intra-industry purposes. That they cover single domains and genres, with repetitive and trainable authors, facilitates their efficacy."*

Additionally, he remarks that formalistic languages (Type F) exhibit high precision but limited expressivity, while naturalistic ones (Type C and potentially T) are highly expressive but not as precise so as to leave room for ambiguity. *"The two strategies might converge, but no project has bridged the gap yet, and it remains unknown how a controlled natural language can achieve precise, yet broadly expressive, meaning representation."*

3.1.3.4 Proscriptive and Prescriptive

Furthermore, CNLs could be based on prescriptive or proscriptive rules. Ones that are based on only proscriptive rules (ones that describe what is not allowed) must be based on an original INL, and ones that are based on only prescriptive rules (what is allowed) could possibly start from scratch such as formal logic languages (e.g. every A is a B), but even then prescriptivists rely to a certain extent on previously acquired language competencies. Nonetheless, it is not uncommon to find controlled languages that have both prescriptive and proscriptive rules, and which are based on an INL.

3.1.3.5 Properties and PENS Classification

Wyner et al. (2009) identified about 40 properties of controlled languages and their environments. However, these properties are not very well-defined and they often intertwine, so it is hard to extract clear cut categories. Other authors discussed various controlled language properties separately, but Kuhn (2014) argues that it is difficult to merge these different properties (that are often based on each author's original intent and domain specific usage of CNLs) to be able to categorize existing languages properly. Therefore, Kuhn developed the PENS (precision, expressiveness, naturalness, simplicity) classification scheme that condenses fundamental properties that are mostly independent of one another. In addition to the main uses of a language, he points out that most languages originate from a certain domain such as academia, industry, or from government agencies, they could be either spoken or written, and may have different levels of maturity and use. PENS is designed to measure the nature of a language, not its quality or usefulness. It only describes languages but does not rank them. As the languages are domain specific, the perfect language does not exist and compromises have to be made for different uses. Finally, different weights could be assigned to the different dimensions depending on the needs of each language.

Kuhn (2014) created a letter code for the different properties that we can observe in Figure 7:

Letter codes for properties of CNLs.

Code	Property
C	The goal is comprehensibility
T	The goal is translation
F	The goal is formal representation (including automatic execution)
W	The language is intended to be written
S	The language is intended to be spoken
D	The language is designed for a specific narrow domain
A	The language originated from academia
I	The language originated from industry
G	The language originated from a government

Figure 7. Letter Codes for CNL Properties (taken from Kuhn, 2014)

The dimensions delineated for the PENS classification scheme are:

1. **Precision**

   Clarity from NL (Natural Language, many interpretations) to formal logic (maximal precision), P1 to P5.

2. **Expressiveness**

   Range of propositions that a certain language is able to express, ranging from no quantifications or arity to being able to express everything, E1 to E5.

3. **Naturalness**

   How close the language is to natural language, readability and understandability to speakers of a given language, ranging from unnatural languages (heavy use of symbols, brackets) N1 to very natural (NL) N5. No CNLs exist in N1 and N2

4. **Simplicity**

   Effort needed to implement the syntax and semantics in a mathematical modal (such as computer programs), verified by the number of pages, from virtually indescribable (NL) to described in one page. S1 and S2 are proscriptive (relying on NL), and S3 till S5 are prescriptive therefore simpler, because defined from scratch, and not relying heavily on NL.

Figure 8 is an excerpt example of some controlled languages classified according to the PENS scheme (the rest of the table could be found in the Annex A).

Observed PENS classes and properties of CNLs (sorted by PENS class).

class	properties	languages
$P^1E^5N^5S^1$	C T W I	IBM's EasyEnglish
	C W S G	Special English
	C W A	E-Prime
	C W G	Plain Language
$P^2E^1N^3S^2$	C S D G	CAA Phraseology, FAA Phraseology, ICAO Phraseology, PoliceSpeak, SEASPEAK
$P^2E^1N^3S^3$	C W D I	Airbus Warning Language
$P^2E^5N^4S^1$	F W A	AIDA
$P^2E^5N^5S^1$	C T W D A I	ALCOGRAM, COGRAM
	C T W D A	CLCM
	C T W D I	ASD-STE, Avaya CE, Bull GE, CTE, CASL, CE at Douglas, DCE, General Motors GE, PACE, Sun Proof
	C T W D	Wycliffe Associates' EasyEnglish
	C T W I	iCE, SMART Controlled English
	C W D I	AECMA-SE, CFE, CASE, CE at Clark, CE at IBM, CE at Rockwell, EE, HELP, ILSAM, KISL, NCR FE
	C W D G	Massachusetts Legislative Drafting Language
	C W I	Boeing Technical English, NSE, SMART Plain English
	C W	Basic English
	T W D I	MCE, Océ Controlled English
	T W A	KCE
	T W I	CLOUT
$P^3E^1N^4S^2$	C F W D I	SLANG
	F S D I	Voice Actions
$P^3E^2N^4S^3$	F W D A	RNLS
$P^3E^3N^3S^3$	F W A	ClearTalk

Figure 8. Observed PENS Classes and Properties (Kuhn, 2014)

3.1.3.6 Some Influential CNLs

If we take the example of Basic English (P2 E5 N5 S1, C W) which is the first reported instance of a controlled language (by current definitions), the C stands for comprehensibility oriented, and the W for written language, P2 stands for Precision 2, E5 for Expressiveness 5, N5 for Naturalness 5, and S1 for Simplicity 1.

It was presented in 1930 by Charles Ogden and it aimed to improve communication among people around the globe. It influenced Caterpillar Fundamental English, which became itself a very influential human oriented controlled language in industry. Only 18 verbs are supported: put, take, give, get, come, go, make, keep, let, do, be, seem, have, may, will, say, see, send. These verbs can be combined with prepositions to form more specific relations such as "put in" to express "insert". Many texts have been written using Basic English. Kuhn (2014) notes that *"the drastic simplifications on the lexical level together with the grammatical restrictions constitute a significant gain in precision compared to full English."* Note that Siddharthan (2003) defines text simplification as any process which reduces the syntactic or lexical complexity of a text while attempting to preserve its meaning and information content.

Flesch (1944) in his article "How Basic is Basic English?" claims that Basic English *"is neither basic nor English"* and starts off with an example *"If I were Mr. Churchill, I would not like being reduced to calling Hitler* "a very bad man", *or a bomber* "an air plane sending down hollow balls full of substance with a tendency to go off with a loud noise"", in reference to Basic English's arbitrarily selected 850-word vocabulary. He criticizes Ogden for *"deliberately avoid[ing] the scientific approach and not [being] lucky enough to find the key to simplicity by accident"*. According to Flesch (1944), linguists criticized Basic English in an issue of the Saturday review of Literature for being *"a kind of quack based on a faulty analysis of the language process."* Nonetheless, he concludes by saying that *"Basic English is the first attempt in the history of mankind to create a simplified language within a language [. . .] and that simplified English is bound to come [. . .] in a generation or two [...] and will be taken over by whatever system of simplified English we are going to adopt"*. Evidently, it is in fact the case.

Caterpillar Fundamental English (P2 E5 N5 S1, CWDI) is an influential controlled language which was one of the first languages to be designed for industrial use. It was introduced in 1971 and was based on Basic English. It was put in place to facilitate translation of Caterpillar machinery manuals across the world.

Another influential controlled language is the ASD-STE or ASD Simplified Technical English (P2 E5 N5 S1, CWDI), one of the most complete, widely used comprehension-oriented controlled languages (a language that has survived the test of time and is still used in the aircraft maintenance domain, and across different aircraft manufacturers). It was created in the 1980's and was previously named AECMA SE (European Association of Aerospace Industries – Simplified English). It was originally introduced to help translation and make aviation maintenance manuals' texts easier to understand by non-native speakers. It included lexical, syntactic and semantic rules. As Goyvaerts (1996) states about controlled languages destined for use in industrial settings: *"industry does not need Shakespeare or Chaucer, industry needs clear, concise communicative writing – in one word Controlled Language."*

Lastly, Plain Language or Plain English (PLAIN, 2011) (P1 E5 N5 S1, C W G) originated in the 1970's and was firstly initiated by the US government and other organizations. The main goal was to make official documents easier to understand. The studies on readability and the different readability formulae (cf. Section 3.3) were also introduced in the same decade by the US army. Plain language guidelines included rules such as "use pronouns to speak directly to readers" and "avoid double negatives" and "avoid exceptions to exceptions". Since 2010, US government agencies have been required to comply with the plain language rules. However, Kuhn (2014) states that *"with the focus being on human understandability and acceptance, documents in Plain Language do not seem to be considerably more precise or simpler from a computational point of view, when compared to full English."*

3.1.3.7 Controlled Languages, Evolution in Time

Kuhn (2014) provided a timeline of various English controlled languages. It includes bars that symbolize the "life" or usage of a given CL and, when possible, an approximation of its birth/death dates (when the CL stopped being used). The full timeline could be found in the Annex B, but we will provide an extract of the figure below (cf. Figure 9) that shows comprehension-oriented controlled languages, and particularly highlight the notable ones we mentioned here.

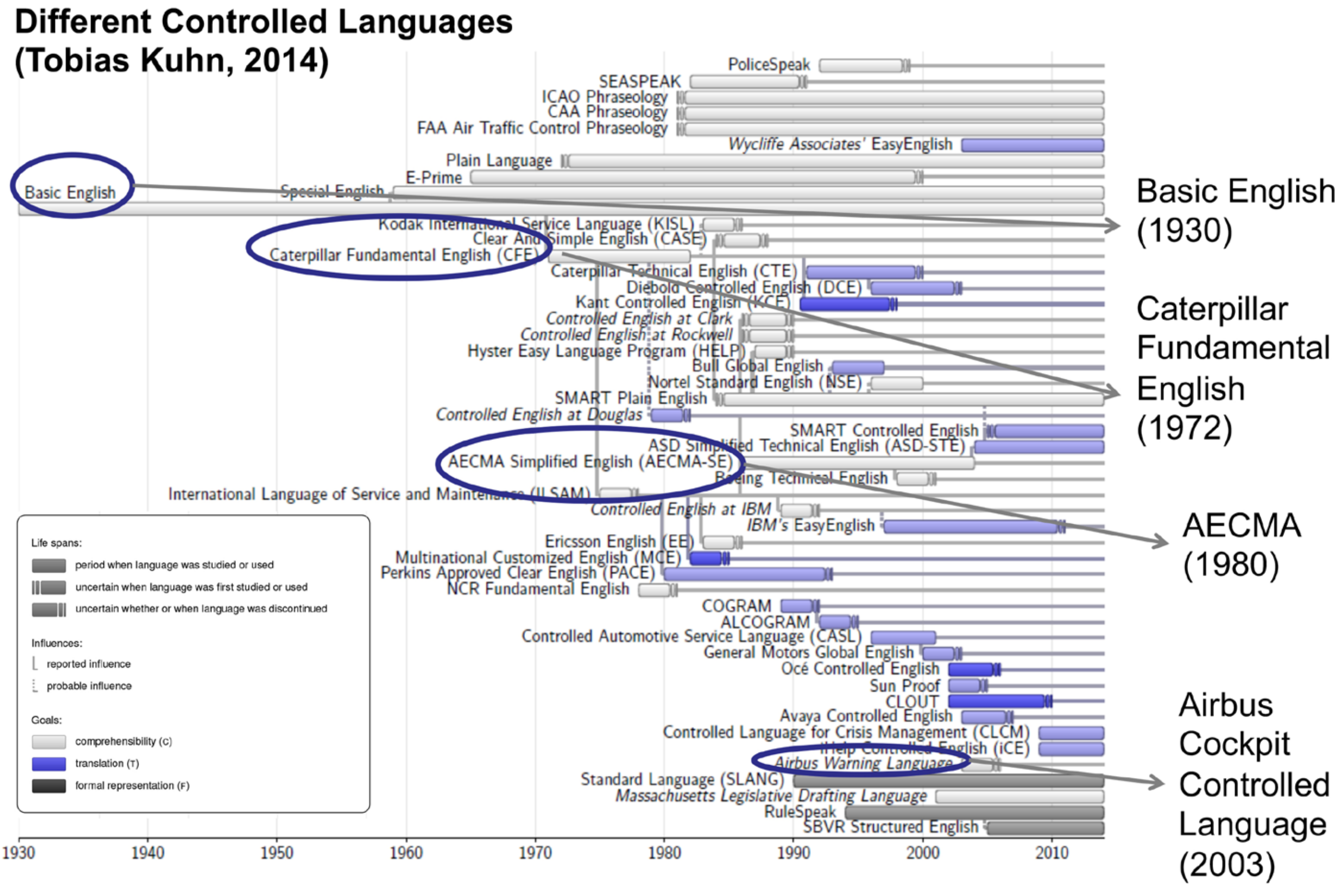


Figure 9. Annotated Sample (highlighting significant Controlled languages) of a timeline proposed by Kuhn (2014)

3.2 Airbus Cockpit Controlled Language

As mentioned earlier, there are several controlled languages that are put in place in Airbus. The one that will be of most interest to us in this research is the Airbus Cockpit Controlled Language (Spaggiari et al., 2003). It is the language used in the Airbus Cockpit on different monitors (such as ECAM (Electronic Centralized Aircraft Monitor)) and for physical displays (names of buttons for example). Before introducing this controlled language and the research that has been done on it, we will give a brief overview of the context in which it is used and some of the corpus' specifications.

3.2.1 ECAM

The ECAM is a system developed by Airbus that informs the crew about aircraft status. It shows detailed messages about possible function or equipment failure, and guides pilots with specific procedures to deal with them. According to Skybrary [14] the ECAM monitor shows:

- Primary engine indications, fuel parameters
- Warning and caution alerts, or memos
- Synoptic diagrams of aircraft systems, and status messages
- Permanent flight data

The ECAM is similar to the EICAS (Engine Indicating and Crew Alerting System) used in Boeing and Embraer aircraft. The EICAS simply displays flight information and failures without offering corrective action procedures or limitations to be undertaken by pilots, like the ECAM does. The ECAM makes use of color coding to guide pilots' actions with the aim of easing stress in abnormal or emergency situations. It replaces lengthy paper procedures that preceded its introduction in the cockpits. Aircraft sensors provide key parameters to monitor abnormal situations with the help of a flight warning computer.

The ECAM displays memos and advisory sections that show equipment or functions being used or diverse affected parameters. Additionally, there are three levels of failures ranging from least to most critical. The ECAM ranks and displays failures according to the urgency/criticality of the situation:

- Level 1 failures include cautions and faults. It could mean a loss of a function or equipment when it is not being used. It requires pilots to monitor the situation without any real risk present.
- Level 2 failures require the crew's attention and performance of actions (such as pressing on a button etc.).They have no direct consequence on flight safety in general (but ignoring a level 2 alert could lead to the appearance of a level 3 critical alert) and are accompanied by a brief sound or chime.
- Level 3 failures are time critical warnings that require immediate crew actions, and have a consequence on flight safety such as a loss of engine. These alarms are usually accompanied by aural alerts and continuous sounds and chimes to call more attention to the situation.

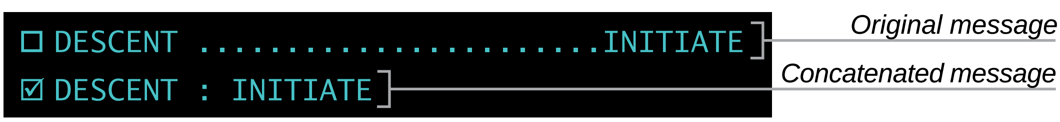
14. Skybrary Aviation Safety. [online] Available at: https://www.skybrary.aero/index.php/Main_Page [Accessed 27 Nov. 2018].

In the A350 cockpit (latest generation of cockpit design), 17 consecutive lines could be displayed. This is subject to change on different plane versions. Pilots could scroll up and down to view the different messages displayed, or to have an overview of the entire procedure.

There are sensed and non-sensed actions. Sensed actions are ones that when performed by the pilot (for example pushing a button on the overhead panel) are automatically dealt with by the aircraft. Non-sensed actions are ones where pilots need to tick a box, as one would a checklist, since the aircraft does not automatically detect that the action has been accomplished. This is similar to what private pilots have to do when they run checklists in non-commercial planes, as smaller aircraft do not have ECAM monitors, nor the capability of sensing information related to pilot actions.

For example:

Non-sensed action: (concatenation: check mark and colon in between words when action accomplished by pilot)



Sensed action: (concatenation: no check mark or colon when action accomplished by pilot)

3.2.2 Airbus Linguistic Corpus: Numbers and Color-coding

The Airbus A350 ECAM linguistic corpus contains about 18000 lines which constitute about 1500 different alarms. We will focus on this corpus as it is the latest generation.

The color coding of the messages themselves is as follows:

- **Red:** Titles of alarms that require immediate action. Time critical warnings. E.g.: Engine on fire.

- **Amber:** Titles of alarms that require crew awareness. Doesn't require immediate action since it is not time critical, but could potentially lead to a red warning. E.g.: Hydraulic High Temperature or Air Bleed Failure.
- **Green:** Information, advice
- **Blue:** Actions to be performed and limitations
- **White:**
  - Conditions: indented and underlined (When/Before/After /If)
  - Names of significant steps
  - Close ended questions that trigger different procedures depending on the "yes" or "no" reply of the pilot (58 different questions)
  - Choice answers of questions
- **Grey:** Dispatch message DA-item: could be dealt with after landing, mostly messages destined for maintenance technicians.

Figure 10 is an example of messages with the color-coding scheme. It is not an exact replica of the ECAM at a given time. The lines are assembled from different alarms, and they are representative of the various types of information in the corpus. Color coding is used to facilitate the identification of the different data types in the corpus. The colors are semiotic signs that aid text comprehension and need to be learned by users.

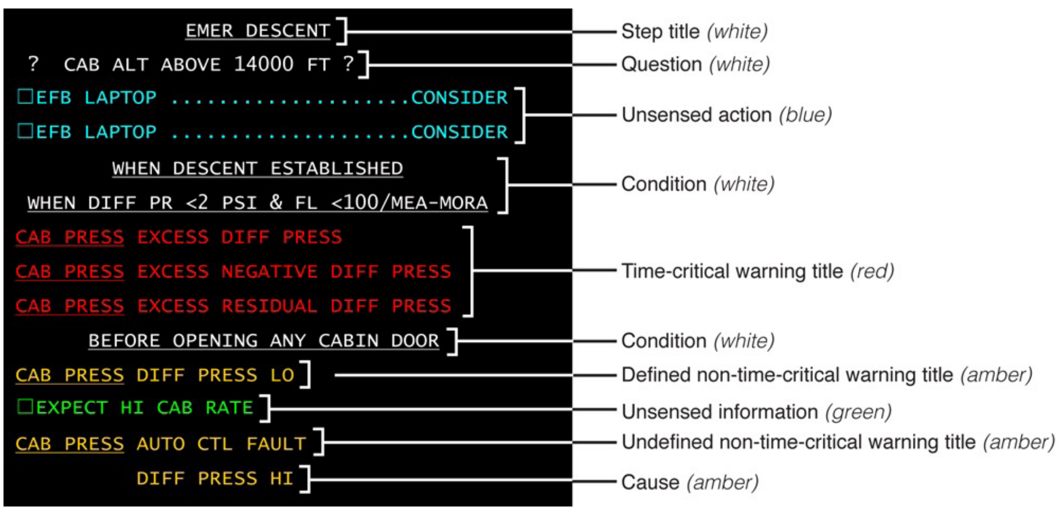


Figure 10. Text Example of different messages in cockpit corpus

The second example (cf. Figure 11) is taken from Skybrary[15]; it shows an real example of an A320 ECAM monitor.

Figure 11. Example of A320 ECAM monitor

3.2.3 Airbus Linguistic Corpus: Categorization of Data

We categorized the different types of data in the corpus. This eventually helped us put in place experiments that were adapted to the categories of data and their intended use. If the current controlled language is to be optimized, various methods of testing adequate alternatives need to be adapted to the information that is transmitted by each message. For instance, a message that instructs the pilot to perform an action does not have the intent, and consequently the desired reaction, as one that merely informs him or her.

Figure 12 is a diagram that shows the different types of data present in the ECAM cockpit controlled language corpus depending on the urgency of the situation (time critical or not) and if it requires the pilot to perform an action or not. It also follows the color-coding scheme of the corpus data. We should note that at this stage we are merely presenting the intent behind messages and not the CL linguistic rules that form them.

15. Skybrary Aviation Safety. Electronic Centralized Aircraft Monitor (ECAM) [online] Available at: https://www.skybrary.aero/index.php/Electronic_Centralized_Aircraft_Monitor_(ECAM) [Accessed 27 Nov. 2018].

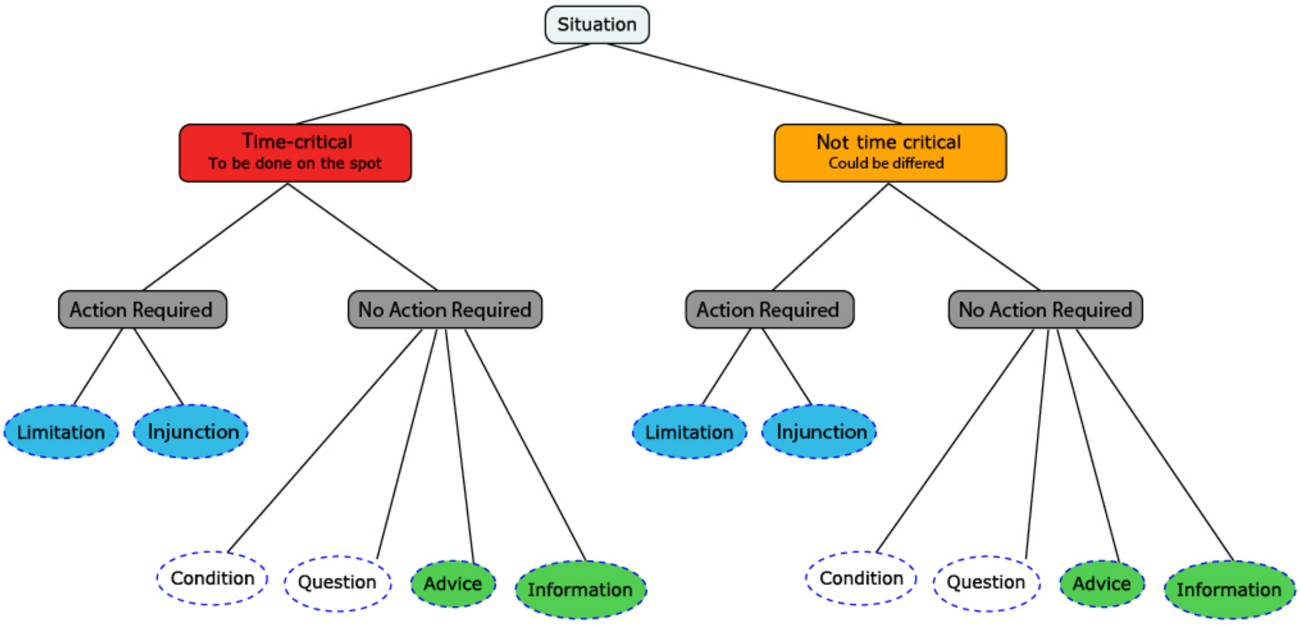


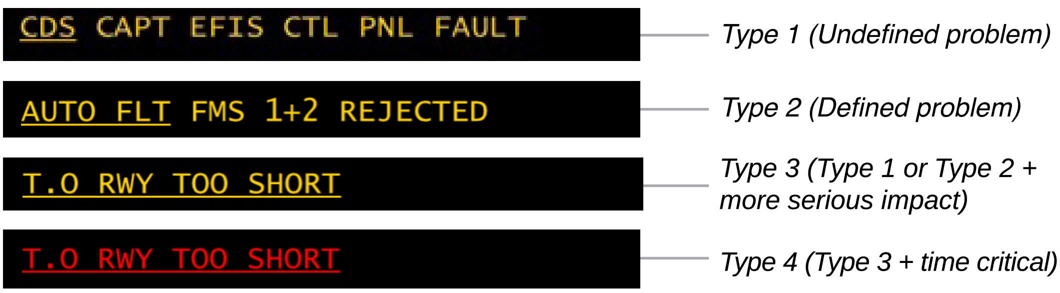
Figure 12. Different Types of Data in Cockpit Controlled Language in ECAM monitor

As we can see in the diagram, the finality of alarms could be summarized in 6 different categories (Limitations-Injunctions-Conditions-Questions-Advice-Information). Examples of each could be found in the ECAM Figure 11 (same color codes as the diagram).

A subsequent categorization of types of data was made. For instance, for the information category we can discern 16 different types (2 outside of alarms and 14 within):

I. Information outside of alarms (white, amber or red)

1. Information introducing steps: (see example in Figure 10)
2. Information in titles:



II. Information within alarms (green)

1. Information in tables:

ACFT CRZ FL	CABIN ALT TRGT
430	6000
400	5900
350	5000
300	4200
250	3300
200	2500
150	1600
100	800

2. Information announcing a risk, non-sensed (requires pilot validation):

☐ RISK OF REDUCED CABIN AIR FLOW
☐ RISK OF UNDUE STALL WARNING
☐ MINIMUM LIMITED TO RNP 0.30

3. Information announcing availability + time component (from T0 till further notice) + non-sensed (requires validation):

☐ LAV & GALLEYS EXTRACT AVAIL IN FLT

☐ NORM BRK AVAIL ON ALL WHEELS
☐ FMS PRED UNRELIABLE WITHOUT ACCURATE

☐ SLATS AVAIL
☐ FMS PRED UNRELIABLE WITHOUT ACCURATE

☐ AFS CTL PNL KNOB AVAIL FOR BUG SETTING
☐ GREEN DOT, S, F, VAPP AVAIL ON FMS

☐ FOR SYS PAGES : "ALL" AVAIL
   ☐ DU RECONF: CAPT OIS ON CTR P/B AVAIL
   ☐ DU RECONF : F/O OIS ON CTR P/B AVAIL
   ☐ DU RECONF : DISPLAY CYCLE P/B AVAIL
   ☐ DU RECONF: CAPT OIS ON CTR P/B AVAIL
   ☐ DU RECONF: CAPT OIS ON CTR P/B AVAIL

4. Information announcing an active function:

```
☐ BKUP GUIDANCE ACTIVE
☐ MINIMUM LIMITED TO RNP 0.30
```

5. Information about an upcoming possibly expected future event + action:

```
☐ EXPECT HI CAB RATE
```

6. Information + advice:

```
☐ TO VACATE RWY : TOWING ........CONSIDER
```

7. Situational information:

```
☐ IN DES : CAB ALT REGULATED TO 7000 FT
```

8. Limitation + consequences:

```
☐ BELOW 7000 FT : CAB ALT = ACFT ALT
☐ CAB ALT REGULATED TO 7000 FT
```

```
☐ ENG 1 BY GRVTY ONLY
☐ ENG 1 BY GRVTY ONLY
```

9. Condition + action on two lines:

```
☐ IF TCAS ALERT :
  AP & FD ............................OFF
  FLY MANUALLY TCAS RA ORDER
```

10. Information and list on second line:

```
   ☐ FMS DATA AVAIL :
             VLS, GREEN DOT, S, F, VAPP
```

11. Information conditioned by another:

```
☐ FMS PRED UNRELIABLE WITHOUT ACCURATE
            FMS FUEL PENALTY INSERTION
```

12. Simple information:

☐ ALT SINGLE SOURCE
☐ CAPT BARO REF : STD ONLY

13. Action in progress:

☐ CONF 1+F AFFECTED FOR T.O
☐ FLAPS ALIGNMENT IN PROGRESS

14. Recommendation:

☐ AUTOLAND RECOMMENDED
☐ FOR LDG : USE DIFF BRAKING AS RQRD

Other types of data (actions, limitations, conditions etc.) were also categorized in the same manner in order to get a clear sense of how the messages were constructed for each category.

After showing the manner in which some of the different data types are categorized in the corpus, we will present the linguistic research that has been done since the creation of the Airbus Cockpit Controlled Language, before we proceed to talk about controlled language evaluations in general.

3.2.4 Airbus Linguistic Corpus: Syntax, Terminology, and Abbreviation

According to Spaggiari et. al (2003), a cockpit controlled language project for warnings was launched in 1998. The goal was to enhance language quality for end users, to facilitate the cockpit designers' job while respecting the stringent safety criteria, and to create rules and standardize the language use in the cockpit.

The project was divided into three parts that dealt with terminology, syntax, and the use of acronyms. As some safety incidents occurred on commercial planes due to non-compliance with official procedures, it was concluded that misinterpretation of oral and written messages could be a contributing factor. Therefore, the obvious choice was to create standardized rules to form a coherent less complex CL that would help native speakers of English (international language of aviation) and non-native speakers alike. They started by doing an overview of the different controlled languages that existed at the time, along with the domain in which they were used and the rules they enforced.

As the CL was a form of writing guide, it included syntactical and lexical recommendations. It consisted in standardization rules such as:

- eliminating synonymy (the use of one word for one concept),
- standardizing syntactic structure for different data types,
- standardizing the use of ellipsis structure for injunctions only,
- improving abbreviations and acronyms to be intuitive and unambiguous.

It also took into consideration the possible different language interferences with English as the CL is to be used by pilots all over the world who have different native languages. The CL was designed in conjunction with linguists, domain experts, end users by means of interviews and assessments in order to consider their operational needs and experience.

On the terminological front, from the work done in Gavieiro-Villatte's PhD (2001) [16], the process consisted in morphological reduction and terminological standardization. The latter consisted in allowing terms already in use in the cockpits to stay in the new CL provided they fit into the new fixed criteria, and that these terms are not synonymous with other existing terms. A corpus analysis of 3000 sentences was performed and a decision tree was created with four branches:

1. A derivational and flexional branch that consisted of enforcing -ed and -ing morphemes for processes that have ended or are in progress respectively, keeping the plural -s morpheme, and replacing the negative prefixes such as "un" with "not".
2. A homophonic and homographic branch that consisted in locating and limiting ambiguities linked to those phenomena in English and other languages.

16. Work continued since then in the Airbus Human Factors department by Florence Beaujard, Emmanuelle Cannesson, and Laurent Spaggiari

3. A sociolinguistic and etymological branch that consisted in giving preference to terms used in American English (as it is the official dialect for aviation English) and/or terms that contained Latin roots.
4. A documentation branch that cross referenced the different ways in which the same terms were used in different aeronautical references such as in regulations, maintenance, operational documentations and air traffic control. This helped converge on terms that are more frequently used in the domain so they are recognized and adopted by the biggest sample possible.

When a term respects all the criteria of the decision matrix, it is accepted in the new official terminology. If it fails to respect one of the decision branches then it is replaced with another term that does comply with the terms and conditions.

Concerning synonymy, experts helped identify valid, invalid, and groups of synonyms to be confirmed:

1. Invalid synonyms are ones that have subtle differences in meaning, and an expert gives a precise definition of use for each term to be included in the new accepted terminology.
2. Valid synonyms are ones that could be used interchangeably. An expert recommends the use of only one of the terms. The decision matrix helps validate the choice that will be representative of the group of synonyms in the new terminology. The others will appear as non-recommended.
3. Synonyms to be confirmed are ones that are not easily established by an expert, so interviews with 8 different experts are conducted (airline pilots, test pilot, and flight instructors) to confirm the most adequate term.

Secondly, concerning morphological reduction, respect of the "uniqueness criterion" is key: "one word, one meaning, one short form". A reductional matrix is used. It consists in a collection of abbreviation rules that depend on word length. Abbreviated terms were then assessed by pilots. A transparency criterion is determined when an abbreviation is correctly identified in limited time and without any context. New short forms were only proposed when the existing terms fail the transparency criterion. When it is impossible to generate a new transparent term (or if a term is rare, frequently used, or safer) the full form is recommended.

On the syntactic front, due to the lack of room on the screens, *"one of the main characteristics of the corpus is the quasi-systematic lack of function words such as "in", "of", "by", etc."* Spaggiari et. al (2003). As such, a critical step is making sure that the right word order is respected in order not to create syntactical ambiguities due to the linguistic economies in function words. Slobin (1985) notes that *"it is likely that elements such as case inflections, verb inflections, pre- or postpositions, and conjoining and subordinating particles provide major orienting points for the perception of the structure."*

For instance, "young horse breaker" has two meanings:

- <young horse> breaker: A breaker of young horses
- <young> horse breaker: A horse breaker who happens to be young

Therefore, with the absence of function words, writing "horse <young breaker>", while unnatural sounding, eliminates the first interpretation in the first sentence.

As with the terminological process, corpus analysis was necessary to determine the scope of potential structural ambiguities and to adopt a standardization principle. *"Consistency is one of the most basic usability principles. Therefore, the same information should be formatted in the same way to facilitate recognition."* Spaggiari et. al (2003)

Different linguistic phenomena are dealt with such as the use of negatives, coordination, conditions, etc. A set of 20 rules were provided to form homogeneous content and format. Figure 13 below provides a few examples of rules.

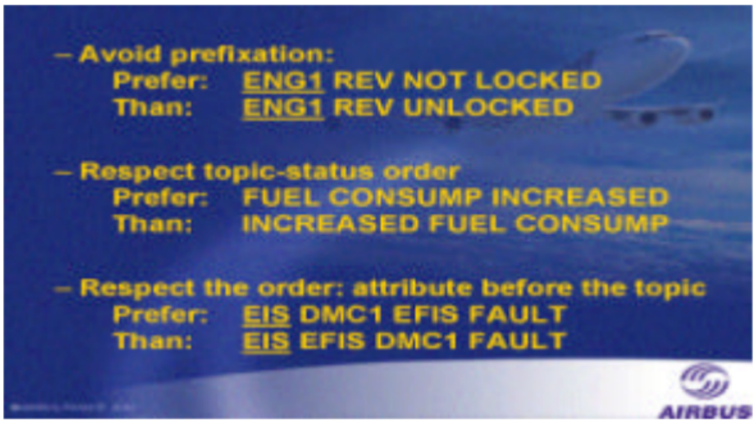


Figure 13. Example of Rules in Current Controlled Language

Interviews with experts, pilots, and people outside of the aeronautical domain (who had different mother tongues) were conducted at the time of Spaggiari's PhD (2002) to confirm the established rule, such as the word order validity by proposing several word order options, and then verifying the interpreted meaning for every sentence. For instance:

- young horse breaker tired
- horse young breaker tired
- tired young horse breaker
- tired horse young breaker

Results showed that even among different native language groups, people tend to follow the rule of association by syntactic proximity, especially when the "rheme", (or "focus", a grammatical category that determines which part of the sentence contributes new, non-derivable, or contrastive information Halliday, M. (1967)) is on the right of the syntagma: "Kitchen small door" (a small door of/in the kitchen) or "Door small kitchen" (door of/in small kitchen) instead of "Small kitchen door" which could have both interpretations. The proximity of the adjective to the noun disambiguates the noun phrase.

After discussing the origins and context of controlled languages, as well as giving an overview of the Airbus warning CL and how it is used in the cockpit, we should mention the notion of readability and text complexity, and how they influence the understanding of controlled languages in general, before tackling evaluations which show if CNLs effectively achieve the goals they were designed for.

3.3 Readability and Text Complexity

3.3.1 Definition

Readability is an essential notion in understanding CL evaluations, as it is a measure of testing for original text complexity. Note that readability is distinct from legibility, a measure of how easy it is physically to read a text.

DuBay's (2004) general definition states that readability is *"what makes texts easier to read than others"*.

According to Van Oosten et al. (2010), *"the concept of readability has been defined in a wide variety of ways, typically dependent on the author's intentions. For instance, Staphorsius (1994) defines readability of a text as the reading proficiency that is needed for text comprehension. The author's intention of designing a formula to determine the suitability of reading material given a certain reading proficiency is not without its influence in that definition".*

McLaughlin (1969), the author of the influential SMOG (Simple Measure of Gobbledygook) formula, on the other hand, defines readability as the characteristic of a text that makes readers willing to read on.

3.3.2 SMOG Formula

The SMOG grade test readability formula (cf. Figure 14) was introduced by McLaughlin in 1969. It estimates the years of education required to understand a text. Along with Flesch Kincaid readability formula, it is one of the most influential readability formulae. It is widely used to estimate the difficulty of public health materials.

The SMOG formula:

$$\text{grade} = 1.0430\sqrt{\text{number of polysyllables} \times \frac{30}{\text{number of sentences}}} + 3.1291$$

The SMOG Readability Formula

Step 1: Take the entire text to be assessed.

Step 2: Count 10 sentences in a row near the beginning, 10 in the middle, and 10 in the end for a total of 30 sentences.

Step 3: Count every word with three or more syllables in each group of sentences, even if the same word appears more than once.

Step 4: Calculate the square root of the number arrived at in Step 3 and round it off to nearest 10.

Step 4: Add 3 to the figure arrived at in Step 4 to know the SMOG Grade, i.e., the reading grade that a person must have reached if he is to understand fully the text assessed.

Figure 14. Steps Taken to Apply SMOG Readability Formula[17]

17. Readability Formulas. The Smog Readability Formula. [online] Available at: http://www.readability-formulas.com/smog-readability-formula.php [Accessed 10 Dec. 2018].

3.3.3 Flesch Kincaid Formula

The Flesch Kincaid readability formula is another influential formula. It was originally developed under contract to the U.S. Navy in 1975 by Peter Kincaid and his team. It is now also officially used by the US department of Defense as a military standard, but is also used in mainstream applications such as Microsoft Word.

It is designed to indicate how difficult a reading passage in English is to understand. Similar to the SMOG grade, Flesch Kincaid also includes a grade level formula. However, it also includes a reading ease branch. The formula for the Flesch reading-ease score (FRES) test is the following:

$$206.835 - 1.015\left(\frac{\text{total words}}{\text{total sentences}}\right) - 84.6\left(\frac{\text{total syllables}}{\text{total words}}\right)$$

The score we obtain from the formula is then compared to the following table (cf. Figure 15) in order to situate the score, its equivalent grade level and the notes on the ease of reading:

Score	Grading Level	Readability
100.00–90.00	5th Grade	Very easy to read. Easily understood by an average 11-year-old student.
90.0–80.0	6th Grade	Easy to read. Conversational English for consumers.
80.0–70.0	7th Grade	Fairly easy to read.
70.0–60.0	8th–9th Grade	Plain English. Easily understood by 13- to 15-year-old students.
60.0–50.0	10th–12th Grade	Fairly difficult to read.
50.0–30.0	College	Difficult to read.
30.0–0.0	College Graduate	Very difficult to read. Best understood by university graduates.

Figure 15. Scale of Readability and Grading Level for Flesch Kincaid (Flesch, 1979)

"The results of the two tests correlate approximately inversely: a text with a comparatively high score on the Reading Ease test should have a lower score on the Grade-Level test. Rudolf Flesch devised the Reading Ease evaluation; somewhat later, he and J. Peter Kincaid developed the Grade Level evaluation for the United States Navy."[18]

To give a few concrete examples taken from the Wikipedia page[19] on the Flesch Kincaid Readability Formula:

- Time magazine scores about 52.
- Harry Potter books have an average of 72.83.
- Harvard Law Review has a general readability score in the low 30s.

The highest (easiest) readability score possible is around 120 (e.g. every sentence consisting of only two one-syllable words; "The cat sat on the mat." scores 116).

While Amazon calculates the text of Moby Dick as 57.9, one particularly long sentence about sharks in chapter 64 has a readability score of −146.77.

One sentence in the beginning of "Swann's Way", by Marcel Proust, has a score of −515.1.

Flesch even provided a graphical tool (cf. Figure 16) to easily situate the complexity of a text and calculate its readability score. After counting the total number of words, sentences and syllables, we could apply the answer by drawing a straight line and thus retrieve the readability score and reading ease.

Like the Flesch–Kincaid and SMOG grade level, there are other readability scales like Gunning fog index (the less words and syllables the more readable a sentence is), Fry readability formula, Coleman–Liau index, automated readability index (ARI), etc.

18. Wikipedia. Flesch–Kincaid readability tests. [online] Available at: https://en.wikipedia.org/wiki/Flesch%E2%80%93Kincaid_readability_tests [Accessed 10 Dec. 2018].

19. *Ibid.*

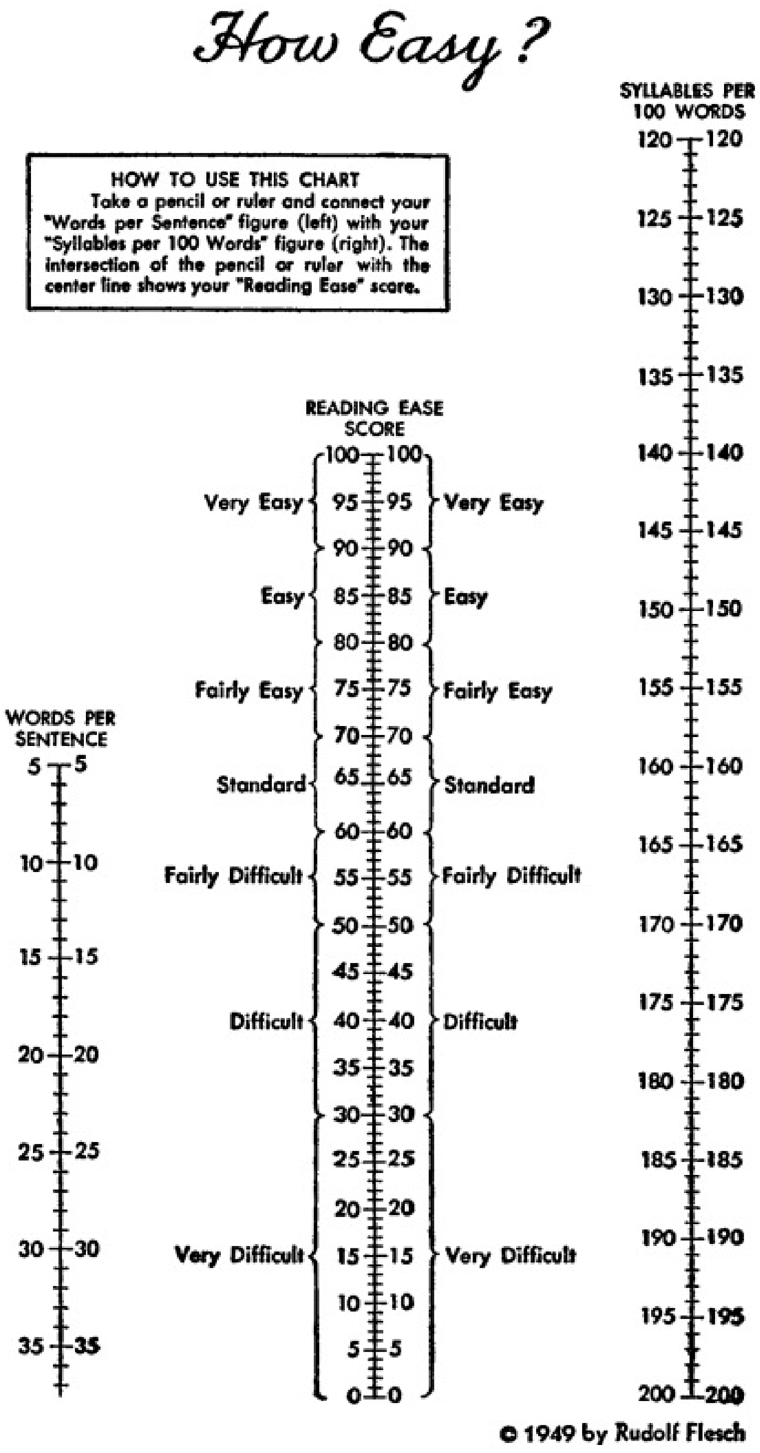


Figure 16. Graphical Tool to Situate Reading Ease in Flesch Kincaid Readability Formula[20]

20. Stuart Mill English. Reading Grade Level Doesn't Mean What You Think It Means. [online] Available at: http://stuartmillenglish.com/reading-grade-level/ [Accessed 10 Dec. 2018].

3.3.4 Readability and CNL Rules

The different formulae are all more or less based on sentence length and numbers of syllables in a sentence, in addition to a mathematical constant which would fit the characteristics of the corresponding score and its significance with respect to reading ease or grade level. They mostly serve as a guide to different professionals to be able to judge the difficulty of a certain text such as teachers, parents, librarians, researchers, and communication experts.

Unsurprisingly, the only rule that was shared by all the CNLs investigated in O'Brien's (2003) study (which compared different lexical, syntactic, textual, and pragmatic rules in 8 influential CLs) concerned restriction of sentence length, and even this rule varied widely in the degree of control specified, i.e. the number of words allowed and the arbitrariness of that choice. Siddharthan (2003) and Harley (2013) assert that long sentences can cause processing difficulties, because they overload working memory. This is further proof of how the classical definition of readability is tightly connected to CNL construction.

Our own definition of readability for the purposes of this research does not involve ease of reading, reading proficiency, or the characteristics that make readers willing to carry on reading. Readability in our sense is about usability of the text. Usability is defined as the *"extent to which a system, product or service can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use"* (ISO/DIS 9241-11.2 :2016).

Nielsen (1993) asserts that usability is made up of 5 components (Learnability (how easy is it to accomplish a task the first time?), Efficiency (once learned, how quickly are tasks performed?), memorability (how easy is it to keep the task in memory?), Reliability (how many errors, how severe, and how to recover?), Satisfaction (how pleasant is it to use?)). He also included property attributes such as utility (is all needed information provided?) and usability (generally speaking, is it easy and pleasant to find the needed information?), and usefulness (is the information usable and has utility for the user?). Tricot (2001) writes that what is considered as a good *useful* document must be both usable and have utility at the same time. If something is easy to find/use but is not what you need then it is essentially useless. On the other hand, it is equally useless to have a tool that contains the information you need but that is too difficult to manage.

Therefore, readability in our sense is when language becomes the product or service that should be used by specified users to achieve specific goals with effectiveness, efficiency, memorability, learnability, and satisfaction in a specified context of use. And in order to make a language more usable, we would need to ascertain what the inherent linguistic qualities of a text are, that make it comprehensible (effectiveness)? By comprehension, we mean that the information we want to transmit has been fully understood (learnability and memorability), the consequences of which should be the correct reaction to the information and the writer's intended meaning in the most optimal manner (efficiency and reliability: fast and accurate comprehension and reaction).

Therefore, one cannot ignore the strong relation between our readability ends (usability and comprehension of text) and the nature and goals of the Airbus Cockpit Controlled Language (whose ultimate general aim is to guide pilots into performing actions) on the one hand, and procedural texts on the other, which will be presented in the following section.

3.4 Procedural Texts

The purpose of a procedural text is to instruct and guide the reader with a task in order to always obtain the same result under the same circumstances. The presented information is usually sequential and broken into steps. According to Heurley (1997), « *[un texte procédural] est un texte dont la fonction principale est de communiquer une procédure, c'est-à-dire un ensemble d'opérations et/ou d'actions à exécuter dans le but d'atteindre un but donné* ».

Kern (1985) mentions that in procedural texts we are no longer merely addressing a reader but an end user, since reading has become a secondary task that helps the fulfillment of the main task. Furthermore, Nickl (2018) states that *"to instruct implies the participation of two parties, i.e., the person providing instruction (the instructor) and the one receiving it (the instructed). It also implies a disparity in the participants' respective knowledge bases, with the instructor possessing knowledge that the instructed needs. The act of instruction seeks to equalize the general knowledge level. Instructions relate to a differential in process-oriented knowledge, aim at enabling people to do things rather than merely providing knowledge of what or why things are. Unlike many other forms of publication, instructional text design is mostly not aimed at producing an emotional reaction or reception."*

He carries on to say that instructions belong to Searle's class of directive speech acts (Searle, 1969). We can differentiate them from representative speech acts, which seek to inform people about things (and which would also cover argumentative or explanatory texts, although these also target different knowledge levels). Many different speech acts could interlock in the composition of instructional texts. Typically, instructional texts use comparatively rigid structures, which are usually organized sequentially.

Nickl (2018) gives the following example to show the speech act sequences used in technical writing and other fields of instructional writing to standardize instructions:

"(warn) Never use the names of existing accounts. Existing accounts will be erased!

(describe-prerequisites) You need to be an administrator to install new user accounts.

(instruct 1) 1. Open the system control centre.

(instruct 2) 2. Activate the control "new user".

(instruct 3) 3. Allocate user rights (read, write, execute).

(describe-result) You will find the new user profile in the "user control centre".

In order to standardize the output of a multi-author team, you might stipulate that basic instructions must always follow this exact speech act structure. This would assure that a) warnings would always be the initial element, b) prerequisites are always mentioned, c) each activity concludes with a result, thus providing the reader with confirmation whether or not the activity was successful."

Rychtyckyj (2002; 2005) talking about SLANG (Standard LANGuage), Ford company's controlled language, also stipulates that sentences in the imperative mood which are meant to instruct must start with a main verb followed by a noun phrase.

3.4.1 Procedural Texts and Cognitive Processes

What is even more interesting to us is the cognitive processes involved in using procedural/instructional texts and how they influence readability. Ganier (2002) and Barcenilla and Brangier (2000) invoke several processes involved in the translation of procedural texts into actions:

1. **Developing and maintaining a goal in memory:** The goal remains in the working memory until the task is completed and allows control over the execution of the procedure.

2. **Reading and comprehension:** This data collection involves 3 levels of interdependent processing:
   a. **Processing of linguistic units:**
      i. *Encoding of visual stimuli:* First pre-lexical processing which could be influenced by several factors such as character size, word length, familiarity of the words by the end user, and the graphical aspect.
      ii. *Lexical access*: Lexical units are recognized and word meaning is identified from context. For example, end users could find the use of specialized jargon and abbreviation difficult at this stage.
      iii. *Syntactic analysis:* The structure of a text and its grammatical functions that give meaning to every word in the sentence.
   b. **Processing of propositional units:** The literal significance of a sentence. In the same way one could read a word and not understand its significance, one could read a sentence and not understand its meaning. Propositional processing allows the reader to form a semantic representation of what was read on the basis of the syntactic structure. The reader constantly ascertains the coherence of a text by linking what is being read and his or her domain knowledge. This process is accompanied by an inferential processing and a memory search to create a coherent whole. Those inferences could be anaphoric, interpretative or logical.
   c. **Development of a mental modal** or a referential representation of information (what the text refers to): This process requires significant inferential activity by the user (producing new information from existing ones), integrating at the same time new information from the text and pre-existing outside information.
3. **Development of an action plan**: During this phase, the previous mental modal will be transformed into a procedural representation that will trigger the executions of the actions.
4. **Execution of actions.**
5. **Proceduralization**: Knowledge and information accumulation that will create a procedural-like representation which will in turn be stocked in the long-term memory. This process will allow for the general learning of the task and its mechanical execution in the current

assignment and on the long term (if one performs a similar procedural task).

Based on these steps, we created the following diagram (cf. Figure 17) summarizing the cognitive processes involved in the execution of a procedural text. The three main steps of the cycle include reading, comprehension, and execution. Inside the cycle are the sub steps that account for the linguistic information processing, the development of a mental model, and action plan that precede execution. The proceduralization step is represented in the cyclical and iterative nature of the diagram.

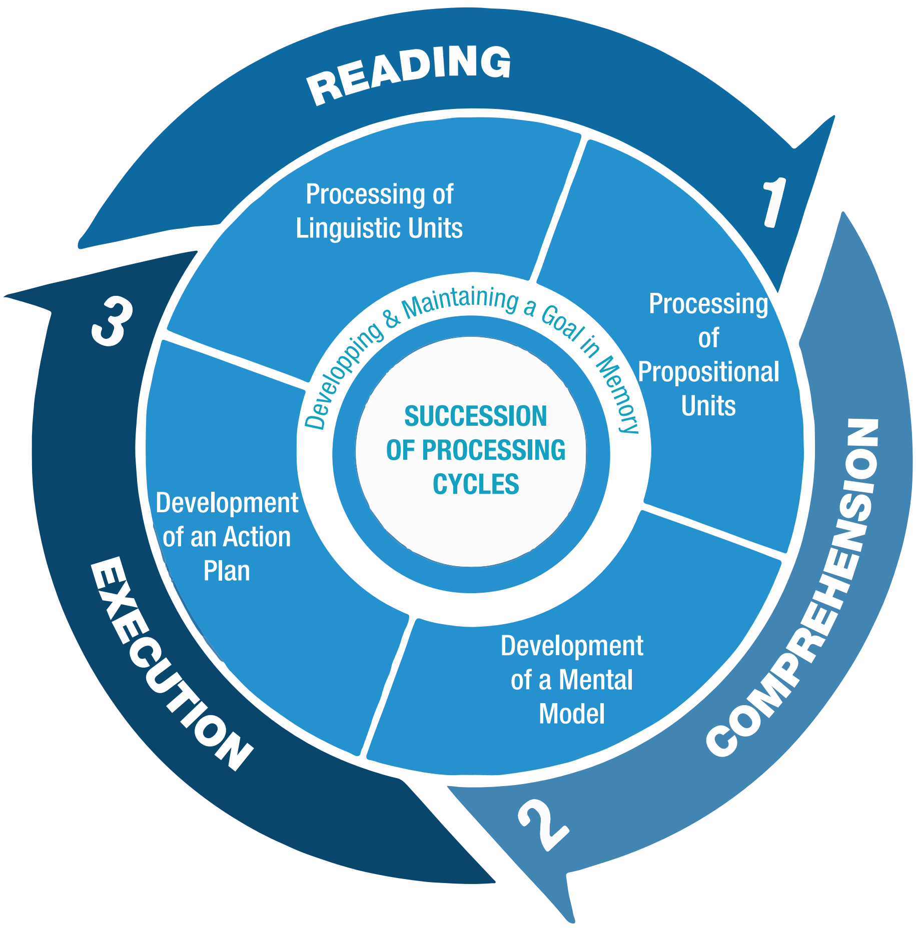


Figure 17. Cognitive Processes Involved in the Execution of a Procedural Text

In this way, we provided an overview of the cognitive processes involved in completing a procedural task. We could infer that a lower score of readability (more linguistic complexity) would potentially lead to slower reaction times in the execution of actions, perhaps even errors.

As mentioned before, our main goal is to establish a more optimized (better comprehension, faster reaction times, and limited training needs) controlled language for the cockpits. Therefore, if we consider readability to be first and foremost about the usability of a text, as we do, then we must consider the linguistic aspects of a sentence that make it comprehensible in the most optimal manner, and how those linguistic aspects influence comprehension and execution.

In order to do so, and since, across domains, the finality of comprehension oriented CNLs is the accomplishment of a task, we delved into CNL evaluations and their actual efficacy in achieving the goal they set out to accomplish. What are the advantages of using CNLs? Has there been any scientific evidence to attest to the efficacy of CNLs over their more natural counterpart?

3.5 CNL Empirical Evaluations

In this section, we are going to present some of the evaluations conducted by different entities with the goal of finding empirical proof of efficiency in established comprehension oriented CNLs, or for specific language rules.

3.5.1 Linguistic Redundancy Effects on Pilot's Comprehension

Carol Simpson (1976) studied the effects of linguistic redundancy on pilot's comprehension of synthesized speech (a study done for Human Factors research in aviation in a psycholinguistics context in NASA's Ames Research Center). The study does not fall into the domain of CNL construction, as it does not evaluate an established controlled language, but it psycholinguistically evaluates different language forms used in aviation, and makes a point about linguistic economies.

Simpson (1976) showed that by taking the time to form clear unambiguous sentences using the same original keywords, the message was detected more accurately and pilot's reaction times were faster. For instance, the message "fuel low"" was inserted in a sentence in the same order "The fuel pressure is low"" and "gear down"" was inserted in "The landing gear is down". The same goes for "Autopilot disengaged" and "The autopilot is disengaged". Response times to sentences were approximately 1 second shorter than response times to two-word messages. The results take into account the duration of the messages. That is to say, even though the duration of the stimuli containing the keywords in sentences was longer, the reaction times in total were still faster in the case of keywords in sentences than in the case of simple keyword messages. The experiment also showed that keywords in sentences were approximately 20 percent more intelligible than keywords presented alone.

Moreover, Hart and simpson's (1976) concurrent study for NASA also showed that sentence-length messages appeared to require less attention to comprehend than two-word keyword messages.

The authors explained that cockpit alarms tend to be presented in the form of short keyword messages rather than in the form of long sentences, as brevity is usually preferred because of the small window of time that the pilots have to react in time-critical situations. Therefore, the obvious way to economize on the time of stimuli presentation was to make the messages as short and precise as possible so as to keep only the relevant information, and eliminate redundancy provided by a sentence structure, i.e. the suppression of syntactic sentential elements, function words, etc.

However, it was concluded in Simpson's research that the syntactic and semantic constraints provided by a sentence frame (which adds redundancy and explicitness) reduced the possible interpretations of keyword alerts. Furthermore, the pilot participants mentioned that *"the longer pattern of the sentence with extra words between the critical ones gives you more time to understand the words"* and in their case react faster to the alert.

While these results are based on aural alerts, one could hypothesize that the same argument would work on written alerts. Simpson's study results showed that certain language structures (non-simplified natural language structures) actually decreased response time, which is a factor that is particularly of interest to us for optimizing comprehension. Additionally, Simpson and Hart's experiments are some of the only experiments that tested accuracy of comprehension and time in short injunctive messages as opposed to long chunks of text.

3.5.2 The Search for Empirical Proof in AECMA SE Evaluations

Established controlled languages (BASIC English, PLAIN English, AECMA SE, etc.) have often been criticized for lack of empirical research that justify their rules and existence such as in Flesch (1944), Shubert et. al (1995), and Eckert, D. (1997).

Additionally, Hinson (1988) in his article "Simplified English-Is it really Simple?" states that *"AECMA's Simplified English claims to be founded on readability research. It would be interesting to establish the nature, validity, and appropriateness of the research used. It would also be helpful to know of any research carried out on Simplified English manuals in use."*

Furthermore, Holmback et. al (1996) point out that the level of difficulty in a document at which SE becomes beneficial has not been identified in the literature.

O'Brien (2006) states that the lack of empirical research is also due to the fact that a lot of controlled languages are developed in proprietary environments (mainly in industry or government, and probably for reasons of intellectual property and judicial accountability in case of misuse) and studies attesting to their efficacy (or in this example, production costs), if any are available, are not publicly accessible: *"few empirical studies on CL have been published. This can be attributed to the fact that the implementation of CL is most often executed in a proprietary environment. Thus there is little published evidence that using a CL reduces information production costs."*

More recently, Ryan (2018) concludes that: *"there remain a number of issues concerning the efficacy of controlled languages in actual use and the quantitative evaluation of the practical gains they convey."*

As experts in the field criticized CLs for claiming anecdotal evidence, they argued that CL rules were not empirically tested by linguists or cognitive scientists, and are sometimes created directly by engineers, end users, and technical writers who often recycle good practices and writing rules without scientific evidence that those rules offer better comprehension than natural language rules. In safety critical domains, where miscommunication and misinterpretation could lead to potentially dangerous situations, and where reaction times are essential to optimal task completion, it is crucial that CL rules are evaluated. To this effect, a wave of large-scale experiments in the mid-90's (Shubert et al., 1995; Chervak, 1996; Chervak et al., 1996; Eckert, 1997; Stewart, 1998) was launched to acquire the empirical evidence that AECMA SE lacked. These studies will be of direct interest to our research. The experiments conducted and relevant results will be exposed here.

3.5.3 Comprehensibility of Simplified English

Shubert et al. (1995) were interested in testing the effects of AECMA Simplified English on comprehension, location of information on aircraft maintenance work cards and response times. They compared pre-SE work cards (non-SE/uncontrolled language) and their SE version. As a reminder, AECMA SE is used to write aircraft maintenance work cards, which mainly consist in describing tasks and procedures to be undertaken by maintenance technicians when working on aircraft.

An example from the warning listed in the non-SE version is shown below:

> "**Do not stand** on **controls bay access door**, 313L, or service access door, 311AL. The weight of personnel on these doors could cause their spring-loaded latches **to release**, and personnel to be injured by falling through the opening."

The same warning listed in the SE version is listed below:

> "Stay off [**shorter imperative**] the service access door, 3 11 AL, and the access door to the controls bay [**breaking down of noun cluster**], 313AL. Your weight can release [**active**] the spring-loaded latches on the door. If you fall through the door, injuries can occur."

In the non-SE version there was a 24-word long sentence while in the SE version, two 9/10-word sentences which use SE rules. They also added an explicit warning consequence with "if" condition: "If you fall through the door, injuries can occur." Stating the possible hazards is important to the effectiveness of a warning (Wogalter et al., 1987) and SE seeks to be more effective by including information about risks and hazards.

Another example of the same sentence in SE and non-SE versions:

- Non-SE: Center section of diaphragm must be found and **removed**, or reservoir could malfunction.
- SE: Make sure you remove [**imperative suggests action must be done**] and discard the center section of the diaphragm (18). If the center section of the diaphragm (18) is not removed [**repetition stresses consequence of action not done**], it is possible that the reservoir will not operate correctly. [**explicit consequence**]

The procedure consisted in a reading comprehension task. After participants read the text, they answered a multiple-choice question about the material they read. For example:

"What are the consequences of standing on the access doors?"
The answer choices for this question are listed below:
(a) the door may be damaged
(b) the door may be jammed
(c) the person standing on the door may be damaged
(d) both a and b

A second part of each question asked subjects to identify where in the document they found the answer. The researchers were particularly interested in how quickly and accurately the subjects would locate their answers because technicians were allowed to refer back to their instructions (the point was not to memorize). Questions were randomized and controlled for difficulty (when questions were deemed too easy, they were replaced with more difficult ones).

Participants consisted of 90 native English speakers and 31 non-native English speakers from 6 Engineering classes (Advanced Technical Writing and Oral Presentations). Subjects were randomly assigned to read one of four documents, split into two procedures A and B, in either the SE version or the non-SE version. Procedure A was more difficult than procedure B.

Native speakers of the SE documents performed significantly better than the readers of the non-SE documents $F(1,86) = 11.082$, $p < 0.05$ Means 17.878 (SE) and 16.653 (Non-SE). However, there was a significant interaction of language by procedure type $F(1.86) = 24.515$ $p < 0.0001$:

Figure 18 illustrates that there is a substantial difference in mean scores between SE and non-SE in procedure A, the more difficult procedure (18.65 SE Vs.15.5 Non-SE) but not procedure B, the less difficult procedure (17.14 SE vs. 17.8 non-SE).

Concerning location of information for native speakers, a similar significant general effect was observed, with a similar significant interaction by document type (procedure B not significant).

Similar effects were observed on non-native speakers: General significance for comprehension and location, with a significant interaction by document type (procedure B not significant).

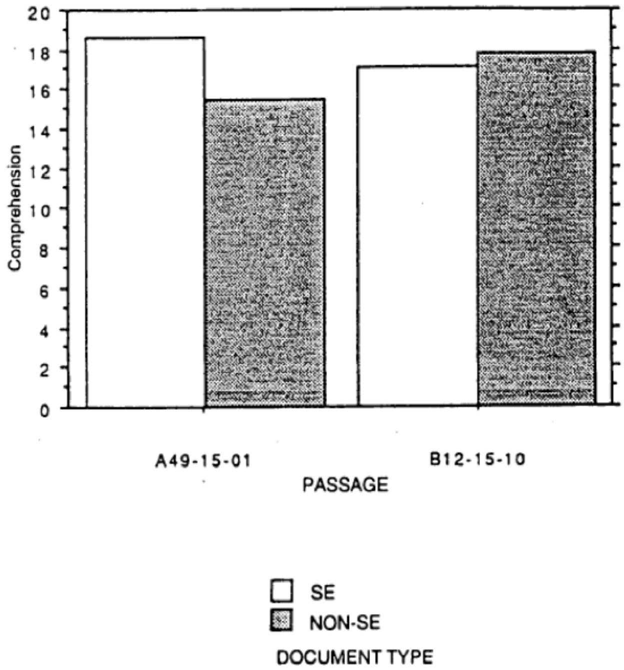


Figure 18. Mean Scores of SE and Non-SE scores in Procedure A and B

However, concerning reaction times neither natives nor non-natives' results showed significant results privileging SE. What is interesting also is that for procedure B (easier) the subjects reading SE versions of that document took slightly longer than those reading the Non-SE versions.

Therefore, SE was apparently more comprehensible and content was easier to locate for the longer more complicated procedure (by number of paragraphs and sentence length, higher Flesch Kincaid score), but content was not significantly more comprehensible or easy to locate for the subjects working with the shorter easier procedure. This study shows that the benefits of using SE may be document, and difficulty, specific. Further research on when and why it is useful would offer more answers. And finally, SE did not significantly improve time scores for native or non-native speakers.

3.5.4 Field Evaluation of Simplified English

Chervak et al. (1996) experiment is based on the previous experiment conducted by Shubert et al. (1995).

The main difference between the two is that Chervak et al. (1996)'s participants are aircraft maintenance technicians (AMTs) who use SE work cards on a daily basis, and Shubert's are Engineering students who have had technical writing classes.

175 AMTs from 8 major air carriers were given a reading comprehension test of 4 different work cards that were actual Boeing work cards:

- 2 work cards written in SE and 2 in Non-SE

17 potential work cards were analyzed. The non-SE versions were analyzed in terms of total words, mean words per sentence, percentage of passive voice, the Flesch-Kinkaid reading score, and a task difficulty rating of each work card by an experienced engineer. 4 work cards were finally chosen: 2 easy and 2 difficult.

Each AMT was given written instructions for completing:

- a demographic questionnaire
- a reading comprehension test (Flesch Kincaid reading levels) and vocabulary test for general English comprehension
- the actual work card comprehension task

Of the 175 AMTs, 157 were native and 18 non-native speakers of English: Natives took on average 20.5 minutes while non-natives took 24.7 minutes.

Results showed that SE was superior with regards to general accuracy increasing from 76% for non-SE texts to 86% for SE texts. The effect was most marked for non-native English speakers (67% to 87%). Therefore, SE helped non-native speakers to reach the same level of performance as native speakers.

The two easy work cards did not show significant differences between SE and non-SE, but for the two difficult work cards, SE was significantly more accurate.

Concerning reaction times, each work card had a somewhat different effect, none of them significant. Work cards easy-1 and difficult-2 gave slower performance times on average and the others faster performance times on average. All in all, reaction time averages were inconclusive.

Therefore, the experiment did not show general SE superiority over non-SE since it is document, and difficulty, specific, and does not economize on reaction times. It was concluded by Chervak et al. (1996) that the effectiveness of SE is greatest where it is most needed: for non-native English speakers and for difficult work cards. With native English speakers and easier work cards, SE will not adversely affect performance.

3.5.5 Effects of Simplified English on the Performance of a Maintenance Procedure

Chervak (1996) measured the effect of SE on the actual physical activity of the task performed, to examine whether SE can actually improve a person's ability to not only comprehend but also perform a task. The tasks were also split into easy and difficult for SE and non-SE.

The experiment consisted in accomplishing a task based on lawnmower engine maintenance. Participants were all native speakers composed of 9 automotive maintenance mechanics students and 9 experienced mechanics. Participants were randomly assigned to perform both easy and difficult tasks using work cards that were the same version (either both SE or both non-SE). Participants were then timed and filmed while performing tasks.

Results showed that expert mechanics made significantly fewer errors and completed the tasks in significantly less time than the student mechanics, and that the easy task was completed in a significantly shorter time than the difficult one. However, SE did not significantly increase accuracy rate or decrease reaction times.

3.5.6 Effects of Simplified English in a Non-native Speaking Environment

This research, done by Eckert (1997) was set to determine if the use of SE improved task comprehension of non-native English-speaking aviation maintenance technician students in a non-English speaking environment (Mexico): 148 aircraft maintenance technician students from 4 aviation schools in Mexico City.

Similar to Chervak et al. (1996) and Shubert et al. (1995), a standardized test was administered to the participants to determine their English reading comprehension before the randomly assigned SE and non-SE task cards. The task cards and tests measured subject comprehension of maintenance procedures.

The results indicated that there was no statistically significant improvement in task card comprehension when using SE.

3.5.7 Effect of Simplified English for Non-native English Speakers from Different Countries

This study conducted by Stewart (1998) looked for an overall difference between the comprehension of a controlled language (SE) and the comprehension of standard English (non-SE) by non-native English speakers who were electronics technician students. It also compared the effect of English reading level on readers' comprehension, ability to locate information, and task completion time.

There were 41 non-native English speakers. They were students from 21 different countries enrolled in electronics technician programs at a technical school in British Columbia. 63% were enrolled in aviation-related programs.

An accuracy level test was first administered to participants to measure their English-reading ability by grade level. Participants then had to do a comprehension test for which they were randomly assigned to read either the SE procedure or the non-SE procedure.

Only one document was used with two variations, an SE version and a non-SE version. The document contained an aircraft maintenance procedure.

Results showed that overall there was no significant difference between the comprehension of SE and non-SE by non-native English speakers, even though means followed the usual trend (average accuracy of SE higher than non-SE). No statistically significant relationship was found for time scores for either SE or non-SE.

3.5.8 Text Complexity and Text Simplification in the Crisis Management Domain

Temnikova (2012)'s PhD thesis deals with the complexity of the crisis management sublanguage. She studies methods to produce new clear texts, and rewriting pre-existing crisis management documents which are deemed too complex. She defines *"Text Complexity (TC) (or "Text Difficulty", G. Leroy et al., 2010) as the internal characteristic of a written text which affects human comprehension during reading or the performance of computer applications processing text."*

Temnikova performed two evaluations, one on extrinsic tasks (consisted in testing the impact of simplification on reading comprehension and manual and machine translations), and the other tested user acceptability. She mentions that "*although thousands of CM* [controlled languages for crisis management] *texts do already exist and more and more of them are currently being produced, the contribution of the Natural Language Processing (NLP) field and of Linguistics to the field is under-developed.*"

Temnikova performed an evaluation on CLCM (Controlled language in Crisis Management). Temnikova explains that *"CLCM is a mixed-purpose CL designed mainly to improve human comprehension of written text in emergency situations, but it can also be used to ensure good translation results."* Figure 19 is an example of an original text (left) simplified with CLCM rules (right).

Original	Simplification
If you suspect there is something embedded, take care not to press on the object. Instead press firmly on either side of the object and build up padding around it before bandaging to avoid putting pressure on the object itself.	**How to treat severe bleeding** If you suspect there is an embedded object: 1. Avoid pressing on the embedded object. 2. Do the following actions simultaneously: • Press firmly on either side of the embedded object. • Build up padding around the embedded object. *Explanation: In order to avoid putting pressure on the object itself.* 3. Bandage the wound.

Figure 19. Comparative Example of an Original Text Simplified by CLCM Rules

The reading comprehension experiment involved 104 volunteers. It was an online experiment which consisted in participants reading simple and complex emergency instructions in random order and replying to multiple-choice questions after each instruction. They had limited response times. Four different complex texts of the same length and similar text complexity levels were used, and their equivalent manual simplifications. No participant was shown both the complex and simplified versions of the same text. Each participant read four texts in total: two complex and two simplified.

The effect was measured by comparing the proportion of correct answers given to the complex and simplified texts and by comparing the time necessary to provide correct answers for both kinds of texts. Participants had different native languages and most of them were university students or from the research community.

The overall results showed no general significance for accuracy nor for time of the CLCM simplification compared to the more complex language.

3.5.9 Recapitulation and Discussion of Evaluations' Results

To recapitulate the results of these evaluations, as seen in the comparative Table 1, Chervak et al. (1996) and Shubert et al. (1995) are the only two studies that showed general SE superiority significantly. In Chervak (1996), Eckert (1997), and Stewart (1998)'s studies there were no significant results to substantiate SE superiority over non-SE versions. Furthermore, in Chervak et al. (1996) and Shubert et al. (1995) (the only experiments showing general SE significance) there was a significant interaction of comprehension of SE and non-SE by document difficulty: The easy work cards (ones that described short and easy procedures as opposed to long and difficult ones) did not show any comprehension significance for SE, and only the difficult ones did. Therefore, content is not significantly more comprehensible or easy to locate for the subjects working with the shorter easier procedure. Chervak et al. (1996) showed that only certain work card types showed significant SE superiority over non-SE, which suggests that SE superiority, is document specific.

Finally, none of the experiments showed that SE significantly improved reaction time. Shubert et al. (1995) even noted that in the easier work cards the subjects reading SE documents required more time to respond.

All of these studies concluded that while the superiority of SE did not show general significance except in certain documents and difficult conditions, it did not adversely affect comprehension in the other conditions. Therefore, Chervak et al. (1996) concluded that SE was suitable for use especially where it is needed most: in hard and long work cards and for non-native speakers. However, most interestingly, Eckert (1997) and Stewart (1998) who only tested non-native speakers did not find any SE significance.

Temnikova (2012)'s experiment is different from the previous studies since it was done 15 years later, and was testing a different CL: Controlled Language for Crisis Management (CLCM). It is relevant here because, like the previously mentioned 5 experiments, it also tests a human-oriented CL in a behavioral experimental protocol. Results showed that there was no statistically significant global superiority of the simplified CLCM over the "complex" natural language. It was significant in certain sets of text (again, document-specific) and it did not show any significance with regards to response time. We summarized all of these results in Table 1.

Author/year	Shubert et al. 1996	Chervak et al. 1996	Chervak 1996	Eckert 1997	Stewart 1998	Temnikova 2012
Native and non-native	Both	Both	Native	Non-native	Non-native	Both
Participants: natives	90 natives	157 natives	18 natives	0 natives	0 natives	22 natives
Participants: Non-natives	31 non-natives	18 non-natives	0 non-natives	148 non-natives	41 non-natives (21 different countries)	83 non-natives
Profession	Engineering students	AMT's	9 maintenance students and 9 experienced mechanics	Aviation maintenance students	Electronics technician students	All walks of life(because not testing SE, but CLCM)
Country	English speaking	English speaking	English speaking	Non-English speaking (Mexico)	English speaking	N/A (Online experiment)
Procedure	Reading comprehension, between subject	Reading comprehension, between subject	Performing maintenance, between subject	Reading comprehension, between subject	Reading comprehension, between subject	Reading comprehension, between subject
Tested for English comprehension	No	Yes (but not specifically for non-natives)	No	Yes	Yes	No (only self-evaluation and not used in analysis)
General SE Significance: doc type	Yes	Yes	No (means followed trend)	No (means followed trend)	No (means followed trend)***	No (means followed trend)
Significance SE comprehension: easy	No	No	N/A	N/A	N/A (only 1 workcard)	N/A
Significance SE comprehension: difficult	Yes	Yes	N/A	N/A	N/A (only 1 workcard)	N/A
General Significance: location	Yes	N/A included in comprehension	N/A	N/A	No	N/A
Significance location: Easy	No	N/A included in comprehension	N/A	N/A	N/A (only 1 workcard)	N/A
Significance location: Difficult	Yes	N/A included in comprehension	N/A	N/A	N/A (only 1 workcard)	N/A
Significance SE: native speakers	N/A	N/A (will not adversely affect)	N/A	N/A	N/A	N/A
Significance SE: non-native speakers	N/A (means yes, but not tested for significance)	No (more marked)	N/A	N/A	No	N/A
Significance: time/SE	No**	No (will not adversely affect)	No	N/A	No	No
Significance: time/native speaker	N/A	Yes (normal)	N/A	N/A	No	N/A
Significance type of workcards	N/A (only easy/difficult was tested)	Yes (only certain workcards)	N/A	N/A	N/A	Yes (only certain sets of text)

What is interesting, however, is that for procedure B (easier) the subjects reading **SE versions of that document took slightly **longer** than those reading the Non-SE versions.
***The study also concluded that the SE participants required higher mean English-reading ability to obtain a mean task card test score similar to the non-SE participants.

Table 1. Comparative Table Summarizing Most Relevant Results of Different CL Evaluations Taken from Jahchan et al. (2016)

Results in Table 1 are relevant to our study for two main reasons:

1. We are interested in the optimization of reaction time and these AECMA SE and CLCM evaluations show that simplifying a language does not economize time and
2. because our corpus is made of short relatively uncomplicated sequential procedures (one action per line, etc.), and these results do not show simplified English superiority (in accuracy or time) when it comes to easy procedures.

Consequently, to answer one of the questions that we evoked at the beginning of this section ("Has there been any scientific evidence to attest to the efficacy of CNLs over their more natural counterpart?"), these results could be considered inconclusive as no significant effects were observed in more simplified texts regarding reaction times. The results observed in the first two experiments (Chervak et al. (1996) and Shubert et al. (1995)), which showed accuracy significance for the simplified text in certain conditions and for certain populations, failed to be reproduced in similar subsequent experiments (Chervak (1996), Eckert (1997), and Stewart (1998)), as no accuracy significance was observed in any condition.

3.6 Evaluation Techniques

While the evaluations in Section 3.5 are a good starting point, reading comprehension tasks (with multiple choice answers) do not accurately evaluate the real understanding of a certain text (especially a predominately procedural text), as the results will strongly rely on memory and skill, and do not show whether the actual performance or execution of a task will be done correctly.

Additionally, with reading comprehension tasks we open ourselves to many uncontrolled biases such as the unlimited time that participants have to answer after they have read a whole text with many details. In these evaluations, the texts were always about a maintenance procedure or an emergency task to be performed, yet the participants did not perform the task, but merely replied to questions about certain steps in the procedure.

In other words, we do not know whether the actions that are described in the text are accurately understood, whether they would have been correctly performed as such. We could only conjecture to the potential comprehension of a text that describes an action that the participants will not be performing. Nickl (2018) writes that *"instructional texts aspire to enable the reader to perform actions. And as the saying goes: the proof of the pudding is in the eating – and the proof of instructions is in the performance. With usability testing, a firm connection between text and users can be established. Comprehensibility therefore is no longer a trait of a given text but becomes firmly connected to target groups and to real world effects."*

Connatser (1999) also writes that *"most audiences of technical documents read to do. Therefore, usability testing of a document seems much more appropriate for measuring how effectively a text conveys technical information than a formula."*

Therefore, these evaluations' primary shortcomings are due to the nature and assessment of the task itself. Proper behavioral evaluations that accurately test human comprehension and performance are an aspect that is missing in the human-oriented CNL domain. Before we discuss our proposed evaluation approaches, we will give a brief definition and history of psycholinguistics as a field, its possible applications, and how we will use psycholinguistic tools and methods for our linguistic ends.

3.6.1 Psycholinguistics: Definition and History

The term "psycholinguistics" also known as the "Psychology of Language" was introduced by American psychologist Jacob Robert Kantor in his book *An Objective Psychology of Grammar* (Kantor, 1936).

Garnham (1985) defines *"psycholinguistics [as] the study of the mental mechanisms that make it possible for people to use language. It is a scientific discipline whose goal is a coherent theory of the way in which language is produced and understood."*

A landmark event for Psycholinguistics as a field was an interdisciplinary summer seminar at Cornell University in both Psychology and Linguistics. It was held from June 18 to August 10, 1951. The seminar identified the possible relationships between Linguistics and Psychology, and made several recommendations for studying and advancing psycholinguistics as a scientific domain.

Carroll (2008) states that the main psycholinguistic concern is determining *"the cognitive processes [that] are involved in the ordinary use of language.*

By 'ordinary use of language' I mean such things as understanding a lecture, reading a book, writing a letter, and holding a conversation. By 'cognitive processes,' I mean processes such as perception, memory, and thinking. Although we do few things as often or as easily as speaking and listening, we will find that considerable cognitive processing is going on during those activities."

O'Grady et al. (2001) make a link between studying Linguistics as a science (analysis of syntactic and lexical structures, phonetic and phonological composition, morphological derivations, etc.) and language processing in the brain: *"[...] an account of language processing also requires that we understand how these linguistic concepts interact with other aspects of human processing to enable language production and comprehension."*

Moreover, psycholinguistics is an interdisciplinary field that uses psychological and neurobiological factors in the cognitive sciences domain to enable us to study how the brain processes, comprehends, and acquires languages, etc.

Modern psycholinguistic research makes use of biology, neuroscience and neurolinguistics (how language is represented in the brain), cognitive science, linguistics (phonetics, morphology, syntax, semantics and pragmatics, etc.), information science and artificial intelligence (AI) to study how the brain processes language. *"Indeed, much of the early interest in language processing derived from the AI goals of designing computer programs that can turn speech into writing and programs that can recognize the human voice."* (Field, 2003)

Psycholinguistics traditionally collects behavioral data through different tasks that test subjects' language abilities. Thus behavioral tasks are a means and not an end. These tasks are used to test theoretical hypotheses related to the way we imagine the cognitive linguistic system to be. It is therefore a hypothetico-deductive method that infers conclusions from behavior with respect to linguistic stimuli in experimental paradigms. These methods test the impact of experimental factors that are expected to have a significant effect on comprehension.

Experiments based on these tasks consist in analyzing the response of participants to various stimuli or inputs that could be internal or external, on a conscious or subconscious level and the reactions could be voluntary or involuntary. Behavioral tasks often involve analyzing measures of performance such as reaction times, accuracy of response to stimuli, analysis of eye-tracking data (beginning with Rayner (1978)), etc. Cognitive processes are then inferred from those behaviors as a response to tightly controlled experimental designs and isolated variables.

However, *"the advent of neuroimaging opened new research perspectives for the psycholinguist as it became possible to look at the neuronal mass activity that underlies language processing. Studies of brain correlates of psycholinguistic processes can complement behavioral results, and in some cases [...] can lead to direct information about the basis of psycholinguistic processes."* (Pulvermüller, 2009)

Brain surgery (if an illness made it indispensable) used to help researchers discover how language works in the brain until the advent of non-invasive techniques of neuroimaging which include PET (positron emission tomography) scans – FMRi (functional magnetic resonance imaging) – ERPs (event related potentials) – EEG (electroencephalography) – MEG (magnetoencephalography) – MRI (magnetic resonance imaging), etc.

Computational modeling is also used to practice testing cognitive processing models such as the mechanisms involved in reading and word/sentence recognition. Executable computer programs are used to test the proposed models. The DRC[21] model proposed by Coltheart et al. (2001) is a good example of computational modeling in relation to language perception.

3.6.2 Psycholinguistics: Themes and Methods

Psycholinguists study many different topics, but these topics can generally be divided into answering the following questions Esenova (2017):

1. *how do children acquire language (language acquisition)? (*Caillies and Le Sourn-Bissaoui (2013)*, among others)*
2. *how do people process and comprehend language (language comprehension)? (*Gernsbacher (2013)*, among others)*
3. *how do people produce language (language production)? (*Bock and Levelt (2002)*, among others)*
4. *how do people acquire a new language (second language acquisition)? (*Giraudo and Hathout (2012), *among others*)

21. DRC model: Dual-Route Cascaded model

In behavioral psycholinguistic experiments subjects are usually presented with linguistic stimuli and asked to make a judgement about it. They may be asked to make a judgment about a word (lexical decision), reproduce the stimulus, or name a word presented on a screen aloud.

Researchers use priming effects, where a priming word/letter/phrase/picture/sound speed up the lexical decision for a related target that appears subsequently, and which could also be a word/letter/phrase/picture/sound. The tasks usually require a yes or no response. The responses are recorded to analyse reaction times and correct answers related to each stimulus.

Examples include reading a word on a screen which could be a valid or invalid word in a specific language, or evaluating if a morpheme or a sound constitute an acceptable sequence for different populations such as native speakers or bilinguals of a given language/languages, etc.

To give a more concrete example, Fischler (1977) found that related word pairs such as cat/dog were recognized faster (faster response times) when compared to unrelated word pairs such as "bread/stem". This facilitation suggests that semantic relatedness can make word recognition and memorization easier. In other words, the word "cat" is the prime (it is shown first, could be subliminal or supraliminal) that facilitates the recognition (reaction times for pressing "yes" (I recognize the word) on a button) of the target word "dog" (shown after) because of the close semantic relations that the two words share. However, showing the prime word "bread" does not facilitate the recognition of the word "stem" as those two words are not semantically related. Therefore, it could be concluded that when the brain is prepped for a target word by a semantically related prime word, it is more ready to understand and recognize a word as part of a specific language.

In such experiments, the stimuli are naturally separated by pairs of non-words such as "acornfly" and "vonk" in order to make the experimental design more rigorous (include both "yes" (I recognize) and "no" (I do not recognize) alternated answers to be able to test for word recognition). Moreover, psycholinguists are also interested in the morphological processing during word recognition to establish how morphologically complex or simple words are analysed by the brain and stocked in long term memory (Giraudo & Voga, 2014). Some believe that word recognition happens on a decompositional level (morpheme-based approach) while others believe it happens on a whole word level (word-based approach). Morpheme-based approaches rather rely on analyses in linguistic theories of morphology and lexicology. These theories are also tested using mainly priming paradigms and lexical decision tasks.

Other relevant psycholinguistic experiments that include judgment tasks involve semantic congruence/ congruency/ congruity. These terms originate from Linguistics, and more particularly from the Semantics and Pragmatics subdomains to define the proximity of words in "semantic space" (Pollio 1964). It was later introduced in cognitive sciences to replace what was known as the "cross-over effect" (Audley & Wallis, 1965) and "affective value-distance" (Shipley et al., 1945; Dashiell, 1937). Oxford English dictionary[22] defines congruence as "*agreement or harmony; compatibility*". It implies that two notions correspond to one another, are in agreement, and equivalent, but nonetheless rarely implies that they are identical.

Therefore, the semantic congruity effect observed in congruence tasks stipulates that stimuli are coded in the brain on different qualitative dimensions such as (meaning, size, color, loudness, heaviness, etc.). When one compares two stimuli (or a prime and a target), the reaction times are faster when the two stimuli are congruent than when they are incongruent, as the perceptual system has to do an extra decoding step in order to make a judgement between the two. A classic example involves the Stroop effect introduced by Stroop (1935) and replicated abundantly through the years in the cognitive sciences domain. If word meaning and font color are congruent, subjects will name the font color significantly faster than when word meaning and font color are incongruent. For example, naming the font of the word "red" when it is written using red ink (congruent word meaning and word appearance) is an easier and faster task than naming the font of the word "green" when it is also written using red ink (incongruent word meaning and appearance). It is also faster to name the font color of a semantically neutral word such as "school" written in red ink than when the meaning of the word interferes with its appearance (such as the word "green" written in red ink).

Semantic congruency tasks are particularly of interest in this research as we will be applying a form of congruency tasks in our first evaluations to test hypotheses involving comprehension of controlled languages (cf. Chapter 5).

22. Oxford Dictionaries. Definition of congruence in English. [online] Available at: https://en.oxforddictionaries.com/definition/congruence [Accessed 10 Dec. 2018].

3.6.3 Psycholinguistics and Link with CNL Evaluations

When we use psycholinguistic tools in CNL evaluations, we are merely confirming or denying linguistic hypotheses using psycholinguistic methods (behavioral tasks, Event Related Potentials, etc.). We are not learning about the function of the brain via models of psycholinguistics but rather, using psycholinguistic and psycho-cognitive methodology to satisfy linguistic ends, in this case, the effectiveness of CNLs and the rules that make them.

The two disciplines must come together in a more effective manner, one that would reap the benefits of a tightly controlled psycholinguistic behavioral protocol evaluating reaction times and accuracy of comprehension in real-time participant performance.

Psycholinguistic evaluations do not usually deal in the performance of an action or the accomplishment of an instruction. This is because what interests psycholinguists most is to understand how comprehension happens or how the brain deals with information and makes links between linguistic notions and concepts. The impact of language specificities is studied to explore the organization of the cognitive linguistic system in the brain by building models to understand and describe the cognitive processes involved in language acquisition, or for instance in sentence comprehension.

Psycholinguistics emits hypotheses about how our cognitive system integrates/represents/codes the language and the nature of the linguistic elements that are actually coded. In other words, psycholinguistics seeks to verify whether what is described in linguistics for a given language is represented in long term memory. It makes assumptions about the architecture of lexicon and grammar in terms of mental representations, and the interactions between them that allow access to meaning.

On the contrary, our aim in this research is to check how language as a tool (in different forms) influences behavioral reactions (which, outside experimental paradigms, are also the end and not only the means), since what we as linguists and ergonomists are interested in, is how language works, or more concretely what exactly is in a language that allows us to react in a certain manner to a given task. In this sense, our methodology falls effectively within the realm of ergonomic linguistics.

This is why we cannot use psycholinguistic experimentation tasks and research designs exactly as they are used in the field, since our hypotheses and aims differ from those intended for this kind of experimentation. We cannot, on the other hand, ignore the virtues of the science responsible for conducting tightly controlled psycholinguistic experimental paradigms to prove narrow and detailed hypotheses. This is why it is essential to use the tools and methods of such experiments but adapt them to fit linguistic (function and structure of language) and ergonomic (performance and usability) ends, without falling in the trap of making questionnaire-like evaluations based on very broad linguistic hypotheses.

3.6.4 Psycholinguistic Tools in the CNL Domain, an Overview

We believe the limits of simplification must lie in the systematic behavioral evaluations of any established CNL and its various rules.

To this date CNL evaluations are not systematically enforced and very rarely put in place for human-oriented CNLs. There have been some evaluations of CNLs using NLP (natural language processing) tools in corpus linguistics-based approaches such as the verification of conformity of requirements (Condamines & Warnier, 2014; Warnier, 2018) or for text complexity (Tanguy & Tulechki, 2009), and machine translation (O'brien & Roturier, 2007; Aikawa et al., 2007), or for syntactic transformations and corpus alignment of specialized corpora with existing simplified corpora (Cardon & Grabar 2018), etc.

There have also been evaluations based on ontographs for knowledge representation and formal languages Kuhn (2010). In this paper, Kuhn (2010) contends that *"user studies are the only way to verify whether CNLs are indeed easier to understand than other languages"*. He argues that it is difficult to obtain reliable approaches with task-based and paraphrase-based evaluation approaches, and offers an alternative method for evaluating formal logic-based languages.

Task-based approaches consist in entering statements written using a CNL in a given tool that is pre-programmed to transform knowledge representations. An example from Bernstein and Kaufmann (2006) is the task "Create a subclass Journal of Periodical" for which the participants are expected to write a CNL statement in the form of "Journals are a type of Periodicals". The latter statement would be checked to see if it contains this information.

However, this type of approach mostly checks the "writability" of a sentence or the ability to write the statement in a given CNL by the help of a specially made tool for a specific controlled language. In a way, it is essentially an evaluation of the usability of a tool based on a CNL.

Paraphrase-based approaches could be tested independently from specific tools. Hart et al. (2008) present a task to test their CNL (the Rabbit Language[23]). They conducted an experiment where subjects were given one sentence written in Rabbit CNL and had to choose from one possible paraphrase written in natural language (English). Only one of the options is correct. They used made-up words like "acornfly" in order to prevent participants from using their own knowledge to answer correctly. An example from the task in which option 1 is the correct answer:

Statement: Bob is an instance of an acornfly.

Option 1: Bob is a unique thing that is classified as an acornfly.

Option 2: Bob is sometimes an acornfly.

Option 3: All Bobs are types of acornflies.

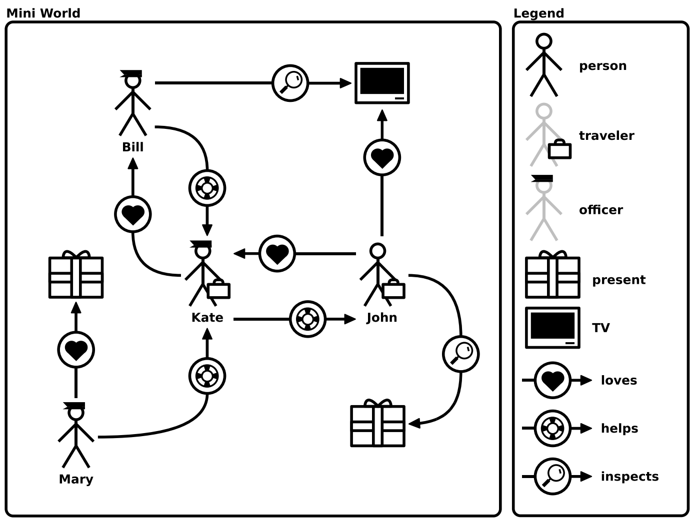
Option 4: All acornflies are examples of Bob.

Nonetheless, paraphrase-based tasks have some drawbacks, in the same way reading comprehension tasks presented in Section 3.5 are not adequate for testing actual user comprehension. One cannot be sure that subjects understood the paraphrases written in natural language, and slightly ambiguous phrases such as "is classified as" or "are types of", and the polysemy in "unique" and "sometimes". Therefore, Kuhn (2010) proposes an ontograph (contraction of ontology and graphs, a graphical notation that enables a tool-independent evaluation of human understandability of knowledge representation languages.). Kuhn writes that *"the basic idea is to describe simple situations in this graphical notation so that these situation descriptions can be used in human subject experiments as a common basis to test the understandability of different formal languages."*

Figure 20 is an example of an ontograph taken from Kuhn (2010). These ontographs are designed for testing comprehensibility of formal notation languages. In order to do so, an ontograph and several statements written in the controlled language to be tested are shown to the participants who have

23. A CNL designed for a scenario where a domain expert and an ontology engineer work together to build an ontology. The construction process is supported by a text-based ontology editor.

to decide which statements are true or false with respect to the shown ontograph. If participants manage to classify the statements correctly then it could be concluded that subjects understood the statements and the ontograph. This approach introduces an interesting way of evaluating CNLs but is limited to simple forms of logic and formal representations restricted to unary and binary predicates (not suitable for comprehension oriented CNLs). Additionally, depending on the complexity of the diagram, this method could in some cases create further ambiguities of deciphering the ontograph itself, on top of formal logic statements.



This is an example of an ontograph. The legend on the right hand side defines the types and relations. The mini world on the left hand side shows the actual individuals, their types, and the relations between them.

Figure 20. Example of an Ontograph Taken from Kuhn (2010)

Consequently, the previously discussed evaluations fail to enlighten us on the effectiveness of comprehension-oriented CNLs on the human cognitive processes of language comprehension, for instance by measuring reaction times and accuracy in performance. We argue that the relative lack of cognitive behavioral evaluations is equivalent to rendering CNLs mere style guides or good authoring practices, and the reasons for adopting certain rules over others are unreliable.

After giving an overview of the literature and the various CNL evaluations, we would like to discuss our core question, the theory behind it, and the approach we adopted.

4

Approach and Methodology

As a reminder, the main research goal of this study was to optimize comprehension, perception, and use of controlled languages in the cockpits (faster and more accurate comprehension, limited training needs). In order to do so, we looked into various evaluations that sought to prove the efficacy of controlled language simplification.

The results of those evaluations concerning AECMA SE and CLCM (simplification does not automatically equal better and faster comprehension) and Hart and Simpson (1976)'s research (linguistic redundancies (theoretically less simple) helped comprehension) have led us to give a more concrete form to some of our more central questions:

To what extent does text simplification improve overall comprehension and task performance? Does using a more controlled language (as opposed to an INL, or a less controlled more natural version) accomplish its comprehension-oriented goals? Do the many simplification rules in different controlled languages need to be tested separately to obtain empirical evidence of their validity?

4.1 Core Question

"Natural language being such a breeding ground for ambiguity, to communicate just one set of meanings while excluding many others is often impossible." Crystal and Davy (1969). In a sense, this statement is true since natural language has theoretically infinite possibilities of expression and interpretation,

but in another sense, natural language is the most common and constant tool in our cognitive processes of everyday life. Let us consider the following example, and the different possible interpretations associated:

I saw a man on a hill with a telescope:

1. "I saw a man, who was on a hill and had a telescope."
2. "I used a telescope to see a man who was on a hill."
3. "I was on a hill, and saw a man who had a telescope."
4. "I was on a hill, and used a telescope to see a man."
5. "I saw a man, who was on a hill, and the hill had a telescope on it."
6. "I saw a man, while I was on a hill which had a telescope on it."
7. "I use a telescope to saw [verb to saw] a man on a hill."
8. "I saw [verb to saw] a man on a hill that has a telescope."
9. Etc.

As we can see, this relatively simple uncontrolled natural language sentence could have many interpretations, which creates ambiguity if not paraphrased.

On the other hand, syntactic constructions, morphological derivations, way of thought, all come naturally in the way we acquire them at an early age; or the way non-native speakers of a given language, let's say English, first learn the language at its most basic form and construction: naturally, implicitly, and without any control.

Bisseret (1983) regards natural language as a *"universal tool of representation and of thought communication."* and Fodor (1975) states that it represents the *"language of thought"* that bears close resemblance to our surface language. Similarly, Delacroix (1924) writes *«la pensée fait le langage en se faisant par le langage. [...] Le langage est la première science, étant le premier instrument dont notre esprit se sert pour construire l'univers mental. »*

Harley (2013) suggests that understanding syntax allows us to better understand cognitive processes: *"In particular, the syntax that governs the language of thought may be very similar or identical to that of external language. Studying syntax may therefore provide a window onto fundamental cognitive processes."*

Consequently, uncontrolled natural language is ambiguous and unsuitable for use in domains where ambiguity may be dangerous such as the aviation industry, but on the other hand, it represents an intricate part of our cognitive processes and its rules must not be excluded.

We therefore hypothesize that the exposure to natural language for both native and non-native speakers influences the way people will understand a certain text and respond to it efficiently. In other words, an unambiguous text written in a natural language construction would be more easily understood on a cognitive and behavioral levels than a CL that is coded, overly simplified, and syntactically non-conforming to natural language. This is due to speakers being more exposed to a certain natural language and its constructions in the usage of this language in their everyday life. Or so we hypothesize.

The idea is not to eliminate controlled language altogether. For then, without rules, common linguistic ambiguities would be very easy to come by. The real question is: what is the right balance? Research in the field affirms that "simplification" is the right way to proceed to achieve better comprehension. Readability, text-complexity, text-cohesion research have all focused on the process of simplification/controllability/structuration (DuBay (2004), McNamara et al. (2010), Temnikova (2012), Van Oosten et al. (2010), among others) without necessarily questioning (or at least behaviorally testing) the simplification rules. Ryan (2018) writes that *"language control raises some formal linguistic issues, particularly the question of whether sweeping restrictions on expression for the sake of simplicity and concision may unintentionally impede communication."*

In some CNLs, simplification reduces the sentential elements to the basic essentials, and diminished the scope and complexity to the detriment of information loss.

The following is an example of PLAIN English CL taken from their website: [24]

A. High-quality learning environments are a necessary precondition for facilitation and enhancement of the on-going learning process.

B. Children need good schools if they are to learn properly.

According to the Plain English approach, these two sentences are synonymous, with sentence A being more difficult than sentence B. While that might very well be the case, sentence B does not say everything sentence A intends to say. The semantic field has been highly restricted. For instance,

24. Plain English. Plain English Campaign. [online] Available at: http://www.plainenglish.co.uk/ [Accessed 10 Dec. 2018].

"learning environments" are not strictly limited to "schools", and not universities or home-schooling, tutoring etc. "Facilitation and enhancement" are not accurately summarized by "learning properly". The idea of an "on-going process" has been completely eliminated. In our opinion, those two sentences are in very little ways synonymous. Simplification has led to a substantial change/reduction of meaning that unless it specifically intended to do so, has failed to accurately "simplify". Adriaens and Schreurs (1992), when working on a controlled English grammar checker, wrote in reference to three CNL authoring manuals (AECMA Simplified English, Ericsson English and IBM English, all three derived from ILSAM[25]): *" [...] the linguistic foundation of these manuals is at times very weak: oversimplification often leads to linguistic inaccuracies; frequently linguistic structures are not covered, the instructions are at times vague and ambiguous, and often the rules disregard linguistic reality".*

Codifying, simplifying, and abridging languages, whether by using syntactic or other forms of ellipses could make a language difficult to assess for a lay speaker of a given language. That is to say, a codified language might require prior training and possibly more effort on the end user regarding direct, easy, and intuitive comprehension; a process that might well be exacerbated in situations of stress or danger. Riley et al. (1999) writes that *"in the early days of personal computing, users had to learn some basic computer science concepts and an obscure command language with only passing resemblance to natural language. Because the command language was highly codified and condensed, the user had to commit apparently arbitrary syntax rules to memory, and if the user made an error, such as reversing the source and destination drives in the "copy" command, the results could be catastrophic."*

Therefore, the usefulness and usability of an acquired codified (in the sense of requiring prior learning) controlled language must be put to the test and undergo behavioral scrutiny. How much control was needed to actually achieve better comprehension, and what are the limits that could potentially render this control or oversimplification unsatisfactory/counter-productive?

This brings us to our core question on which all of the experimentations in Chapter 5 and Chapter 6 will be based:

> Might a more natural syntax help pave the way for better pilot comprehension and faster reaction times?

25. ILSAM: International Language of Service and Maintenance

4.2 Approach with Regards to Our Corpus

Our experimentation plan is to go against the tide of common comprehension-oriented CNL construction, in the sense that we will not be taking natural language and simplifying it, but rather taking a highly controlled codified language, the Airbus Cockpit Controlled Language, (therefore theoretically most simple) and "complexifying" it (bring it closer to natural language: theoretically most complex) in order to make it more accessible (less training).

In other words, we want to bring it back to a more natural state: give it a more natural language structure, syntactically and otherwise. Thus, we are going backwards, towards natural language, while making sure not to fall in the trap of ambiguity.

Simplification does not necessarily have to start from an unsimplified text; such is the case with more formal representation languages such as Attempto Controlled English (Fuchs & Schwitter 1996) that start with basic logical relations and gradually add complexity. In these cases, simplification is applied to force writers to write in a simple manner from the start.

In our case, we will go from an initially codified corpus (evoked in Chapter 2) to a more natural one, by using research that has been done on CNLs and evaluations in cognitive sciences and test, bit by bit, how we can add sentential elements that would make the language closer to natural language structure of English. At the same time, by adding a less elliptical sentence structure we would be limiting the different possible interpretations, therefore avoiding, as much as possible, elliptical ambiguities.

4.2.1 Naturality Scale

As we mentioned in Section 3.1.3.5 before, a CNL can vary in its dimension of naturalness on the PENS classification scheme (precision, expressiveness, naturalness, and simplicity, Kuhn (2014)) from N3 to N5, with N3 describing languages that have some natural and unnatural elements, but that are nevertheless understood by speakers of the language to a substantial degree; and N5 on the other end of the scale, describing languages that contain sentences with natural text flow. N1 and N2 languages are not considered CNLs as they are not natural enough to be understood easily and intuitively by native speakers.

We propose a "Naturality scale" (cf. Figure 21) on which CNLs (only the yellow part of the scale could hold a CNL) would be placed on a continuum ranging from "Least naturalistic" or very coded to "Most naturalistic" or natural language in its theoretical state.

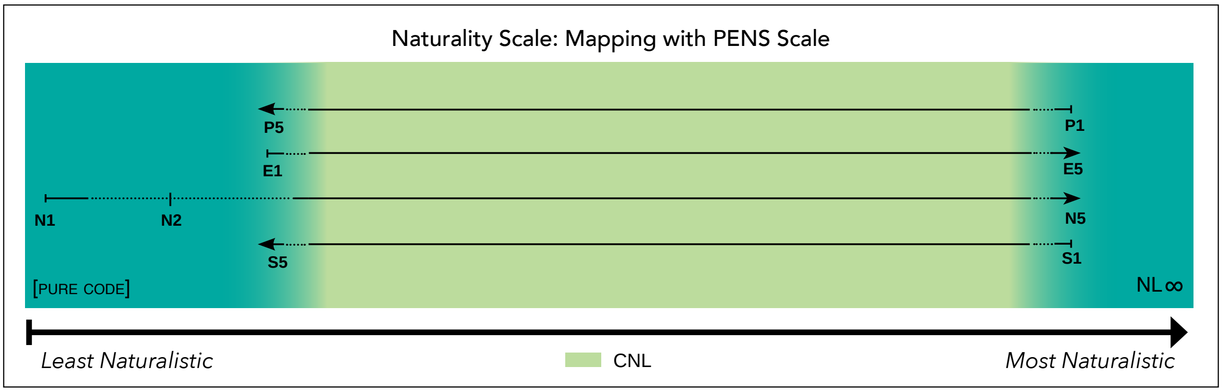


Figure 21. Naturality Scale Mapping with PENS

In other words, the Naturality component could be roughly defined as the naturalness levels present in a language on a boundless continuum ranging from pure code to natural language. In this theory, natural language will always be theoretically unattainable ∞. Language is almost always to some extent controlled. Whether it is the written word or the spoken word, context, audience, aim, social decorum, officially regulated language rules in INLs, Grice's speech maxims, and many other outside factors force the user of the language to control to a certain degree what language he or she produces at a certain period in time.

Therefore, language will not be divided into controlled and natural but should be placed on a naturality continuum with regards to all its aspects and the continuity of its gradations.

This differs from the PENS classification scheme (but does not necessarily exclude it), because here we consider that the "naturality" aspect is the most significant dimension and from which all other dimensions should follow suit. PENS' aim is to describe and give qualifications of CNLs and not rate them, which would fit right along the naturality continuum.

From this we argue that the 4 dimensions that make up PENS (precision, expressiveness, naturalness, and simplicity, displayed in 4 horizontal lines on the graph) could be concatenated and placed onto one dimension of naturality. For depending on whether a language is naturalistic or not and where it should be placed on the naturality continuum, we would be able to extricate whether or not a language is precise (from many interpretations to extremely precise), expressive (from no quantification to able to express everything), or simple (virtually indescribable rules (NL)) to described in one page).

Most importantly the classification of CNLs on the naturality continuum should be fluid because being subsets of natural language means that their application could hardly and fractionally be formalized in a clearly defined range. Additionally, as mentioned in Section 3.1.2, CNLs as any language tend to evolve with time and with the need and application we have for them, and the linguistic norms that make them.

As we can see on the naturality scale (cf. Figure 22), we plotted the Airbus Controlled Language using the PENS classification scheme P2 E4 N3 S2. The Precision, expressiveness, simplicity and naturalness are all plotted on the naturality continuum from least naturalistic [pure code]] to natural language [NL ∞].

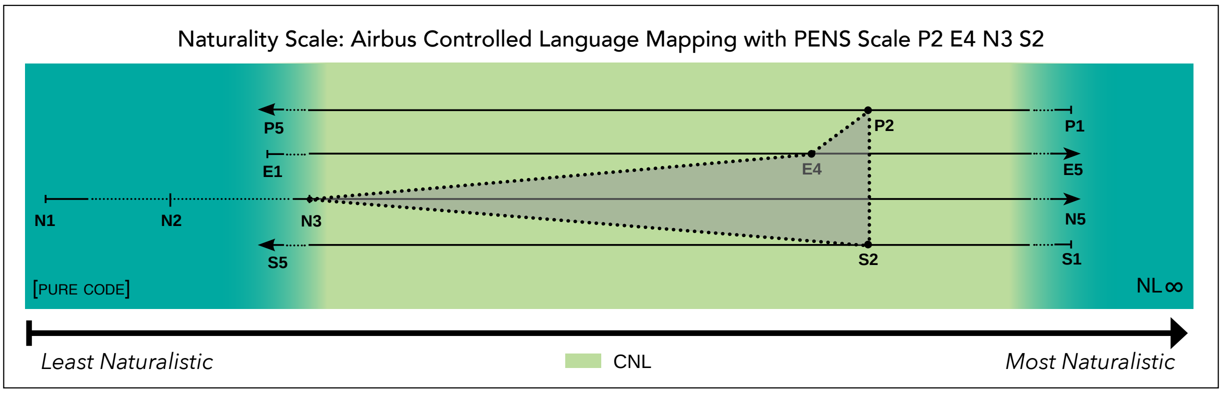


Figure 22. Naturality Scale: Airbus Controlled Language Mapping with PENS

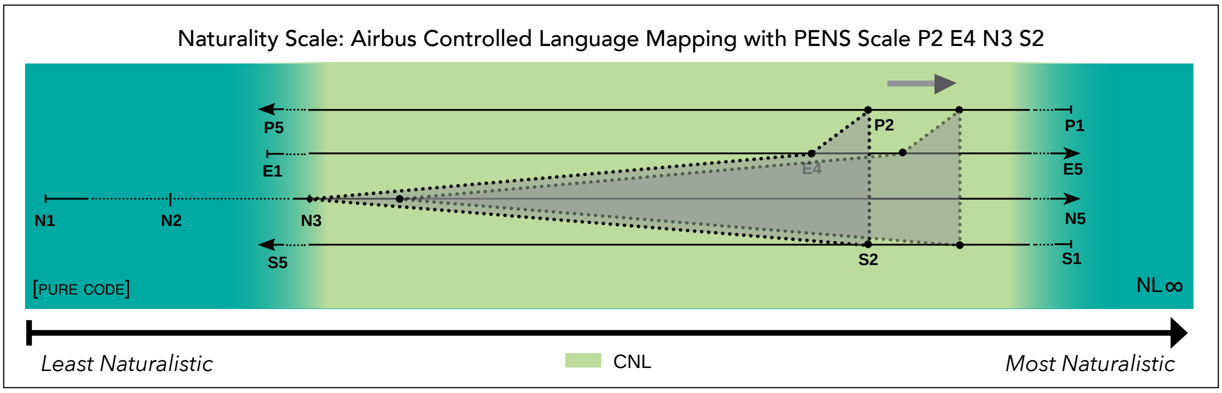


Figure 23. Naturality Scale: Airbus Controlled Language Mapping with PENS and potential shift towards natural language

The Airbus Cockpit Controlled Language forms the shape we see in the middle of the scale. What is interesting and novel about this representation is the fluidity with which a language can travel on the continuum. Considering the fluidity of languages, if a CNL becomes more or less naturalistic (as a result of an evaluation) and thus shifts on the continuum, the entire mapped CNL shape will shift accordingly since the foundation of this scale is the naturality continuum, the x - axis (cf. Figure 23).

Additionally, this scale also gives us a visual dimension of a CNL's naturality and could form grounds for comparison of different controlled languages that differ in their naturality levels and in their naturality evolution in time. Therefore, the Naturality scale is essentially a mapping of the PENS classification and criteria on a naturality based continuum.

In other words, if a controlled language has become more natural as a result of norming and normalization, or of behavioral or other forms of experimentation (for example, if it was shown that there is a need to reduce the use of syntactical ellipses), it will shift on the naturality scale towards the most naturalistic side of the scale (right side), i.e. it becomes more natural.

What this means is that when a language becomes more naturalistic it necessarily also shifts away from all its previous PENS dimensions. In this case (Figure 22 and Figure 23), the new language becomes less simple to explain with traditional language rules (Simplicity dimension shifts from S2 to S1.5, the more natural a language is the less simple it is to explain). It will also be able to express more (Expressiveness dimension shifts from E4 to E4.5) etc.

4.2.2 Towards a New More Natural Controlled Language (MNL)

Figure 24 is an infographic that summarizes our core question and hypotheses. The figure opposes two poles, the natural and the controlled. On one end, in orange we have the natural language which is more naturalistic, and on the other end in purple we have the controlled language which is less naturalistic. If we consider natural language first without any control, we fall into ambiguity, misuse and misunderstanding, which is unsuitable for a human operator and could lead to erroneous actions. Therefore, some control and simplification are necessary to avoid ambiguity. When we do that, we create standardized rules that limit ambiguity and form a controlled language.

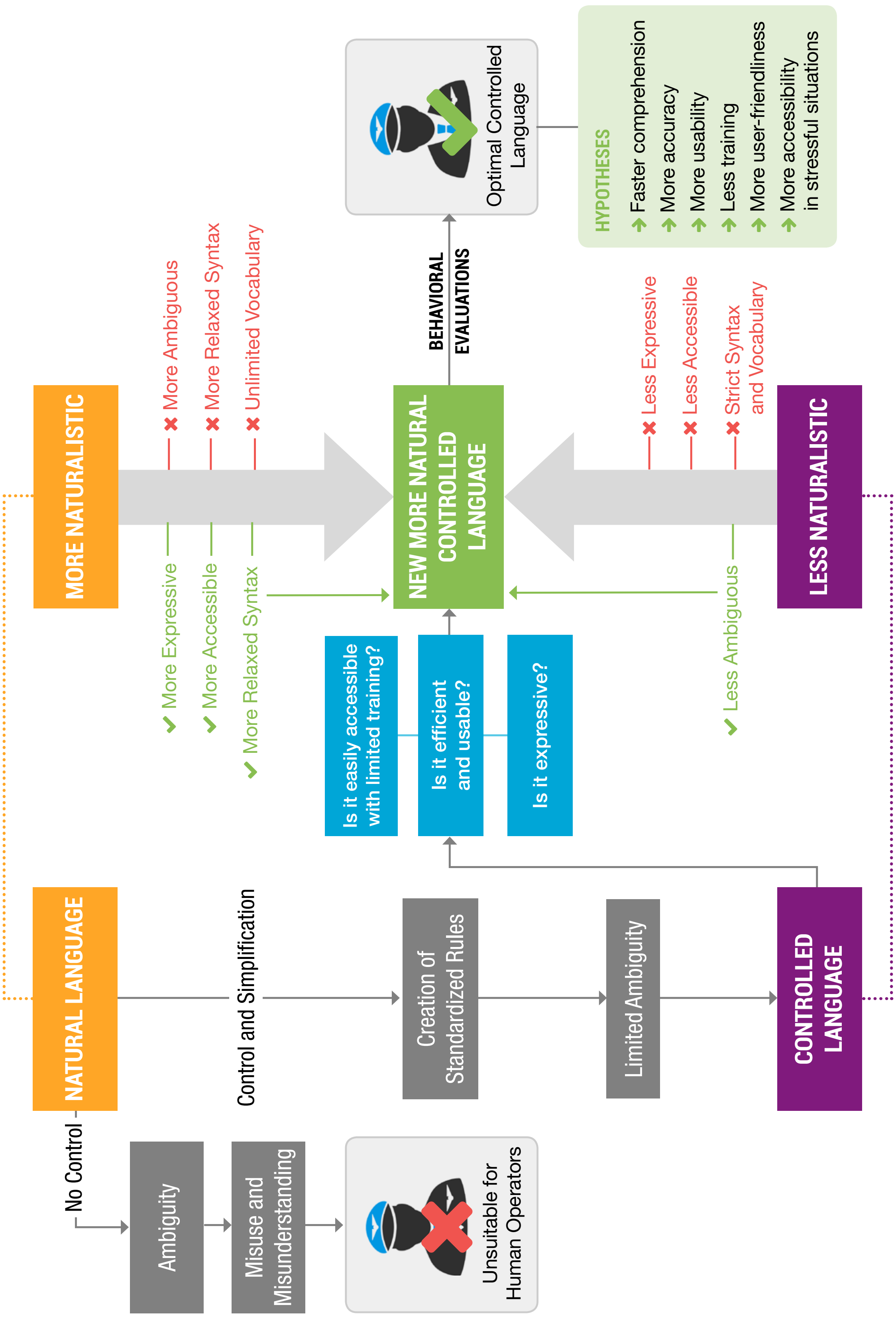


Figure 24. Core Question and Hypothesis Figure

However, when we create a controlled language from standardized rules, we need to know whether that language:

- is expressive enough (are we able to say everything we need to say, with the right words?),
- is it efficient and usable (are we able to efficiently communicate certain information in a clear and coherent manner so it is effortlessly employed?),
- is this language easily accessible with limited training (are we able to teach the language easily, and is this language easily learnt because it has more or less familiar structures, and is it intuitive enough so it does not require memorizing new codes?).

In order to start giving answers to these questions we need to find common ground between the tightly controlled but less naturalistic and the non-controlled but more naturalistic as both poles have positive attributes to offer:

- The more naturalistic, while more ambiguous because of a more relaxed syntax and unlimited vocabulary, is more expressive and more accessible since it is closer to natural language which we use in everyday life.
- On the other hand, the less naturalistic pole, while more restrictive because it is less expressive and accessible, it is also less ambiguous on account of the restricted syntax and vocabulary.

Therefore, in order to create a new more natural (less coded, less restricted but more optimized controlled language) we need to take advantage of the positive attributes (in green) of both poles (the more naturalistic and the less naturalistic). We do that by going towards a more natural language, by de-codifying bit by bit the current coded language, and (paradoxically) complexifying it in order to make it more natural and more accessible with limited training.

After we propose a new more natural controlled language (MNL) which is essentially a more natural version of the syntax of the current more coded controlled language (MCL), we use behavioral methods and tools in order to empirically evaluate its efficiency for comprehension and performance. This new language that benefits from the positive attributes of both poles is hypothesized to display this empirical efficiency by achieving faster and more accurate comprehension, and by being a more usable language, one that is available with less training. This would make it more accessible and user friendly, especially in more trying circumstances.

The following chapters will introduce the experimentation that we conducted for the purpose of testing the structures and syntax of the Airbus Controlled Language that pilots currently use in the cockpits to navigate and operate the planes against a more naturalistic (in syntax and lexicon) controlled language. Empirical results will be presented and analyzed, and conclusions provided.

II

BEHAVIORAL METHODS: EVALUATION AND ANALYSIS

5

Effects of a More Natural Language on Comprehension in Informational Statements

5.1 Introduction

For the purposes of this first experiment, we sought to evaluate passive comprehension. That is, we put in place an experiment in which motor skill reactions were not the main focus. We did not use injunctions and expected an action to be done (as will be the case in the following chapter), rather we used congruency tasks.

In order to be able to use congruency tasks to evaluate comprehension we had to limit ourselves to the use of the “information category” in our corpus, and more particularly, the constative messages informing pilots of the availability of a certain function such as “Galleys extraction available in Flight” or “Expect high cabin rate”. Therefore, we will start by testing the hypothesis on statements in initial coded language vs. a more natural form by evaluating reaction times and accuracy of comprehension (cf. Figure 25).

`Cabin Altitude Regulated to 7000 FT` *Current coded format*

vs.

`The cabin altitude is regulated to 7000 FT` *Proposed more natural format*

Figure 25. Current CL Example of an Information Statement and a Proposed More Natural Format

In order to empirically test the hypothesis, we needed to use psycholinguistic evaluation tools and make use of the syntactical structures that are available in the Airbus controlled language.

5.2 Experimental Design

5.2.1 Information Category Description and Proposed Task

Even though there are different types of messages included in the information category such as titles and expectations etc. (cf. Section 3.2.3), we had to exclude these different types and limit ourselves to the messages which inform users of a certain availability (messages included the word “avail” (cf. Figure 26)) because these were the only messages that allowed us to test the effective real time comprehension of the messages in a tightly controlled experimental paradigm.

```
□ LAV & GALLEYS EXTRACT AVAIL IN FLT
□ L TK 17000 KG MIN AVAIL
```

Figure 26. Information Category Examples with Availability

More specifically, the reason “availability” messages were the only viable candidates was because we could test comprehension in a psycholinguistic congruency task that is represented in Figure 27.

As we can see, the task consisted of the participants reading a text written in either the MCL syntax or the MNL syntax. The text then disappears and an image appears, an image which could be congruent with the previously read text or incongruent. If the text for example says “bus stop available” and the image shows a bus stop then the participant has to press “yes” on the controller to indicate congruency, and if for instance the image shows an image of a car then the participant should press on “no” to indicate that the image is incongruent with the text. Response times and precision in both conditions of language were recorded.

We chose sentences that could show an accurate visual description of a situation or scene. The nature of the congruency task put in place limits the messages that we could use. For instance, if participants were presented with a message written in MCL or MNL followed by an image for which they had to press on the controller to say whether “yes” the image corresponds to the message previously read (congruent) or "no" the image does not correspond to the message (incongruent), the message should be describing the availability of an object in the image. Had we used informative sentences such as "expect high cabin pressure rate" or "risk of reduced cabin airflow", we would have no clear way of testing in real time the comprehension of such a sentence and to collect accurate reaction times; and it would have been impossible for the participant to evaluate the congruency of those messages to any image.

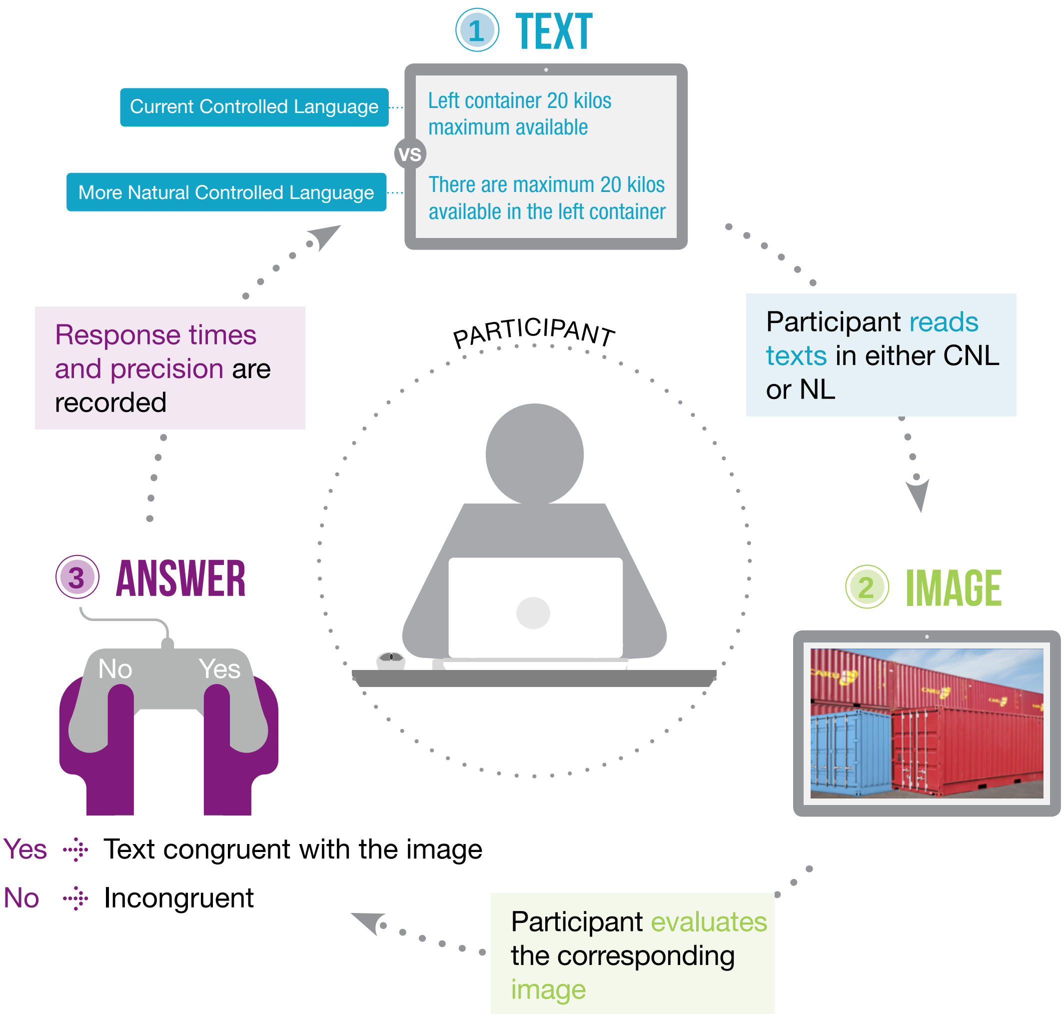


Figure 27. Representation of Task Performance

Moreover, there were different kinds of messages that contained the word "avail" such as with negation (not avail) or "avail" followed by a condition (avail if) or "avail" followed by a list, but we limited the target stimuli to the messages containing simple availability (negatives, conditions, etc. were not included) so as not to create ambiguity in the congruency task. A table showing some examples is available in the Annex C.

5.2.2 Construction of the Messages: Syntactic Difficulty

The messages constructed for this experiment will be tested on naïve participants, (both native and non-native speakers of English) in order to attest to their usability and confirm or deny our hypothesis on human comprehension in general and to avoid expert bias. This is a critical step preceding eventual future testing on end users (pilots) using the aeronautical language corpus. Therefore, in this first experiment, we will not be using the exact aeronautical corpus terms as it will be difficult for naïve participants to understand technical jargon. The stimuli will be made up of everyday life images and sentences such as "parking spot is available" that emulate the syntax and intentions of our original corpus statements.

After studying the various syntactical and semantical relationships in the corpus, we established 6 difficulty categories that represented the syntactical structures of the information availability statements. They went from 1 easiest structure (noun + noun + available) to 6 most difficult (noun + noun + noun + available + in + noun) as length has been proven to be an effective and efficient index of syntactic difficulty (cf. Section Figure 28). According to Szmrecsanyi (2004), sentence length (or a version of the Flesch-Kincaid tests) are as good a means of testing syntactic text complexity as counting syntactic nodes in a sentence. Szmrecsanyi reports comparing three methods of measuring syntactic complexity node counts, word counts, and 'Index of Syntactic Complexity' (which takes into consideration the number of nouns, verbs, subordinating conjunctions, and pronouns). She concludes that the three measures are near perfect proxies since they significantly correlate and can be used interchangeably. Therefore, we can feel safe to use the measure that is most economical to apply.

After the difficulty conditions were set, we had to come up with 6 random sentences for each condition and naturally an image was needed to be congruent with the sentences. In order to use these sentences with non-pilot participants we needed to remove all original corpus abbreviations and color coding, as well as the full upper-case letter font, except at the beginning of every message (we kept the upper-case letter in the first word of messages to mark the beginning of a new message in every stimulus). No punctuation was used and messages were written in AI-B612[26], a font developed by Airbus and PolarSys that facilitates on-screen reading and letter identification.

26. PolarSys. B612 - The PolarSys Font. [online] Available at: https://www.polarsys.org/proposals/b612-polarsys-font [Accessed 9 Dec. 2018].

Non-Aviation Messages Parallel to ECAM Structure Messages	Syntax (Difficulty 1-6)
Chalk board available	1- Noun + Noun + Avail
Mobile car holder available	2- Noun + Noun + Noun + Avail
Emergency exit available in building	3- Noun + Noun + Avail + In + Noun
Office writing supplies available in catalogue	4- Noun + Noun + Noun + Avail + In + Noun
Left container 20 kilos maximum available	5- Adj + Noun + Num + Noun + Noun + Avail
Yellow hall 2 movie posters minimum available	6- Adj + Noun + Num + Noun + Noun + Noun + Avail

Figure 28. Examples of 6 Difficulty Conditions

5.2.3 Construction of the Messages: MNL Messages from MCL Syntax

A critical step was also to decide how to write the MNL corresponding messages to MCL, or in other words how to naturalize the more coded language.

As we mentioned before, natural language is by definition ambiguous and has almost unlimited possibilities of expression. Therefore, we needed to determine the most appropriate linguistic form, standardize it to all our stimuli, and justify its use.

In the following example case, the original coded and abbreviated message is L TK 17000 KG MAX AVAIL which when decoded without abbreviations means "left tank 17000 kilograms maximum available". It was relatively easy to construct the MCL messages since we could keep the same structure and same words when possible, and find or construct an image that is congruent to its meaning. However, constructing the equivalent MNL messages was a little more complicated as we had several options; there was at least 4 different ways of writing the sentence in the previous example in a more natural language. See Annex D for other sentences and their different possibilities.

1. There are maximum 20 kilos available in the left container
2. There are 20 kilos maximum available in the left container
3. The left container has maximum 20 kilos available
4. The left container has 20 kilos maximum available

After careful consideration and in order not to multiply variables, we chose the first option for the MNL structure as the existential clause "there is/are" introduced by the expletive pronoun "there" + predicate "are" indicates the existence or the presence of something in a particular place or time, which in our experiment reinforced the idea of something available or not available in the target picture. The existential clause itself expresses a predicate of existence which sets the tone for the incoming noun phrase. While the second option also includes an existential clause, it was not deemed sufficiently plausible by English native speakers that we consulted. The existential clause introduced in the MNL structures also inverts the theme and rheme structure of the original MCL structure. The current controlled language uses the theme at the onset of the message "left container" followed by the rheme. One of the main differences between both languages is the addition of function words in the MNL stimuli. Leroy et al. (2010) affirms in a study about the effects of linguistic features and evaluation perspectives that *"complex noun phrases significantly increased perceived difficulty, while using more function words significantly decreased perceived difficulty. […] Laypersons judged sentences to be easier when they contained a higher proportion of function words. A high proportion of function words leads to a different cadence closer to spoken language. It may also help space out individual concepts in text to facilitate assimilation."*

Refer to initial and complete list of MCL and MNL messages proposed for the experiment in Annex E.

5.2.4 Image Base Construction

Once the messages were more or less set (there were additional changes along the way), a more challenging task consisted of finding/building/modifying images that corresponded to the messages at hand. It was a constant back and forth between non-copyrighted google images, Adobe Photoshop[27], Adobe Indesign[28], and the messages themselves. We will show 3 examples that illustrate the methodology undertaken. A sample list of images from each difficulty level and their corresponding messages are available in the Annex F.

Figure 29 shows an image of two containers which we altered to fit the message "Left container 20 kilos maximum available", like in Figure 30.

27. Adobe.com. Adobe Photoshop CC | Best photo, image, and design editing software. [online] Available at: https://www.adobe.com/products/photoshop.html [Accessed 9 Dec. 2018].

28. Adobe.com. Adobe InDesign CC | Desktop publishing software and online publisher. [online] Available at: https://www.adobe.com/products/indesign.html [Accessed 9 Dec. 2018].

Figure 29. Original Image: Containers

Figure 30. Altered Image: Containers

Figure 31 shows an image of a police officer interviewing a civilian, but since the message that is congruent with the image needed to have a nominal group of 3 nouns + the word “available” (difficulty level 2) the message became “A **crime scene officer** is available”. Figure 32 does not necessarily show a “crime scene” officer, as it could be interpreted as a police officer handing a parking ticket. The images needed to be unambiguous since we did not want to leave room for interpretation of congruency. Therefore, we photoshopped the image to make it suitable and congruent with the "crime scene officer available" message in Figure 32. As we can see a police yellow tape was added and a body traced in chalk to reinforce the idea of a "crime scene".

Figure 31. Original Image: Crime Scene

Figure 32. Altered Image: Crime Scene

Additionally, we had to make sure that there was a level of coherence in the difficulty of identification of the objects in the images themselves. The message that is congruent to Figure 33 is "School bus available". The image seemed too obvious as it was easy to spot the school bus and understand and identify congruency than in other images.

Figure 33. Original Image: School Bus

Figure 34. Altered Image: School Bus

That is, the object of the images in question, in this case a school bus, needed to be visible enough for participants to identify, but not so obviously in the foreground that it was too easily spotted. Therefore, in Figure 34 we photoshopped out one of the school buses in the foreground of the image so that it is not as glaring, and participants had to scan the image more thoroughly before deciding on congruency. Reconstruction of the background was necessary in this case, such as the trees and pavement.

Any problematic images or ones that were deemed particularly difficult to apprehend were eliminated in the pre-tests.

5.2.5 Integration in DMDX and experimental Protocol

DMDX[29] is a Win 32-based display system used in psychological laboratories to measure reaction times to visual and auditory stimuli. We used this software on a Dell Precision 3510 laptop in order to display the messages and images. For that we developed 6 scripts which consisted of 3 semi-randomized lists of stimuli for right handed participants and 3 for left-handed participants (same lists but the "yes" and "no" buttons were inverted for left handed participants).

Participants started with a practice session composed of a different set of 24 semi-randomized stimuli representative of the difficulty and language conditions, and the same image construction methodology as the target stimuli in the main lists. They had noise cancelling headphones and were set in a quiet room with no distractions.

Each list consisted of 48 target stimuli, split into 24 congruent stimuli (image congruent with the message, correct answer is a "yes") and 24 incongruent stimuli (image incongruent with the message, correct answer is a "no"). In addition to the practice list, each participant performed the task on one of the three lists.

An additional variable that we controlled was the reading time of on-screen messages. As we showed in Figure 27 the messages disappear to make way for the corresponding images. As messages were different in length, the allotted

29. U.arizona. DMDX. [online] Available at: http://www.u.arizona.edu/~kforster/dmdx/dmdx.htm [Accessed 9 Dec. 2018].

reading time was different depending on the number of words. MNL messages necessarily have more words than MCL messages. However, those words were only grammatical words such as "there is" or "a", or "the" etc. We decided to count only lexical words in order to calculate reading time. This choice might have inadvertently given a position of privilege to the MCL messages since MNL messages had more total words (grammatical and lexical) than the equivalent MCL messages yet they had the same reading time (same number of lexical words). We based ourselves on word per minute and reading time research to calculate the time the messages appeared on the screen. Trauzettel-Klosinski & Dietz (2012) found the average speed across 17 different languages to be 184 plus or minus 29 WPM and 228 plus or minus 30 for Latin alphabet languages like English. Ziefle (1998) showed that when proofreading, people read English at 200 WPM on paper and 180 on a screen.

Therefore, a reasonable value would have been 180 WPM but that proved to be too fast for participants to read in our pre-tests. The reason might be because these messages are not part of a bigger text but appear out of context preceded only by a 3000 ms. fixation cross in the middle of the screen. We decreased that value to 150 WPM, so that a message that has 3 lexical words would appear for 1.2 seconds (3 x 60/150) and a message that has 6 lexical words would appear for 2.8 seconds (6 x 60/150), etc.

Participants mentioned that time to read was still short but sufficient when one is paying attention and focused. We purposefully did not leave extra time for reading or re-reading because we wanted participants to respond as intuitively as possible to test for initial comprehension and reaction. Results might differ and could be mitigated or altered by having more time to read and interpret more thoroughly the messages written in one language condition or the other.

Participants had 5000 ms. to respond. In case of a non-answer the next stimulus appears and so on. Once the participant responds the image disappears and the next fixation cross appears.

5.2.6 Follow up Experiment

As the reading speed variable seemed to be somewhat inequitable in both language conditions (MNL having more words but same reading time) we decided to use the same experimental design and materials but change the speed variable (hereby referred to as Experiment 2 or second experiment). We decreased reading speed to 120 WPM and counted all the words in messages (lexical and grammatical) so that the number of total words affected screen reading time. The more a message contained words the longer it stayed on screen.

5.2.7 Participants

Before beginning the experiment, participants filled out different forms: a general ethics and compliance consent form, a data sheet in which they specified their age, gender, dexterity, native language, English placement, knowledge of Airbus CL. Participants were not remunerated for their participation, but were offered an 8Gb USB key for which they signed a receipt. Refer to Annex G to see these forms.

All non-native English speakers also performed a quick English placement test online to determine their CEFR levels (Common European Framework of Reference for Languages)[30]. The levels range from A1 or breakthrough/beginner to C2 or Mastery/Proficiency. This will eventually help us determine whether English placement levels had an effect on our hypotheses.

72 participants took part in the first experiment (12 native speakers of English and 60 non-native speakers whose placement levels ranged from A1 to C2 in CEFR). The non-native speakers' languages included Arabic, Chinese, Dutch, French, German, Portuguese, Spanish, Serbian, and Indonesian, with the overwhelming majority being French (45 out of 60).

38 participants had no knowledge whatsoever of controlled languages. 16 claimed had beginner knowledge of the Airbus controlled language (Airbus employees having rarely worked with the language or its rules). 14 had a more intermediate knowledge of the language.

5 participants had expert knowledge of the language as it could be part of their daily task.

30 participants took part in the second experiment with only 1 native speaker of English and 28 non-native speakers, including 5 French native speakers, 19 native Arabic speakers, and the rest were distributed among

30. The CEFR Levels. [online] Common European Framework of Reference for Languages (CEFR). Available at: https://www.coe.int/en/web/common-european-framework-reference-languages/level-descriptions [Accessed 9 Dec. 2018].

Chinese, Bengali, Spanish, and Japanese speakers. Their English placement levels ranged from A2 to C2 in CEFR. Almost all of the participants had no or very little knowledge of the Airbus Controlled language, therefore it was not a variable that we studied in experiment 2.

5.3 Research Questions and Hypotheses

We will start by listing our hypotheses and research questions and then proceed to show the results and analysis.

Our metrics and dependent variables were reaction times in ms. for correct answers, and the number of errors for accuracy. Since experiment 1 and 2 have the same protocol, stimuli, task and hypotheses, with the exception of the speed variable evoked in the previous section, in order to facilitate the discussion and avoid repetition, we will analyze the results of both experiments for every hypothesis simultaneously, before discussing what those results mean for each of them.

Statistical tools and analysis in this section were conducted using the software R[31], a free software environment for statistical computing and graphics. We developed scripts to clean and manage the data (in addition to the pre-cleaning done with DMDX tools and Microsoft Excel) into appropriate R vectors, and to be able to use different statistical tests and methods that were useful for our analysis. The graphs were also generated based on our data from R scripts.

The list of independent variables that we will evaluate are:

- Language (MCL-MNL)
- Syntactic Difficulty (1 to 6) (Refer to Figure 28)
- Type (Congruents-Incongruents)

Extraneous and participant variables:

- English placement level (Basic Intermediate, Proficient, Mastery, Native)
- Familiarity with Airbus CL (None, Beginner, Intermediate, Expert)

31. The R Project for Statistical Computing. [online] Available at: https://www.r-project.org/ [Accessed 9 Dec. 2018].

Hypotheses:

- MNL messages produce shorter reaction times than MCL ones in different syntactic difficulty conditions.
- MNL messages produce less errors (are more accurate) than MCL ones in different syntactic difficulty conditions.

Research questions:

- Did the language factor play a different role for the different types of congruency responses regarding reaction times?
- Did the language factor play a different role for different levels of English placement (Basic Intermediate, Mastery, Natives) regarding reaction times?
- Did the language factor play a different role for different levels of familiarity with Airbus Controlled language regarding reaction times?

5.4 Results and Analysis

1. **MNL messages produce shorter reaction times than MCL ones in different syntactic difficulty conditions.**
   - **Experiment 1**

We started by verifying if our data follows the normal distribution in order to know what statistical tests to use. We performed a Shapiro-wilk normality test on our reaction times and it showed that the data is significantly non-normal ($p = 2.054e\text{-}05$) with abnormal skew, therefore we used non-parametric tests in order to test the main effect such as the Wilcox signed rank test because the same participants took part in both conditions.

We started by comparing the general effect regardless of difficulty for both language conditions using the Wilcox signed Rank test. There was a significant difference in the scores for MCL (Median=2030.317 ms.) and MNL (Median= 1944.163 ms.) conditions; $v = 1692$, $p = 0.0339$, effect size calculated with Pearson's coefficient $r = 0.24998$. With the hypothesis confirmed, we can conclude that the more natural language helped participants process the stimuli and provoked significantly faster reaction times than the more coded language format.

We then performed a linear regression model to ascertain the influence of the syntactic difficulty condition in both languages.

A simple linear regression was calculated to predict the reaction times of the MCL responses based on the 6 difficulty conditions. A significant regression equation was found ($F(1,1500)=9.211$, $p<0.002447$), with an R2 of 0.006103. Participants' predicted reaction times is equal to 1873.77 + 42.55 ms. for every additional difficulty condition. Therefore, reaction time increased 42.55 ms. for each additional difficulty condition.

A simple linear regression was also calculated to predict the reaction times of the MNL responses based on the 6 difficulty conditions. A significant regression equation was found ($F(1,1450)=12.68$, $p<0.0003822$), with an R2 of 0.008667. Participants' predicted reaction times is equal to 1801.64 + 47.81 ms. for every additional difficulty condition. Therefore, reaction time increased 47.81 ms. for each additional difficulty condition.

Figure 35 is the graph that plots those two linear regression models for both languages in the 6 difficulty conditions. As we can see there is no interaction between the two languages (lines are parallel and do not intersect) but reaction times get slower when difficulty increases in both languages which confirms that syntactic difficulty based on length is a valid measure. With the hypothesis confirmed, we can also conclude that MNL messages produced consistently faster reaction times than MCL messages in all difficulty conditions.

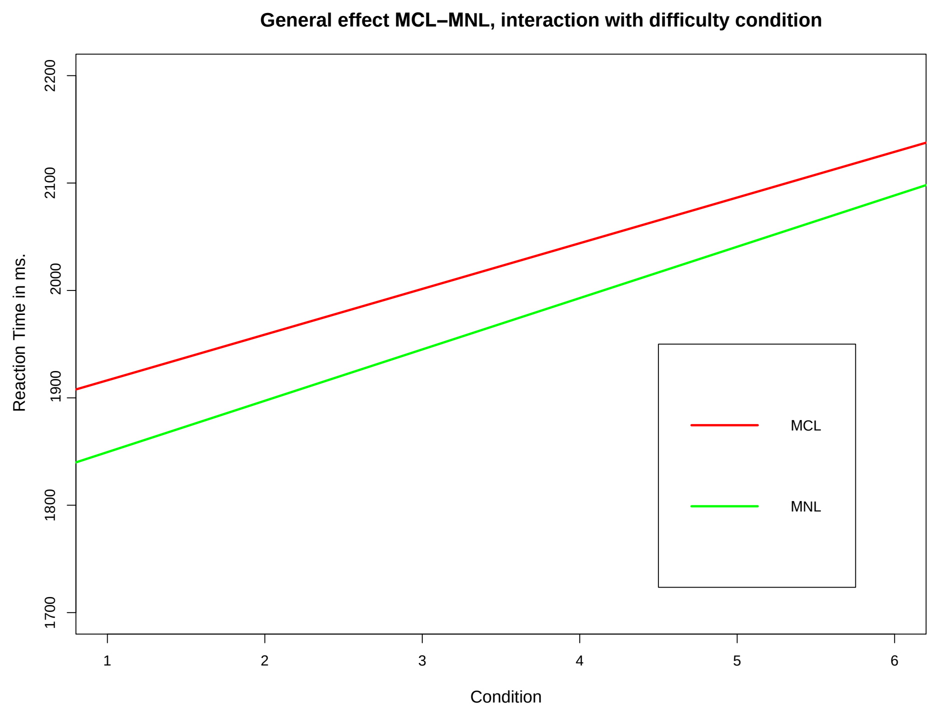


Figure 35. Linear Regression Models for MNL and MCL in the 6 Difficulty Conditions

- **Experiment 2**

The same analysis was done for experiment 2 results. We performed a Shapiro-wilk normality test on reaction times and it showed that the data is significantly non-normal ($p = 0.001297$) with abnormal skew, therefore we again used non-parametric tests.

We started by comparing the general effect regardless of difficulty for both language conditions using the Wilcox signed Rank test. There was no significant difference in the scores for MCL (Median = 1935.758 ms.) and MNL (Median = 1917. ms.) conditions; $v = 8182$, $p = 0.957$. We can conclude that while the more natural language median reaction time is less than the MCL median reaction time, this difference is not statistically significant. This differs from our results in the first experiment where the difference was significant. This is due to the added variable of time in the presentation of the stimuli. We can conclude that speed plays a role in the comprehension linked to language conditions and should be the object of further investigation.

As there was no significant difference in medians in the main effect of experiment 2's MNL and MCL conditions, we cannot perform linear regressions for the different difficulty conditions.

2. **MNL messages produce less errors (are more accurate) than MCL ones in different syntactic difficulty conditions.**

- **Experiment 1**

Accuracy was calculated using the average number of errors. Therefore, we started by comparing the general effect of accuracy regardless of difficulty for both language conditions using the Wilcox signed Rank test. There was no significant difference in the number of errors by subject produced in the MCL (Mean = 2.46 errors) and MNL (Mean = 2.9 errors) conditions; $v = 549$, $p = 0.07121$. There is no true difference in means of the two conditions of language regarding accuracy, therefore the hypothesis is not confirmed. While MNL proved to be superior to MCL with respect to reaction times, both languages performed equally with regards to making errors. We could interpret this by proposing that the difference in the syntax of the two languages was not different enough (a lot of the stimuli had only one or two grammatical articles added to them) to cause one language to have better performance with respect to errors, but those subtleties were manifested in the reaction times instead which stand to be more adequate measures of early/ initial comprehension.

Figure 36 is a histogram plot of the errors made in the different conditions of difficulty for both languages. As we can see the number of errors in both languages is not consistent across different difficulty conditions, but there is a tendency for both languages to have more and more mistakes as difficulty increases. The advance that the MCL has over the MNL in the easy difficulty conditions (probably due to having less words to read and the same time as MNL stimuli with more words to read) disappears the harder the stimuli get with the exception of mid-way difficulty level 4.

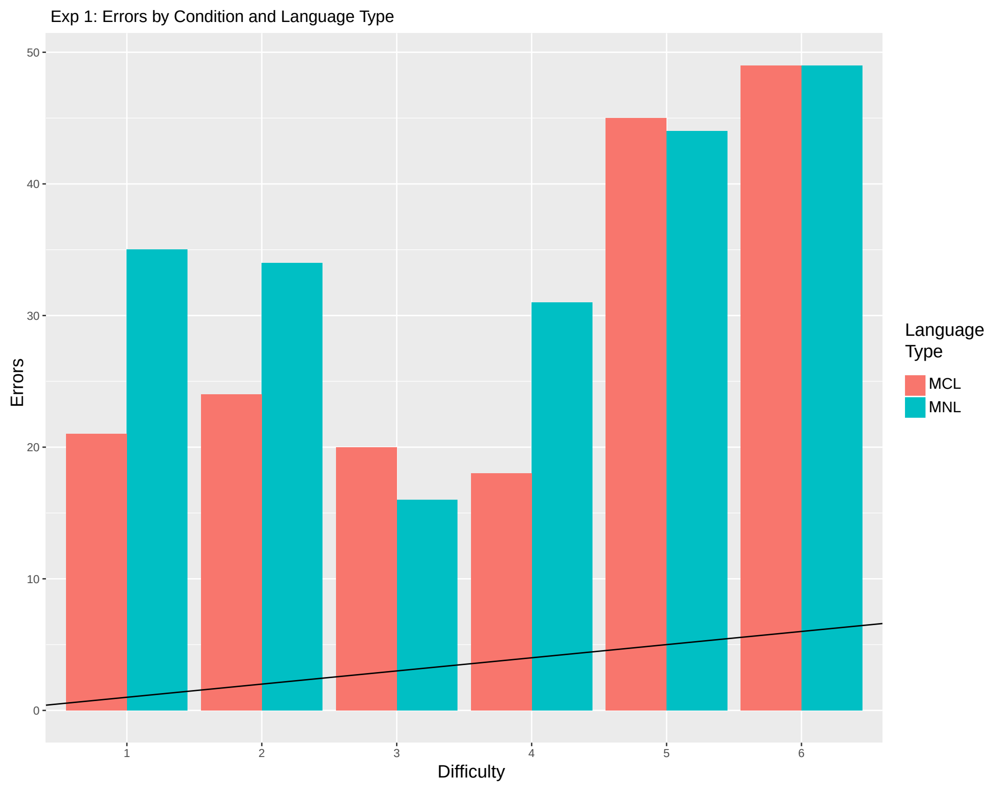


Figure 36. Histogram of Errors in MNL and MCL in the 6 Difficulty Conditions in Experiment 1

- **Experiment 2**

Accuracy was calculated using the average number of errors. Therefore, we started by comparing the general effect of accuracy regardless of difficulty for both language conditions using the Wilcox signed Rank test. There was no significant difference in the number of errors by subject produced in the MCL (Mean = 2.457 errors) and MNL (Mean = 2.448 errors) conditions; v = 176.5, p = 0.9794. There is no true difference in means of the two conditions of language regarding accuracy, therefore the hypothesis is not confirmed. In experiment 2, reaction times as well as accuracy fail to give an advantage to any of the two language conditions. We again argue that having more time to read, interpret, and respond flattened the discrepancy between the two language conditions because the task was not sufficiently hard and the corresponding stimuli in both language conditions did not differ greatly.

Figure 37 is a histogram plot of the errors made in the different conditions of difficulty for both languages. As we can see the number of errors in both languages is not consistent across different difficulty conditions, but there is a tendency of both languages having more and more mistakes as difficulty increases. The advance that the MCL has over the MNL in the easy difficulty conditions (probably due to having less words to read and the same time as MNL stimuli with more words to read) disappears the harder the stimuli get. As we can see, apart from conditions of difficulty 1 and 4, MNL had less errors on average than the MCL conditions. However, those observed differences were not statistically significant.

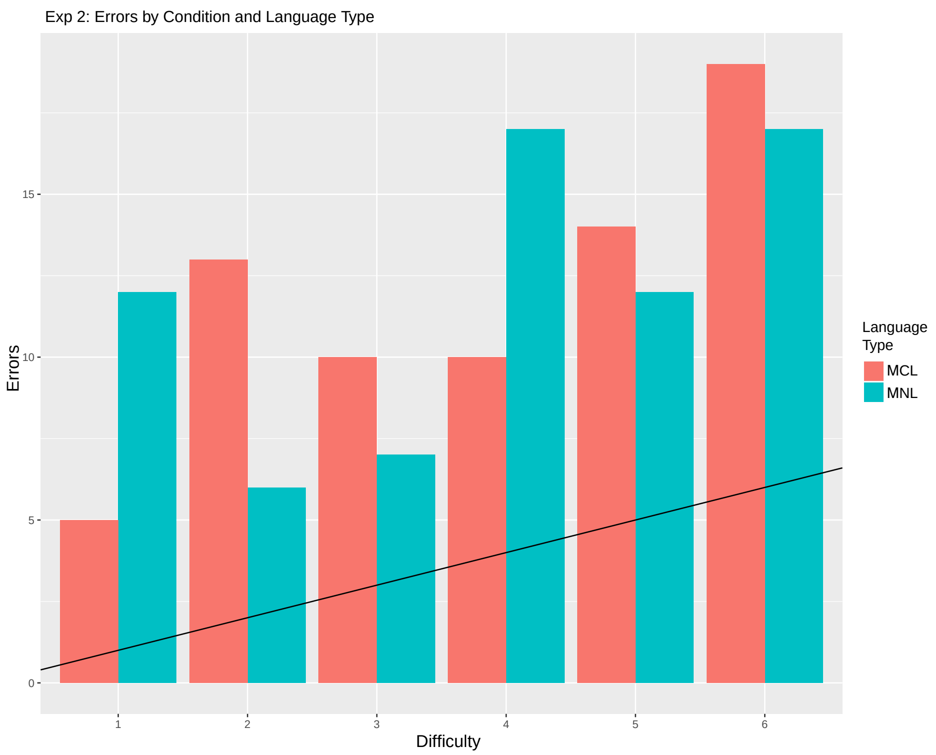


Figure 37. Histogram of Errors in MNL and MCL in the 6 Difficulty Conditions in Experiment 2

3. **Did the language factor play a different role for the different types of congruency responses regarding reaction times?**
    - **Experiment 1**

It was important to us to verify whether there was an effect of congruent stimuli versus incongruent stimuli (to the corresponding image) since congruent stimuli were deemed easier targets than incongruent ones, therefore understanding incongruent stimuli constitutes an extra difficulty condition in and of itself. To illustrate this with a concrete example: An image that shows an empty parking lot with a message that reads "Parking is available" is easier to interpret as a "yes congruent" than an image showing a desk lamp with a message that reads "Ceiling lamp is available" as a "no, incongruent". Confusion might arise from the presence of a lamp in the picture but which is not a ceiling lamp. Most incongruent images were purposefully chosen to include a little forced ambiguity, or an extra "trick" where the participant had to verify thoroughly the image before responding.

Therefore, we started by comparing the general effect of reaction times regardless of difficulty for congruent stimuli in both language conditions using the Wilcox signed Rank test. There was no significant difference in reaction times of the congruent stimuli produced in the MCL (Median = 1888.502 ms.) and MNL (Median = 1879.167 ms.) conditions; v = 1468, p = 0.3875.

However, when performing the same test for the incongruent stimuli we found a significant difference in the MCL (Median = 2241.473ms) and the MNL (Median = 1927.541ms.) conditions; v = 1475, p = 0.0308.

As we can see from Table 2 the difference between medians in the incongruent condition is far superior than the congruent one and is statistically significant. We attribute this difference to the added difficulty in the interpretation of the incongruent stimuli, and we conclude that the MNL syntax helps process information faster than the MCL condition as the difficulty in the task and stimuli increase.

MCL Congruent	MNL Congruent	Difference	MCL Incongruent	MNL Incongruent	Difference
1888.502	1879.167	9.335	2241.473	1927.541	+313.932

Table 2. Medians in ms. of MCL and MNL Reaction Times in Congruent and Incongruent Stimuli in Experiment 1

- **Experiment 2**

The same analysis was conducted for experiment 2 data. We started by comparing the general effect of reaction times regardless of difficulty for congruent stimuli in both language conditions using the Wilcox signed Rank test. There was no significant difference in reaction times of the congruent stimuli produced in the MCL (Median = 1847.912 ms.) and MNL (Median= 1913.910 ms.) conditions; v = 7102, p = 0.3108.

However, when performing the same test for the incongruent stimuli we found a significant difference in the MCL (Median = 2164.427 ms.) and the MNL (Median = 1892.255ms.) conditions; v = 6853, p= 0.0262.

As we can see from Table 3 the difference between medians of MNL and MCL in the incongruent condition is superior than the congruent one and is statistically significant. We can even observe that in the congruent condition the median of the MCL is 65.9ms. shorter than that of the MNL. We again attribute this difference in the MNL's favor to the added difficulty in the interpretation of the incongruent stimuli. We conclude that the MNL syntax helps process information faster than the MCL condition since the difficulty in the task and stimuli increases, even with the addition of more reading time in the stimuli that was introduced in experiment 2.

MCL Congruent	MNL Congruent	Difference	MCL Incongruent	MNL Incongruent	Difference
1847.912	1913.910	+65.998	2164.427	1892.255	272.172

Table 3. Medians in ms. of MCL and MNL in Congruent and Incongruent Stimuli in Experiment 2

4. **Did the language factor play a different role for different levels of English placement (Basic Intermediate, Mastery, Natives) regarding reaction times?**

- **Experiment 1**

We grouped the English placement levels into 3 categories. "Basic intermediate" regroups participants that were placed from levels A2 to C1, "Mastery" has participants that were placed in C2 level and "Natives" are the native English speaker participants.

We did a series of t-tests (as reaction times for those sub-groups were not significantly non-normal so we could use a parametric test) to compare the two different language conditions in each of the English placement groups.

- For basic intermediate level, there was a significant difference in the scores for MCL (Mean = 2246.322 ms.) and MNL (Mean = 2144.104 ms.) conditions; $t = 2.5416$, $p = 0.01644$.
- For mastery level, there was no significant difference in the scores for MCL (Mean=1956.563ms.) and MNL (Mean= 1954.745ms.) conditions; $t = 0.034395$, $p = 0.9728$.
- For native level, there was no significant difference in the scores for MCL (Mean = 1690.904 ms.) and MNL (Mean = 1588.062 ms.) conditions; $t = 1.8301$, $p = 0.09444$.

A summary of means and p-values is available in Table 4.

	Lower Intermediate	Mastery	Natives
MCL mean	2246.322	1956.563	1690.904
MNL mean	2144.104	1954.745	1588.062
Difference	+102.218	+1.818	+102.842
P value	0.01644	0.9728	0.09444

Table 4. Means of Reaction Times for Different English Placement Groups for MNL and MCL

As we can see the only significant result is the basic intermediate level. We could conclude that MNL helps comprehension for the weaker levels of English levels as reaction times are significantly shorter for that group. While the native group does not show statistical significance, most probably because the group is made up of 12 participants only, it is interesting to note the difference in the average of the MNL and MCL which is equal to the difference for lower intermediates. Native speakers often mentioned that they preferred the more natural language, and this is also apparent in their results. In the next experiment, we made sure to have a bigger sample of native speakers to be able to accurately test and analyze the results related to that sample.

A simple linear regression was also calculated to predict the reaction times of the MCL responses based on the 3 English placement levels. A significant regression equation was found (F(2,432) = 21.83, p = 9.275e-10), with an R2 of 0.0918. Participants' predicted reaction times is equal to 2221.92 – 280.14 ms. for every English placement level gained. Therefore, reaction time decreased 280.14 ms. for every English placement level gained (cf. Figure 38).

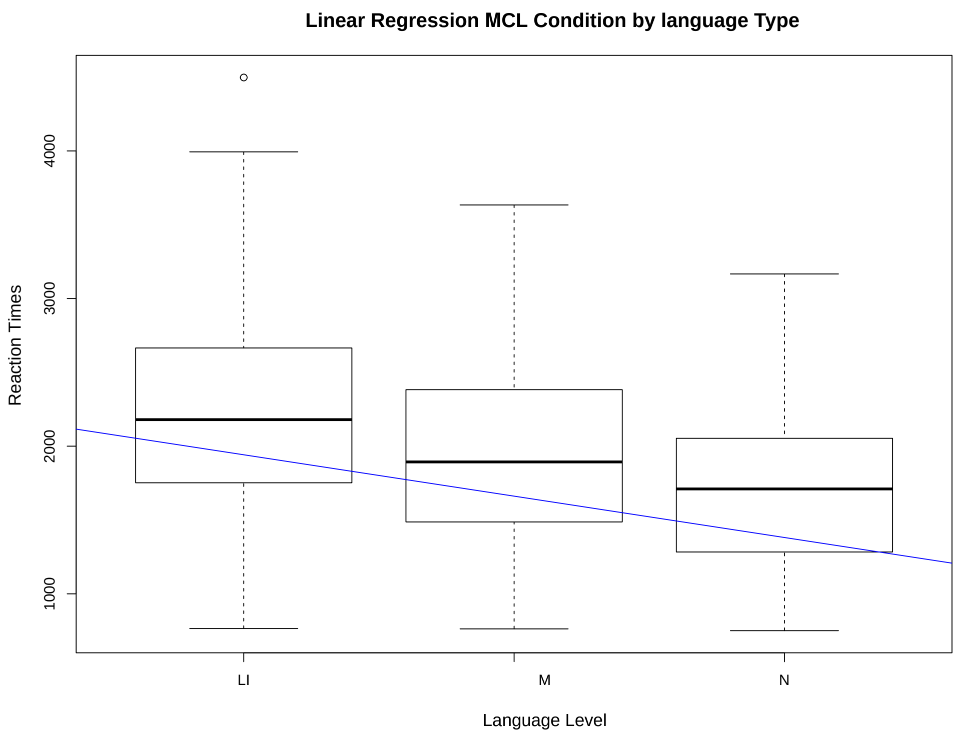


Figure 38. Linear Regression Reaction Times of MCL for the Different English Placement Levels

A simple linear regression was calculated to predict the reaction times of the MNL responses based on the 3 English placement levels. A significant regression equation was found (F(2,430) = 21.38, p = 2.288e-10), with an R2 of 0.0981. Participants' predicted reaction times is equal to 2146.50 – 190.20 ms. for every English placement level gained. Therefore, reaction time decreased 190.20 ms. for every English placement level gained (cf. Figure 39).

A graphical representation of both of those linear regressions is shown in Figure 40. As we can see, there is no interaction between those two languages for all three English level placements, but they both show decreasing reaction times with every additional level of English placement. The MNL proves to have consistently faster reaction times in all English placement levels, and therefore, we can conclude that MNL helps comprehension and information processing more than MCL regardless of participants' English placement level.

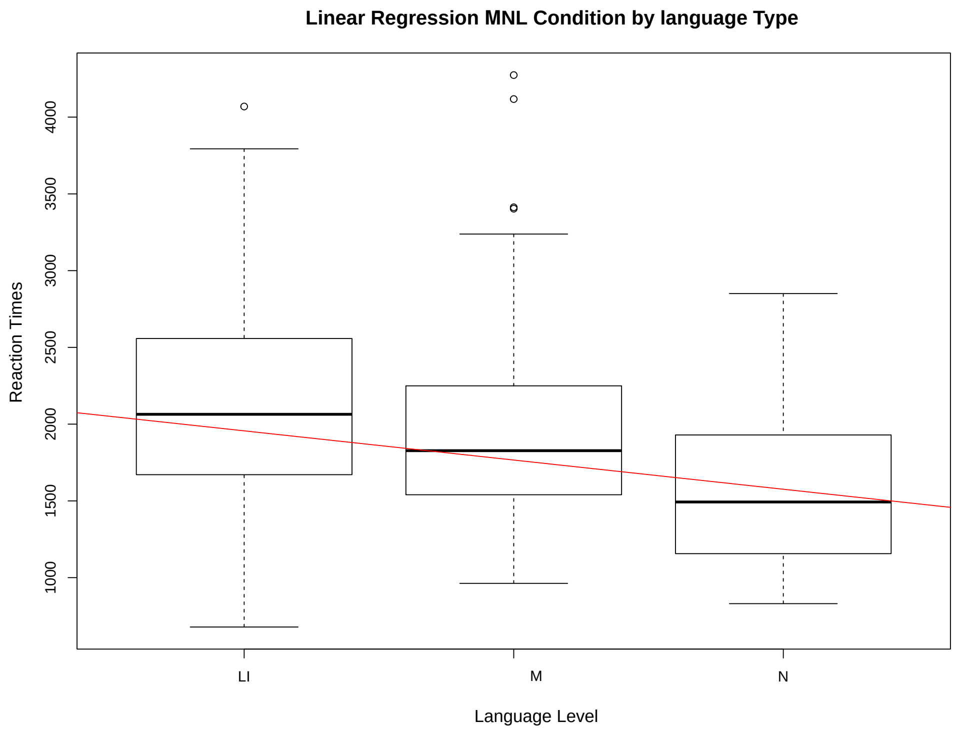


Figure 39. Linear Regression Reaction Times of MNL for the Different English Placement Levels

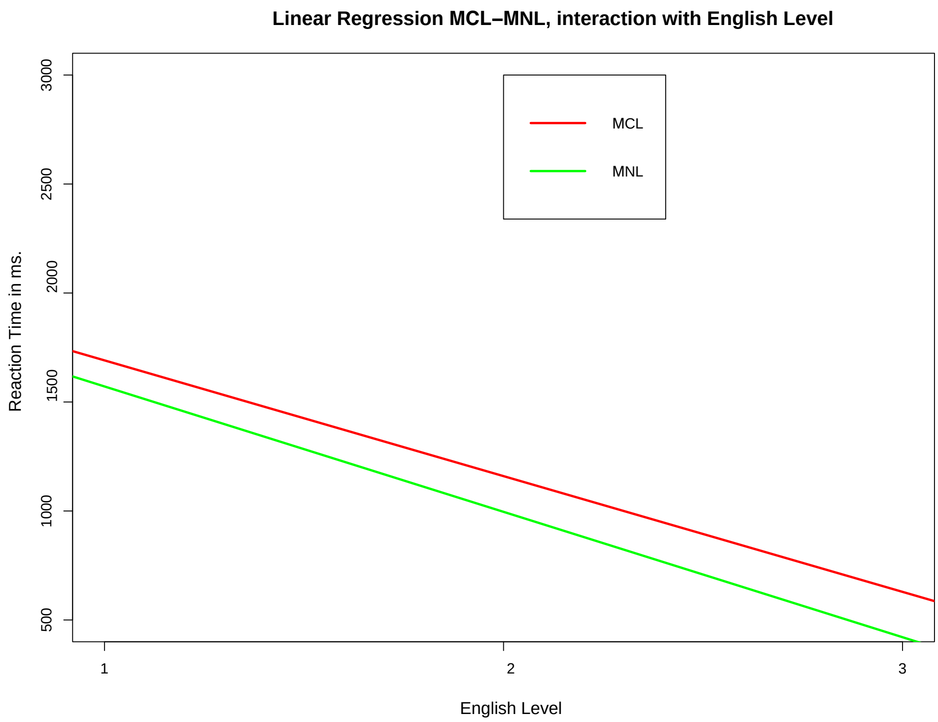


Figure 40. Linear Regression of MCL and MNL Reaction Times for the Different English Placement Levels

We are not going to report the results for the same tests of research questions 4 and 5 for the second experiment as the general effects did not show any significance.

5. **Did the language factor play a different role for different levels of familiarity with Airbus Controlled Language regarding reaction times?**

- **Experiment 1**

A series of t-tests were done on the different levels of familiarity and as we can see from Table 5 there was no significant difference for the two language groups for all the different familiarity groups. There was no influence of familiarity with the Airbus controlled language on language reaction times.

	None	Beginner	Intermediate	Expert
MCL mean	1914.299	2282.857	2117.772	2241.823
MNL mean	1767.780	2124.295	1991.894	1992.262
Difference	+146.519	+158.562	+125.878	+249.561
P value	0.466	0.1127	0.1941	0.5142

Table 5. Means of MCL and MNL Reaction Times for Different Levels of Familiarity with Airbus CL

5.5 General Discussion

As shown in the results of hypothesis 1, MNL condition shows significantly faster reaction times than MCL condition in experiment 1, and both languages performed equally with regards to accuracy (in hypothesis 2). This could be explained by the fact that the syntactic changes between the two conditions did not have enough disparities to warrant observable differences in accuracy, whereas the observed differences in reaction times were able to highlight the subtle syntactic variations that led to faster comprehension.

Experiment 2 did not show any significant effects for the main hypotheses (both reaction times and accuracy) with regards to the two language conditions. This is due to the added variable of time in the presentation of the stimuli. We can conclude that speed plays a role in the comprehension linked to language conditions. In the first experiment, speed and to a certain degree the stress it provoked, accentuated the role of the more natural language in information processing. The absence of that stress by providing abundant amounts of time to respond (such as the case in experiment 2) attenuates that difference, and both languages have almost the same level of performance with regards to reaction times, or at least produce a difference in the observed measure of central tendency (median reaction time of MNL < MCL), but that difference is not statistically significant. Therefore, we would like to investigate the role of speed and stress linked to time pressure on the interpretation of the two language conditions and the resulting performance. This will be one of the important points we will deal with in the third experiment in the following chapter.

Additionally, in experiment 1, there was no interaction between the two languages with regards to the 6 levels of syntactic difficulty but reaction times get slower when difficulty increases in both languages. We can also conclude that MNL produced consistently faster reaction times than MCL in all difficulty conditions.

As we illustrated in research question 3, incongruent stimuli had an additional touch of difficulty and that is reflected in the reaction times' discrepancies for congruency conditions in both language conditions. In experiment 1 and 2, incongruent stimuli showed significantly faster reaction times for the MNL condition over the incongruent MCL condition, while the congruent stimuli did not. Moreover, even though we do not have a significant general effect for reaction times in experiment 2's data, still the incongruent stimuli show significant effects for the MNL condition. Therefore, in cases of increased difficulty the more natural language helps ease comprehension.

Concerning English placement levels (research question 4) in experiment 1, MNL seems to facilitate comprehension for participants in the basic intermediate level placement, and this suggests that weaker English speakers would benefit more greatly from a more natural language than confirmed speakers, or at least we could say that the effect is more obvious. While native English speakers performed better on average in the MNL condition, that effect was not statistically significant and should be the object of further studies with bigger samples.

We could also conclude that there is no interaction between the two language conditions' reaction times and the different English level placement (one language did not start out having better performance than the other but ended up performing worse in different level placements), however we do observe a downward tendency in reaction times the more proficient speakers become. Natives have significantly faster reaction times than basic intermediate English speakers.

Lastly as seen in research question 5, there were no observed effects of participants' Airbus controlled language familiarity or absence thereof on the results of our hypotheses. But we must point out that the majority of the participants had little to no knowledge of the syntax of the Airbus CL, and there were only 5 participants who categorized themselves as experts. Largely unequal sample sizes could explain the absence of heterogeneity in the means of the groups. Additionally, while syntax and structure of the messages were similar to the original Airbus corpus, the experimental context, task, absence of abbreviations, and non-aeronautical stimuli were sufficiently dissimilar to what experts are used to dealing with, that it could easily explain the absence of familiarity effect.

5.6 Limitations and Perspectives

While opinions were divided on preference for either one of the language conditions, certain comments on the participants' behalf were recurrent. Some mentioned that the word "available" in the stimuli was confusing. Since we based ourselves on the original corpus, the messages actually had a real notion of availability, such as informing pilots of the availability of a particular function. However, as the experiment went forward and the congruency task with images was designed to test out comprehension in different language conditions, the word "available" almost lost its original meaning as "able to be used, obtained, or reached" (from Cambridge Advanced Learner's Dictionary) and became a proxy for "is the object present/visible/shown in the picture". We thought about changing the word "available" and replacing it with one of those previously mentioned synonyms but then every potential substitute had its own polysemy problem, such as "visible" which could be interpreted as "can you see it clearly on the picture?" or "shown" which could imply something deliberately displayed by someone, or "present" which implied that it could be absent (which particularly added more issues for incongruent stimuli) or that the object would be somehow linked to a temporality such as being present at the moment of the picture but not necessarily before or after.

Therefore, while "available" was far from perfect, we took the decision to keep it in our messages and warned participants in the practice sessions before the experiment about its use in the task, and how best to interpret its meaning. This was one limitation that was persistent but which was the by-product of the experimental design and task conducted in laboratory conditions.

In the following experiment, we bypassed the congruency tasks and went straight to performance of injunctive messages, which resolved the availability issue. Consequently, these two experiments were limited to passive comprehension that we will supplement with real time performance tasks that include the urgency factor (speed of stimuli, and stress generated by limited response time) in the following experiment. We will also be recruiting more native speaker participants to have a larger panel of the target population, and ascertain whether the different syntactic language conditions reflect equally on native and non-native English speakers.

5.7 Conclusion and Way Forward

The results from this experiment are somewhat satisfactory as they show that our initial hypothesis is validated in a certain number of conditions. In all cases, contrary to popular belief more simplification and linguistic economies never led to better performance (MCL never performed significantly better in reaction times or accuracy than MNL). Furthermore, these experiments brought us first elements of empirically tested data which deeply question controlled language construction, and simplification in general. It showed that what we sometimes mistakenly label as superfluous or empty syntactical elements could go a long way in ensuring better comprehension and faster information processing.

6

Effects of a More Natural Language on Comprehension in Action Statements

6.1 Introduction

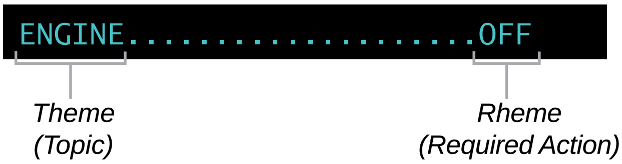
After the interesting results obtained in the first two experiments, specifically concerning the more natural controlled language (MNL), showing significantly better performance (in RTs) in the information category as opposed to the more coded controlled language (MCL), we will proceed to the action statements category.

To recapitulate the main conclusions of the previous experiments:

- There is a significant effect of MNL on information processing.
- Particularly for Lower intermediate English speakers, and more tests should be done with a bigger sample of native speakers (although the means followed the general trend as well).
- There is a significant effect for more complex stimuli: the MNL helps disambiguate and process information faster than a more coded version.
- This effect is reinforced in the more stressful condition when time pressure was a factor
- In all conditions, the MCL never performed significantly better than its more natural counterpart.

These conclusions led us to the construction of the experimental design of this third experiment which will take into account the results of the first two experiments, and which we will in turn apply to the action category elements of our corpus.

As a reminder, the action category statements as used in our corpus stand for commands or injunctions that the user must follow. In other words, these statements consist of injunctions displayed on the ECAM destined for pilots to perform an action linked to a button/a lever/an altitude or various other cockpit interactions. They have the following syntax:



The word "engine" is the theme or the element on which the action will take place, and the word "off" is the rheme or what one is meant to do to/with the "theme", in this case switch the engine off. The theme and the rheme are separated by dots to reach a maximum of 41 characters, spaces and dots included. The dots are supplemented after the words are written so we always have 41 characters on screen (for simplicity purposes, from this point on, we will only use 5 dots to mean whatever number of dots needed to separate the theme and the rheme, for instance: engine.....off).

This typographical ellipsis is unique for the action category, as is the blue color and syntax (Theme/Rheme separated by ellipsis). The standardization of the color coding and format is meant to eliminate different possible interpretations, i.e only injunctive statements (orders expecting an ensuing action) are written in blue with the theme/rheme format. Other categories of information such as titles, questions, conditions etc. have a different color-coding scheme and syntactic and typographical formats. The messages are therefore denoted by different linguistic and semiotic representations that could be combined and are meant to be learnt by the users.

By learning what the standardized format stands for, a user is meant to understand how to react to it, in this case perform an action.

Therefore, once again we hypothesize that while the linguistic economies and simplification in these statements are compensated for by standardization, learning, and training to make sure the statements are understood and reacted to correctly, there might be a more natural way to reduce ambiguities that is translated into better human performance.

Our general hypothesis is the following:

Coded format words inserted in natural language structure sentences produce a faster reaction time and are understood more accurately (less mistakes in performance) than words in isolation (cf. Figure 41).

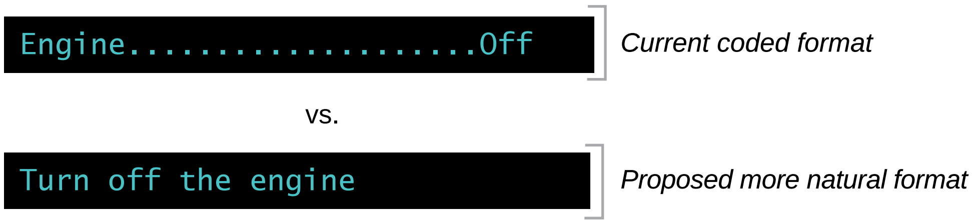


Figure 41. Example of an action statement

We hypothesize that the current coded format could be understood in two different ways, especially under potentially more trying circumstances of duress, fatigue, drowsiness, stress, etc.:

1. The engine is switched off/turned off → Statement informing of the current state of affairs
2. Turn/switch off the engine → Injunctive action statement

Moreover, the sentence structure provides a failsafe way of avoiding ambiguity. The sentence "Switch off the engine" adds two more words to the original statement "engine.....off" yet completely eliminates the second interpretation. Thus, information is solely contained in the linguistic elements, excluding color and typographical separation. There is only one possible way of interpreting this sentence. Therefore, controlling this multiple interpretation situation is made possible by naturalizing (bringing close to natural language format) and "complexifying" a more coded, abbreviated, and originally simplified language.

We carry on with this approach throughout our experimental design, but before the creation of the complete corpus of stimuli, an adequate way was required to empirically test this hypothesis on injunctive statements for which the expected reaction is to perform an action.

In the previous experiments, a judgment task was suitable enough for the informative category i.e statements color coded in green which inform pilots of the current state of affairs without demanding the performance of an action as a result of its being displayed on the monitor. For those statements, for example "left tank is available", comprehension was assessed on the basis of a yes/no congruency judgment task (cf. Chapter 5).

It is quite impossible to design such a judgment task for injunctive statements demanding the performance of an action, as we would not be able to assess whether the participants truly understood the real meaning of the injunctive statement, especially in its hypothetically more ambiguous coded format (engine…..off).

Furthermore, by inserting the individual words in a sentence we are injecting a process into it by giving the action more agency. In Austin's (1975) speech act theory terms, we would be transforming individual words into clear directive sentences, sentences that will have illocutionary power (show the meaning conveyed). A directive sentence clearly has injunctive power, signifies an order, an action to be executed through the illocutionary act and with the perlocutionary act (the actual effect) as a result.

It is our hypothesis that individual words on the other hand, simply evoke the locutionary act (the literal words and their meanings), because the syntactico-semantic context is missing. While the coded language with its literal minimal units of meaning, color, and elliptical format has been thus far functional for, to quote Austin, "securing the uptake" and fulfilling what is being asked, we argue that the illocutionary and the perlocutionary acts are explicitly missing, and by using formats that convey them we will be making comprehension faster and more optimal in various contexts.

It was challenging to find in psycholinguistic literature such tasks as they tend to be judgment tasks, and isolate as much as possible the interference of motor skills in language comprehension tasks, (cf. Section 3.6). On the other hand, AECMA SE evaluations discussed in Section 3.5 do not include performance of an action but rather use reading comprehension to evaluate accuracy.

A new task that is suitable for testing this hypothesis was designed that differed significantly from judgment tasks in traditional psycholinguistic experiments (which lack the need to test performance since the aim is to map out cognitive processes, and not the results of the actual performance linked to comprehension of stimuli); and secondly this proposed task also differed from controlled language evaluations of AECMA SE (evoked in

Section 3.5) which do not use tightly controlled experimental protocols to evaluate subjects' actual performance with respect to the messages, but instead use reading comprehension tasks related to previously read instructional texts.

6.2 Experiment Design

6.2.1 Interface Design

Contrary to the first two experiments, this experiment was designed using aeronautical terms when possible derived from the actual corpus of action statements. Therefore, it was important to have an interface on which to perform these actions that was representative of actual cockpit interactions. Note that we did not use Airbus simulators since we wanted to experiment using laboratory controlled conditions without interference from different variables that inevitably come along with any flight simulator scenario. Additionally, we would have been limited to the small and very occupied test pilot population, and the difficult logistics of booking and securing the simulators all along the experimentation phase.

We designed a visual interface using Sketch[32] and Adobe InDesign software that were more or less inspired from the A350 overhead panel (cf. Figure 42) and other cockpit functions. This phase was challenging to put in motion as we imagined interactions on a touch screen in order to minimize mouse bias in performance, and so far, there are no tactile interactions in Airbus cockpits. The action category statements are messages that appear on the ECAM, and the action to be performed is usually performed on physical buttons or levers. Below is an extract of a photo of some of the overhead panel buttons.

The challenge was to transform those buttons into non-physical digital buttons on a human-machine interface for a tactile screen that had at the same time the function of the ECAM (displaying messages) and the buttons (where the action was to be performed).

32. Sketch. The digital design toolkit. [online] Sketchapp. Available at: https://www.sketchapp.com/ [Accessed 9 Dec. 2018].

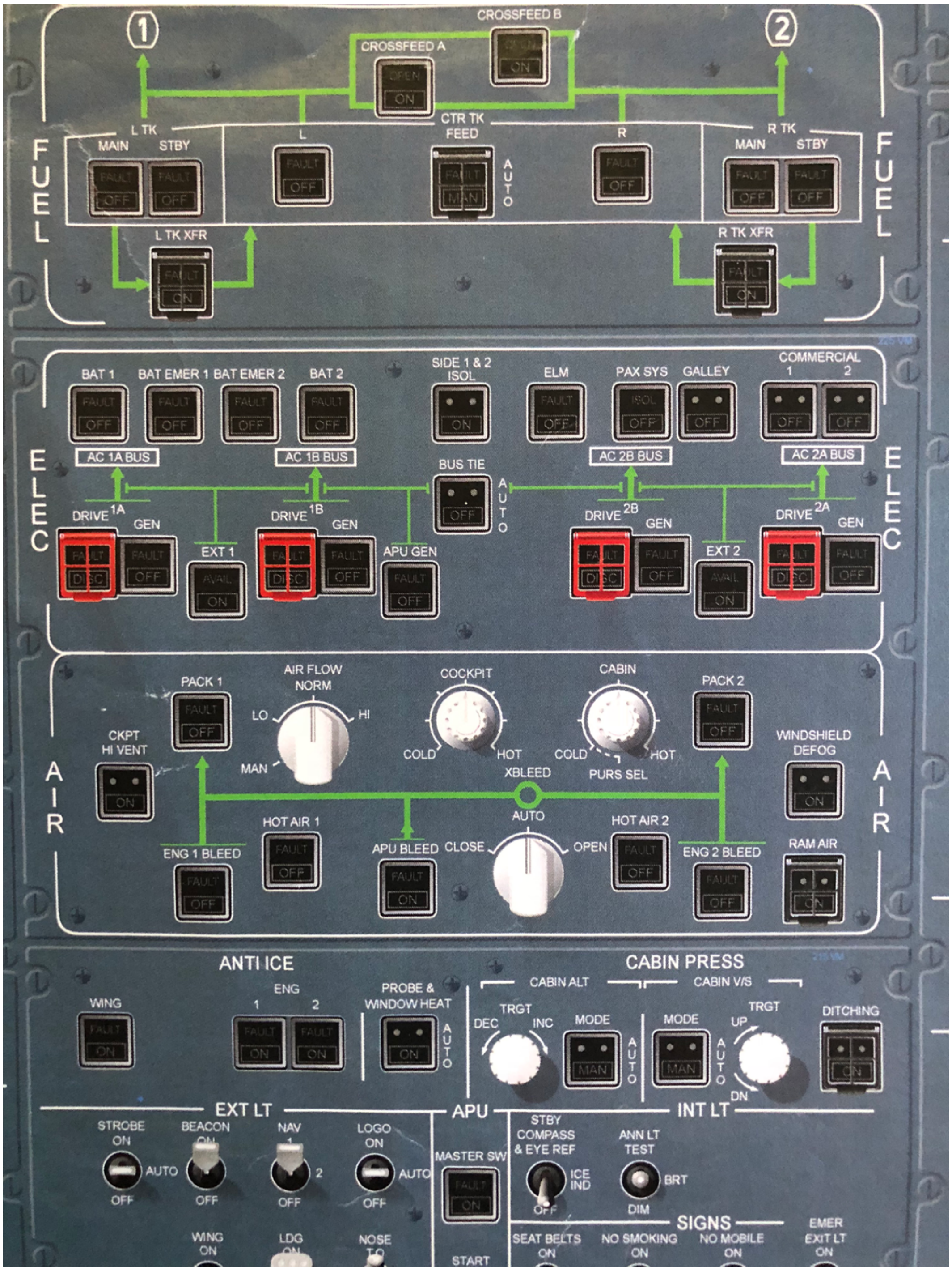


Figure 42. Extract of an Image of the Overhead Panel in the Airbus A350

The first obstacle was the difficulty to remain loyal to the original buttons as digital tactile buttons cannot look the same as the physical ones. The second obstacle was simplifying the interface so as to be easily used by non-pilot participants, and consequently redesigning a whole new interface from scratch. As well as pilot participants, it was important to also have non-pilot participants in order to be able to test our hypotheses on a population that is not trained on the original controlled language, so as to test the intuitiveness of the proposed language formats and avoid expert bias.

We tried when possible to remain loyal to the zone concept already present in the overhead panel, and to select a few significant aspects of piloting such as managing the fuel pumps, the engines, the hydraulics, the brakes, the air, and the speed. What is more, in order to design a usable, easily accessible, and user-friendly interface, we needed to follow a few UX/UI rules (User-eXperience – User-Interface, an up and coming conceptual design discipline).

We mainly used 3 Gestalt design principles of proximity, similiarity, and enclosure, as well as industry standard tactile button sizes (more details in Annex H).

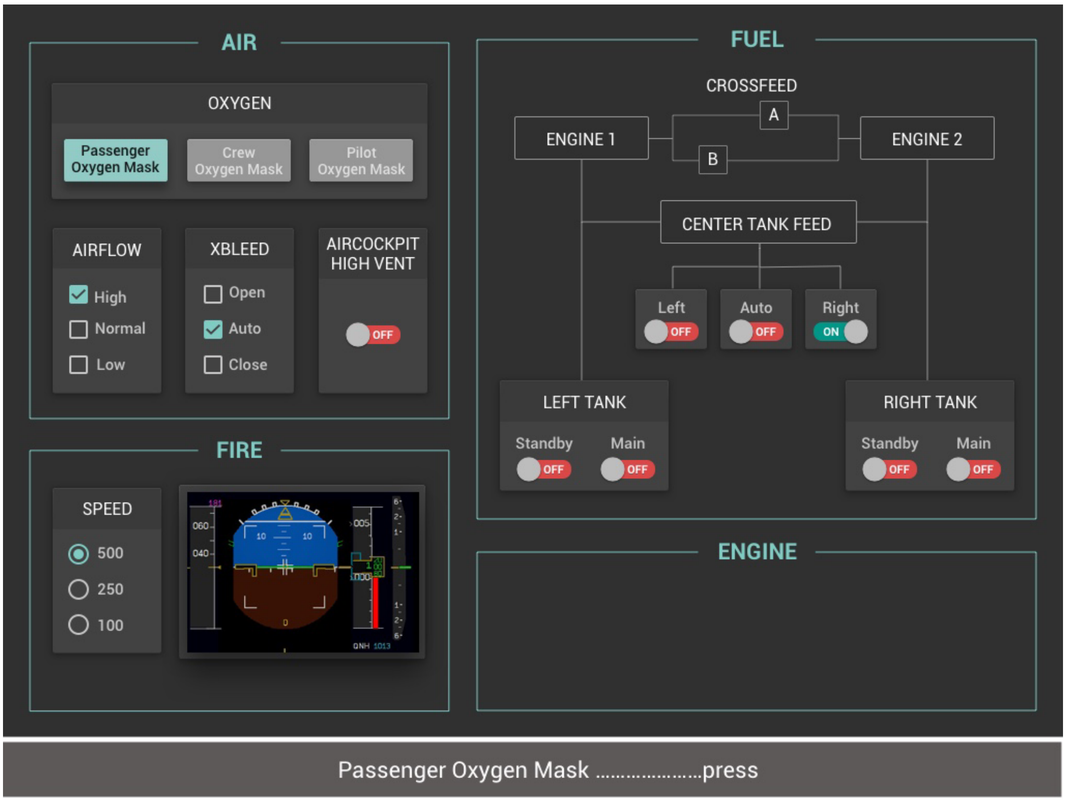


Figure 43. Experiment 3 Interface Intermediate Version

However, there were certain problems in this first attempt (cf. Figure 43). We were designing a user interface that follows suitable visual design principles, but we also needed to make the design suitable for behavioral controlled experimentation that we later on integrated in E-prime 3[33], a comprehensive software for behavioral research. For instance, the on/off buttons could not have stayed in the format in Figure 43 which would require participants to slide the slider to turn off or on, because this would make it difficult to assess reaction times of the initial touch. Therefore, our final version of the interface had to be altered and looked like the following (cf. Figure 44):

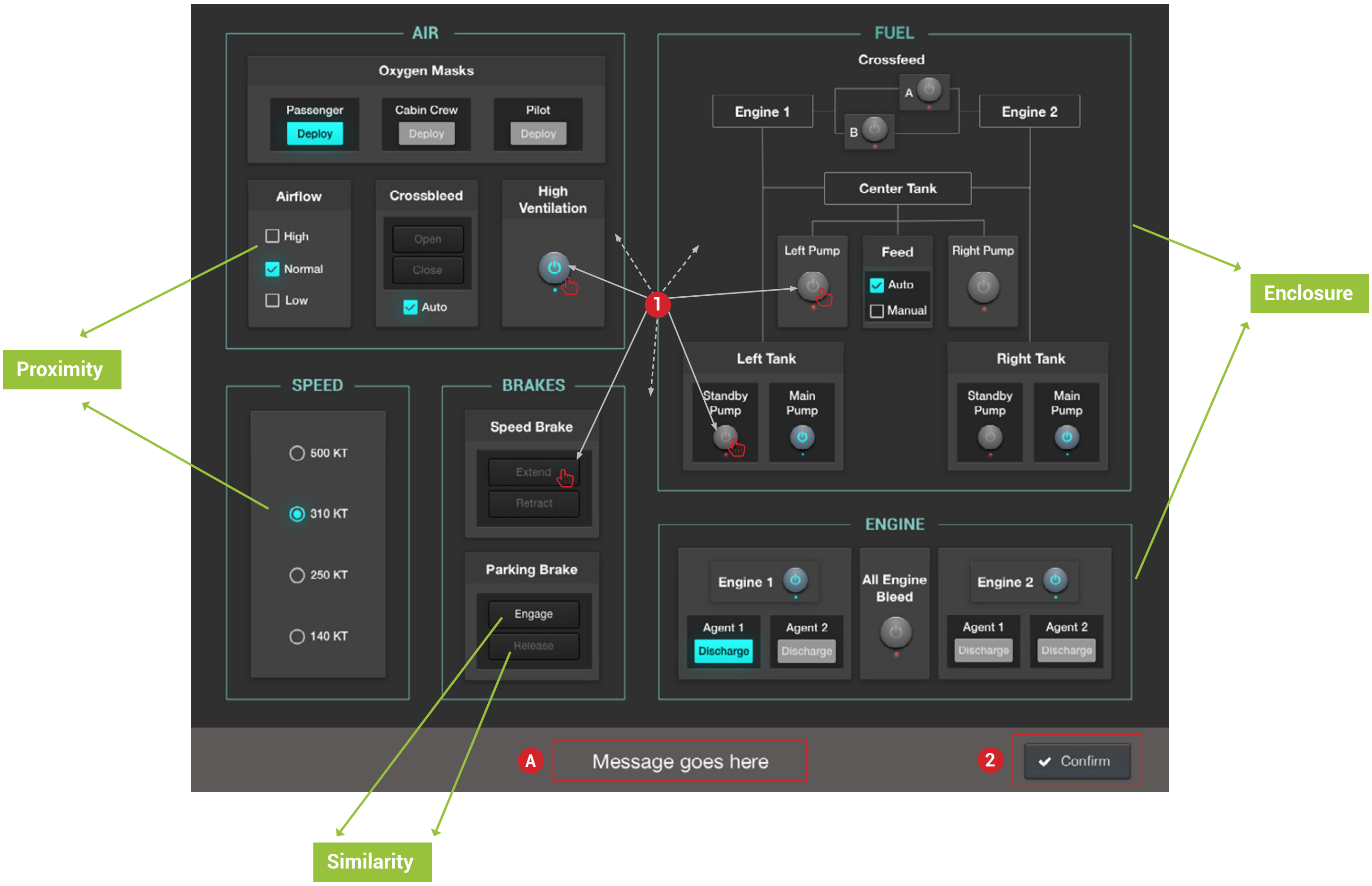


Figure 44. Experiment 3 Interface Final Version

This final version included 5 zones of Air-Fuel-Speed-Brakes and Engine. The size of the interface was 1260 x1015 pixels (which means it did not fill out the entire touch screen). Some of the main differences with the cockpit are:

33. E-Prime® | Psychology Software Tools. [online] Pstnet.com. Available at: https://pstnet.com/products/e-prime/ [Accessed 9 Dec. 2018].

- The On/Off buttons became push buttons from which we could easily calculate reaction times. The button is meant to light blue/green when it is ON and red when it is switched OFF.
- The rotary buttons that had multiple positions needed to be turned into tick boxes because of the touch screen interactions, like in the case of Airflow in the Air zone. The tick boxes were either ticked when selected or not ticked when not selected
- The speed was turned into radio buttons for the same reasons. The radio button was selected when activated and empty when not selected.
- The buttons where actions were required such as "Deploy" in the Air zone or "Discharge" in the Engine zone were made into rectangular push buttons with a light gray background when the action required is unitary (deployed or not, discharged or not), and they turned blue/green when activated. When the action had two possible positions such as open/close in the Air Zone or Extend/Retract in the Brake zone, it had a dark gray background in a separate enclosure and turned white when selected. The distinction between unitary and double position buttons was deemed required at this stage to respect the similarity principle, but as we will see later on, it might not have been the most judicious choice: All action buttons should have been designed to be activated in the same manner so as not to confuse participants, since the end user does not have time to realize the subtle differences between both. More on that in Section 6.4.1.4.
- Finally, in our interface, only the titles of the zones (AIR-ENGINE etc.) were written in all capital letters (as opposed to everything being written in all capital letters in the overhead panel and on the ECAM display). Everything else in the interface including button names, titles, and messages was written in title case lettering. Our choices were based on several studies that concluded that lower case or title case (only first letter in every word is capitalized, except for less important words such as articles and conjunctions like "the", "and", "a" when they do not head a sentence) helped facilitate reading and reading speed. Paterson and Tinker (1946)'s concluded:
  - At a distance of 38.1cm, lowercase was read 18.9% faster than uppercase, 24-point bold face

- At a distance of 182.4cm, lowercase was read 5.3% faster than uppercase, 60-point bold face
- Not until we reach 518cm do we get significant results for uppercase (which is why road signs are usually in all capital letters, distance + size need to be proportional)
- The use of capitalization, especially words written in full upper case is best used to call attention upon the said word, or to signal urgency, for example “DANGER”.
- It is easier for our brains to complete missing parts of letters from words when they are lower-case letters, and the different shapes of the lower-case letters help us recognize words faster as in Figure 45 and Figure 46.

Figure 45. Lower Case vs. Uppercase 1

Figure 46. Lower Case vs. Uppercase 1

6.2.2 Integration in Eprime

After the interface was visually designed, we integrated the interface into a carefully designed E-prime script[34] that emulated an interactive interface

34. The integration of the design and button/tactile interaction was done in conjunction with Pierre Vincent Paubel, CLLE lab engineer, whom we thank for his work and availability.

with real buttons. The experiment was designed and played out on a Dell Precision 3510 laptop and a 23-inch touch screen (1920 x 1080 pixels).

Had we created an application or an actual software, we would be losing E-prime capabilities for measuring very accurate reaction times. Therefore, we chose to use E-prime and simulate actual responsive interactions with the interface, as would any touch screen on a mobile device for example. It was important for us that participants have believable feedback, for example if asked to activate a fuel pump, after the action is performed by the participant, the fuel pump would light in blue/green to show that it is activated along with a correct sound ding.

We achieved that by first creating images of all 36 buttons in both their activated and deactivated status (checked or unchecked, etc.). Afterwards, we simulated feedback as it would be in a real responsive interface (the buttons lighting when the participant pressed on them on the touch screen) by having E-prime display the exact same image (as feedback to the action performed) that we'd had previously altered to have the correct button light up when pressed. In other words, the participant would think that the interface is responsive when they press a button and it lights up almost instantaneously, but in reality, it is just an elementary image feedback that we have programmed. This also meant that for every single stimulus, we needed to design 2 images, one for the target and one for the correct feedback, along with programming E-prime to display a down button function (pre-activating the touch) for every button. This whole process, while very time consuming, allowed us to have a seemingly interactive tactile interface that registers extremely precise reaction times. The responsive interface was important because it made the task seem more realistic and closer to an actual piloting exercise. Moreover, the feedback allowed participants to understand and learn the task, and allowed them to better judge the intent behind potentially ambiguous messages (injunction or statement), something we will clarify in the Research design (cf. Section 6.2.4).

6.2.3 Stimuli Construction

The stimuli were progressively constructed based on the original corpus of action messages (injunctive statements: orders expecting an ensuing action). They were constructed concurrently with the visual interface design, while keeping in mind the possible button interactions.

Apart from the absence of upper capitalization mentioned in the previous section, we also did not keep abbreviations and acronyms (e.g. ENG for Engine). We decided that abbreviations are an object of study in and of itself and adding this extra variable in our current research design might influence results, especially since a lot of the abbreviations used in the corpus would only be understood by a specific population of trained Airbus pilots, and would be difficult to be decoded by the lay person. In fact, the same experiment could in the future be duplicated using abbreviated terms, but the interface must also reflect that with the abbreviated buttons' tags and functions.

This is the case with the current corpus, as the messages take into account the actual tag of the buttons, for example if the button has three positions: LO-NORM-HI, the message on the ECAM would be:

```
AIR FLOW ....................HI
```

Even though both versions of the word exist in the allowed vocabulary list of the current cockpit controlled language, we use "HI" and not "HIGH" because the physical button tag in the cockpit uses it, and not because we are exceeding the 41 characters allowed on the screen nor because it is better for reading and comprehension. Therefore, since we have leeway in this experiment as to what tags the buttons could take on our own interface design, we chose not to complicate matters and add abbreviations when the full word is a better fit.

Furthermore, the task of interface design and stimuli construction were done simultaneously and adapted to one another. The complete list of stimuli could be found in Annex I.

The stimuli were made of two different sets. The MCL ones were either taken exactly as is from the corpus (except for the capitalization and abbreviations), or were slightly adapted to the interface design when needed but still followed the exact same syntax as the original controlled language (theme and rheme separated by dots). An example of messages that we added/adapted in both sets of MCL and MNL because it suited the design and the interactions would be the oxygen mask interactions: "Deploy passenger oxygen masks", "Deploy Cabin Crew Oxygen Masks", and "Deploy Pilot oxygen masks". In the cockpit, we only make the distinction between the crew oxygen masks and the passenger oxygen masks. The cabin crew and the pilots are not two distinct entities as is the case in our experiment. This addition does not however make the task less believable to experts and pilots, and is aligned with the experimental and linguistic strategy that we employed. The MNL contained a parallel set of stimuli to the previous one but used natural language syntax (cf. Figure 47 and Figure 48).

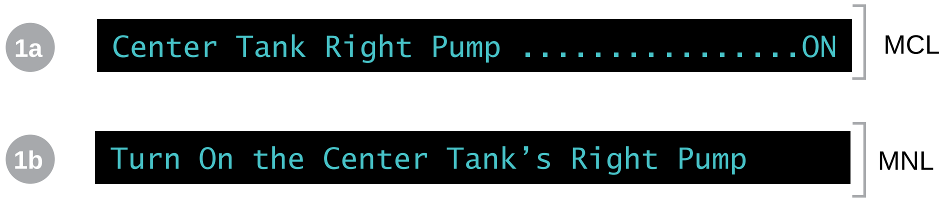


Figure 47. Example 1 of MCL and MNL Action Statement

Figure 48. Example 2 of MCL and MNL Action Statement

The task of naturalizing the MCL into MNL was quite straightforward with regards to injunctive action sentences (as opposed to informative declarative sentences in the previous experiments). The English injunctive mode typically starts with the verb form in the infinitive (Set, Deploy, Turn, etc.) followed by the topic or object and potentially a prepositional phrase (eg: "Set Speed to 140KT", or "Deploy Oxygen Masks"). All the messages followed this logic.

Some messages however had to be changed due to the design of the interface which mimicked the architecture of the fuel tanks on the aircraft. For instance, in example 1, we had a more obvious choice (or a more natural one in the English syntax) of naturalizing the MCL into "Turn On the Right Pump in the Center Tank". This was not possible or at least not ideal because the right pump is visually and physically located inside or under the main center tank (cf. Figure 49). In our pretests, this showed that the participants were confused by the message because they were visually searching for the right pump first as it comes first in the sentence and then moving their gaze upwards to the center tank, whereas it is clear from the interface that the center tank is the bigger entity that supersedes and contains the right pump, and which is visually more accessible. Additionally, the right pump having first position in the message made participants confuse the right pump in the center tank and the main pump in the right tank. Therefore, it was wiser to stick to the same word order logic used in the MCL (Center tank right pump….off) and use the possessive form (which did not come without its problems, (cf. Section 6.4.1.4)) in the MNL (Turn On the Center Tank's Right Pump). One should note that this did not resolve all ambiguities and complexities linked to this fuel section and the abundance of tanks and pumps it included, but it did help the participants' visual search-and-locate-the-right-button part of the task. Choosing this word order for the MNL in those cases also helps not having big syntactic differences between the MCL and MNL set, which was another way for us to control our variables. This was one way the interface design (and de facto the aircraft architecture) influenced and put pressure on the linguistics, something which is quite common in this domain. Another message that was changed due to the same reasons was in the engine section: "Discharge agent 1 of Engine 1" which was subsequently changed into "Discharge engine 1's agent 1".

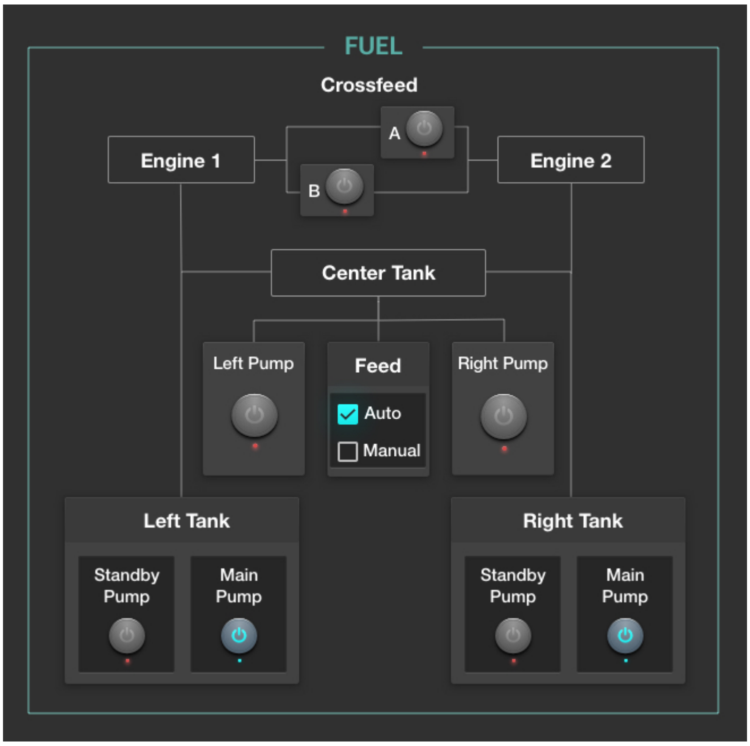


Figure 49. Fuel Zone Architecture in Interface of Experiment 3

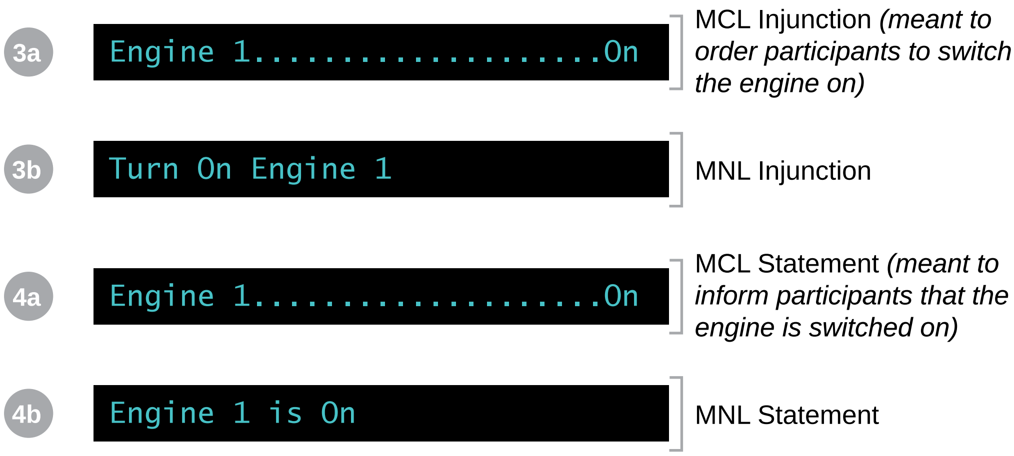
One way linguistics or at least semantics influences interface design would be the fact that center, right, and left have to follow the visual logic of their meaning. In our case, the center tank, while clearly in the center with its proper enclosure was drawn higher than the left and right tanks. This was not optimal, as the "center" in this case was not a proper anchor to the left and right. A better design would have been if the left, center, and right were aligned. This was unfortunately discovered late in the process.

Regarding the MCL set, all the stimuli were made to have exactly 41 characters as this is the actual limitation of the action statements on the ECAM monitor in the A380 and A350 aircraft. We based ourselves on that limitation for the construction of the stimuli. We adapted the number of dots in each message so as to always reach 41 characters (dots and spaces included). We never experienced any issues regarding this limitation, in other words, none of the MCL sentences in our corpus required us to shorten them with respect to this limitation, as could be sometimes the case in the original controlled language. Additionally, there was exactly one space after the theme before the dots were inserted, once again in the same manner as the original corpus.

On the other hand, the MNL set did not always have 41 characters on screen as the absence of dots in the sentences did not allow it. MNL had between 15 and 41 characters in all sentences, therefore, the 41-character limitation was not violated in either of the MNL and MCL sets.

6.2.4 Research Design

After the interface design and the message construction, the task and research design were established. As we can see on the interface design in Figure 44, the on-screen message appears right below the interface, as well as a "Confirm" button in the bottom right corner. This confirm button was added to ensure that one of our initial hypotheses was evaluated: As was briefly discussed before, the coded format of the injunctive statements in our initial controlled language corpus (Theme.....Rheme) could potentially be understood in two ways. One as a statement and one as an injunction: Engine 1.....Off could mean to switch off the engine or that the engine is switched off. Therefore, part of our task was not only to contrast the usage of the coded language and the more natural one, but also to test the adequacy of the format as it is currently being used. In other words, is the "theme..... rheme" format more accurately understood as an injunction or a statement? Therefore, we endeavored to include that query in our experimental design, first by adding a confirm button in the interface, and second by constructing a set of MNL *statement* messages as follows:



The task included messages which required participants to perform an action directly on the buttons as in the case of 3a and 3b. But it also included messages such as 4a and 4b which required participants to press on the confirm button to signal that the action is already performed, i.e. the button is already pressed, or in the case of this example, the engine is already running and the on/off button is on.

Note that 3a and 4a are written in the exact same manner because we would like to test exactly what possible interpretation (injunction or statement) is more naturally understood in this format. 3b and 4b are the more natural equivalent to the coded version with explicitness as their main asset. It is worth mentioning that using the theme.....rheme format for anything but injunctions (in this case a statement as in 4a) could be considered as a violation of its initial sole intent in the cockpit, which in practice is reinforced by training, standardization and color coding; nonetheless, we found it important to question its expected usage for injunctions by basing ourselves solely on syntax while isolating all other variables. This way we could determine the leanings of this format towards one interpretation or the other. Barring general task instructions, by taking away training, standardization, and color coding, we would be testing for the adequacy of the format itself and the intent with which it is better employed.

Consequently, we had 96 target stimuli per participant split into 48 injunctions and 48 statements that were in turn split into MNL and MCL equally. As we were also testing the effect of time pressure, the 96 stimuli were randomly split into urgent (4000ms to respond) and non-urgent conditions (12000ms to respond). The stimuli chosen for the urgent phase were inverted to the non-urgent phase and vice versa (2 different lists) for half of the participants to control for possible message and urgency interaction bias. As we can see in Table 6 we had 48 randomized stimuli in the non-urgent phase split into injunctions and statements in both languages, and again the same architecture in the urgent phase.

Non-Urgent Phase = x48	MNL	MCL
Injunction	x12	x12
Statement	x12	x12

Urgent Phase = x48	MNL	MCL
Injunction	x12	x12
Statement	x12	x12

Table 6. Distribution of Stimuli in Urgent and Non-Urgent Phases, Language Conditions, and Intention

- The messages were also semi-randomly divided into interactions on all 5 zones. Some zones naturally had more occurrences because they included more possible interactions such as the fuel zone which comprised many buttons (in contrast to the speed zone for example, which had only 4 selections).
- The same message was never repeated in the same capacity (twice in the same format for the same intention (injunction or statement)), regardless of the urgency condition. In other words, all 96 messages were unique after randomizing language and intention in both urgency conditions.
- Two messages were considered unique even when only the position of the button was changed. For example, “Turn On Crossfeed A” and “Turn Off Crossfeed A” were considered different even though the interaction was to be done on the same button.
- Concerning statement messages, we did not use any of the interactions which had an action verb in the MNL condition. This means that on buttons like the “deploy” button in the air zone, or “retract”

in the brake zone, or "discharge" in the engine zone, we did not use the format of theme.....rheme in the statement condition. If we compare with the previous example:

3a	Engine 1.............On (MCL Injunction)	5a	Speed Brake......Retract (MCL Injunction)
3b	Turn On Engine 1 (MCL Injunction)	5b	Retract Speed Brake (MCL Injunction)
4a	Engine 1.............On (MCL Statement)	6a	~~Speed Brake.....Retracted~~ ~~(MCL Statement)~~
4b	Engine 1 is On (MCL Statement)	6b	~~Speed Brake is Retracted~~ ~~(MCL Statement)~~

Contrary to 4a and 4b, which were possible messages to use in order to test the hypothesis of intention, 6a and 6b were not possible to use with a statement intention as the action verb "retract" would need to be altered to its past participle format "retracted" in order to be understood as a statement. "Speed Brake.....Retract" in our opinion would hardly ever be understood as the speed brake is retracted (i.e. as a statement) because the action verb "retract" carries all the weight of the injunction in its infinitive form. And conversely, we would very much expect the past participle form "retracted" if we expect to be told a statement.

As the postulated potential ambiguity lies in the double interpretation of the exact same message (as in 3a and 4a being the same message but one is understood as an injunction and one as a statement), the fact that the verb form needs to be altered in cases like 6a invalidates that postulate because the messages would not be exactly the same anymore.

Therefore, in the case of statements we only used interactions on buttons which had a position or a state as a rheme. For example the rhemes "on" or "off", "high", "normal", "low", "auto", "manual", could all perfectly be used/understood as a statement in the theme.....rheme format such as in 4a.

6.2.5 Cognitive Process Related to the Performance of the Task

In Section 3.4 we introduced the cognitive processes involved in the use of procedural texts. Injunctions and informative statements are an inherent part of them. Since these message types constitute our stimuli, we are going to discuss the cognitive processes involved in the translation of procedural texts into the performance of the task. The complete explanation of the steps involved is to be found in Section 3.4.1.

The task involved reading the messages below the interface that instructed subjects to either perform an action on one of the buttons in the interface or to press on the confirm button if the action was already completed (the button was already activated on the interface).

Performing an action on one of the buttons of the interface insured that the messages were understood as commands and performed as such. Confirming that the information provided in the message is true (button already lit for "engine 1 is on" message) insured that the messages were understood as informative statements, and that the subject has understood their meaning, and by pressing the confirm button confirmed their validity on the interface.

Therefore, this task involved reading, comprehension, visual pattern matching of nominal entities (by looking for the location of specific buttons and their tags), and execution.

Figure 50 shows the three main steps of the cycle of using procedural texts, which include reading, comprehension, and execution. Inside the cycle are the substeps that account for the linguistic information processing, the development of a mental model, and action plan that precede execution. The proceduralization step is represented in the cyclical and iterative nature of the diagram. We adapted Figure 50 to the cognitive processes involved in the performance of our task. The additional step of pattern matching is an essential one in this case since the interface itself is an intricate part of the performance of the task, and subjects need to search for the right button that was invoked in the read message. Therefore, we introduced it in Figure 51 firstly as an extra main step along with reading, comprehension, and execution and as a sub-step. It falls after the linguistic information processing and the development of a mental model. It is followed by the development of an action plan in the execution step as the right button has already been located in the visual search and find and execution is due.

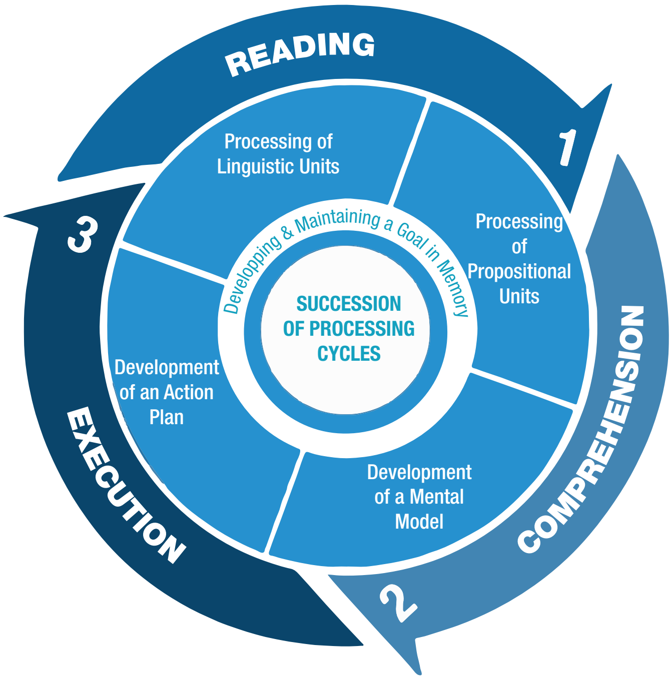


Figure 50. Cognitive Processes Involved in the Execution of a Procedural Text

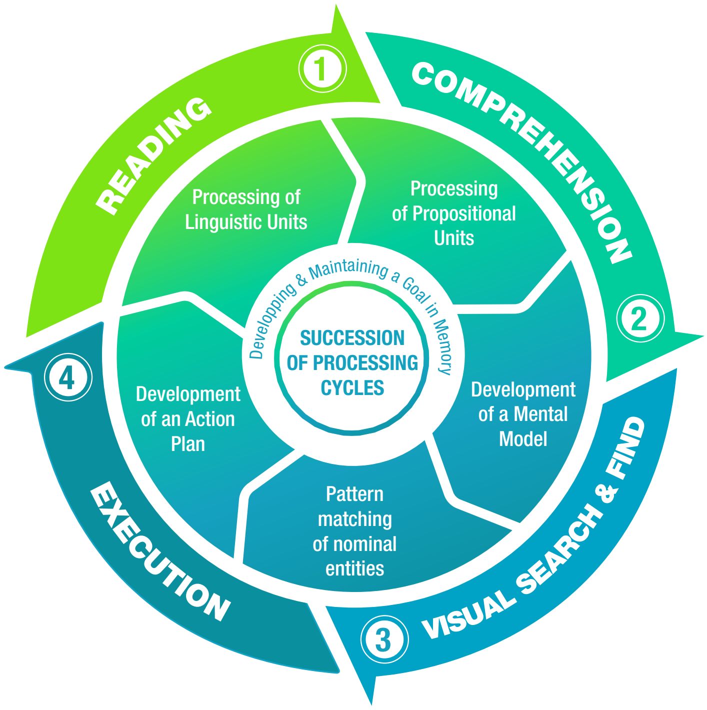


Figure 51. Cognitive Processes Involved in the Performance of the Task in Experiment 3

More concretely, we built an automaton type representation of the task (cf. Figure 52) showing how participants proceed to understand and perform the task correctly. The first state begins with the processing of linguistic units and then processing of propositional units by reading the message shown on the interface. This information would be stocked in short term memory before the development of a mental model to scan the interface in search for the nominal entity (or theme of the message) that has just been read and processed. Once the right button has been located on the interface (which would also include reading the tag, understanding the meaning, and matching it to what has been read in the message), and depending on its state (activated or disactivated) would either:

1. lead to the development of an action plan and execution (confirm button or action directly on button). *(shown in green with a check mark and the number (1) on the automaton)*
2. would lead to going back to step 1 to re-read the message below the interface and process it again, stock it in short term memory to perhaps produce a different mental model in order to go through with the correct process until execution. *(shown in red with an X and the number (2) on the automaton)*

3. would lead to going directly back to short term memory as the message is still memorized in order to alter the mental model before proceeding with the correct process until execution. *(shown in red with an X and the number (3) on the automaton)*

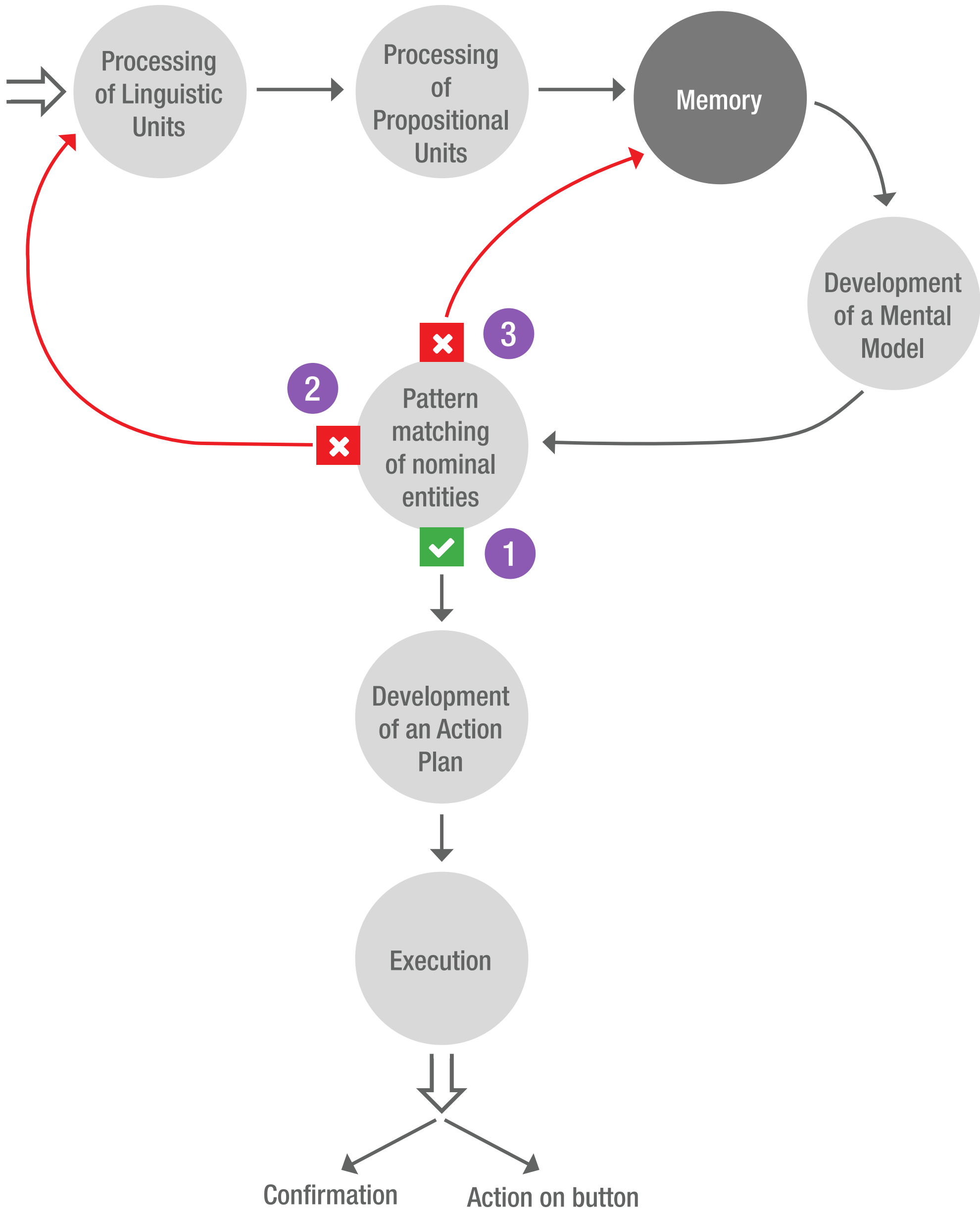


Figure 52. Automaton Representation of Task Execution

In this way, we accounted for the different possibilities that could occur in the performance of the task and through the succession of processing cycles. This task execution model could be more thoroughly verified by analyzing eye tracking data in order to map out subjects' gazes during execution.

If we go back to our previous examples:

3a `Engine 1....................On` MCL Injunction *(meant to order participants to switch the engine on)*

In this example, as the format could have two possible meanings, subjects may have understood this to mean an injunction to switch on the engine (the correct interpretation here) in which case there would be a development of the correct mental model based on the processed linguistic information. Then the subject will have to match the pattern on the interface (location of the engine 1 tag and the corresponding switch) which in this case would be switched off and red, the subject will form an action plan to execute (1) which will result in the pressing of the on/off button in order to switch on the engine and change the state of the button (from red and disactivated to blue and activated).

If, however, the subject originally understood the format to mean an informative statement (the engine is (already) switched on), the same process will occur until the pattern matching step in which the located engine 1 button will show a red switched off (disactivated) button. This would prove to be conflictual/confusing since if the subject understood that the engine is already running but the state of the button on the interface says otherwise, then the subject would have to re-evaluate his/her interpretation of the message/format by going back to memory (3) and reformulating another mental model before execution (the message is an injunctive one and not an information statement), or by going back and re-reading the message (2) to double check what was initially read (on or off), and upon confirmation reformulate another mental model (the message is an injunctive one and not an information statement) before going through the steps again until execution.

The same reasoning would go for example 4a:

4a `Engine 1....................On` MCL Statement *(meant to inform participants that the engine is switched on)*

The path from reading the message to execution on the automaton will depend on the initial understanding by the subject of the format (injunction or statement) and how they re-evaluate the meaning when the pattern matching step on the interface fails (performing path (2) or (3)).

Example 3b:

3b `Turn On Engine 1`] MNL Injunction

and Example 4b:

4b `Engine 1 is On`] MNL Statement

are written in MNL in order to eliminate the double interpretation in the elliptical format of MCL. Therefore, the path from reading the message to the execution on the automaton should be relatively straightforward after the pattern matching step (performance of the correct path (1)) since the interpretation of the message and the developed mental model should correctly match the expected button state on the interface (in 3b the button should be initially disactivated and ready to be activated, in 4b the button should be initially activated and subject should press on the confirm button below and confirm that the statement is correct because the engine is on).

This demonstration of cognitive processes involved in the use of procedural texts could give further support to our initial hypothesis (which we further discuss in Section 6.3) that the linguistic economies in the elliptical format hinder fast and direct comprehension.

6.2.6 Unfolding of the Task

Before beginning the experiment, subjects filled out different forms: a general ethics and compliance consent form, a data sheet in which they specified their age, gender, familiarity with the Airbus Controlled Language, whether they are a pilot/piloting qualifications/hours/, native language and English placement. Participants were not remunerated for their participation, but were offered an 8Gb USB key for which they signed a receipt.

All non-native English speakers also performed a quick English placement test online[35] to determine their CEFR levels (Common European

35. Exam English. Test your level of English Grammar and Vocabulary. [online] Examenglish.com. Available at: https://www.examenglish.com/leveltest/grammar_level_test.htm [Accessed 9 Dec. 2018].

Framework of Reference for Languages). The levels range from A1 or breakthrough/beginner to C2 or Mastery/Proficiency. This will eventually help us determine whether English placement levels had an effect on our hypotheses.

And lastly, they filled out a small questionnaire:

Outside of Airbus context, what do those sentences mean to you?
Please explain on the lines below.

Engine 1 ....On

____

Speed ....250 KT

____

Figure 53. Intention Questionnaire

This questionnaire was added because we wanted to see, mostly for (but not limited to) people who were not familiar with the Airbus Controlled Language and its injunction rules, how participants interpret the theme…..rheme format more instinctively before they perform the task, whether they more naturally lean towards injunction or statement interpretations.

After the different papers were filled, the participants put on headphones (Beyerdynamic DT990 Pro 250 OHM) and faced the touch screen in a quiet isolated room.

As the pre-tests showed, the interface was not easily memorized and the buttons were hard to find, therefore we proceeded to include a familiarization phase where we introduced the interface in details, and zone by zone to each participant while briefly explaining the task. The familiarization included 7 slides introduced directly on the E-prime interface, starting with the global interface introducing all the zones, and then every zone separately, to finally end with the global interface again (cf. Figure 54).

Figure 54. Familiarization Process in 7 Steps

After the familiarization, the participants were presented with the official instructions (cf. Figure 55).

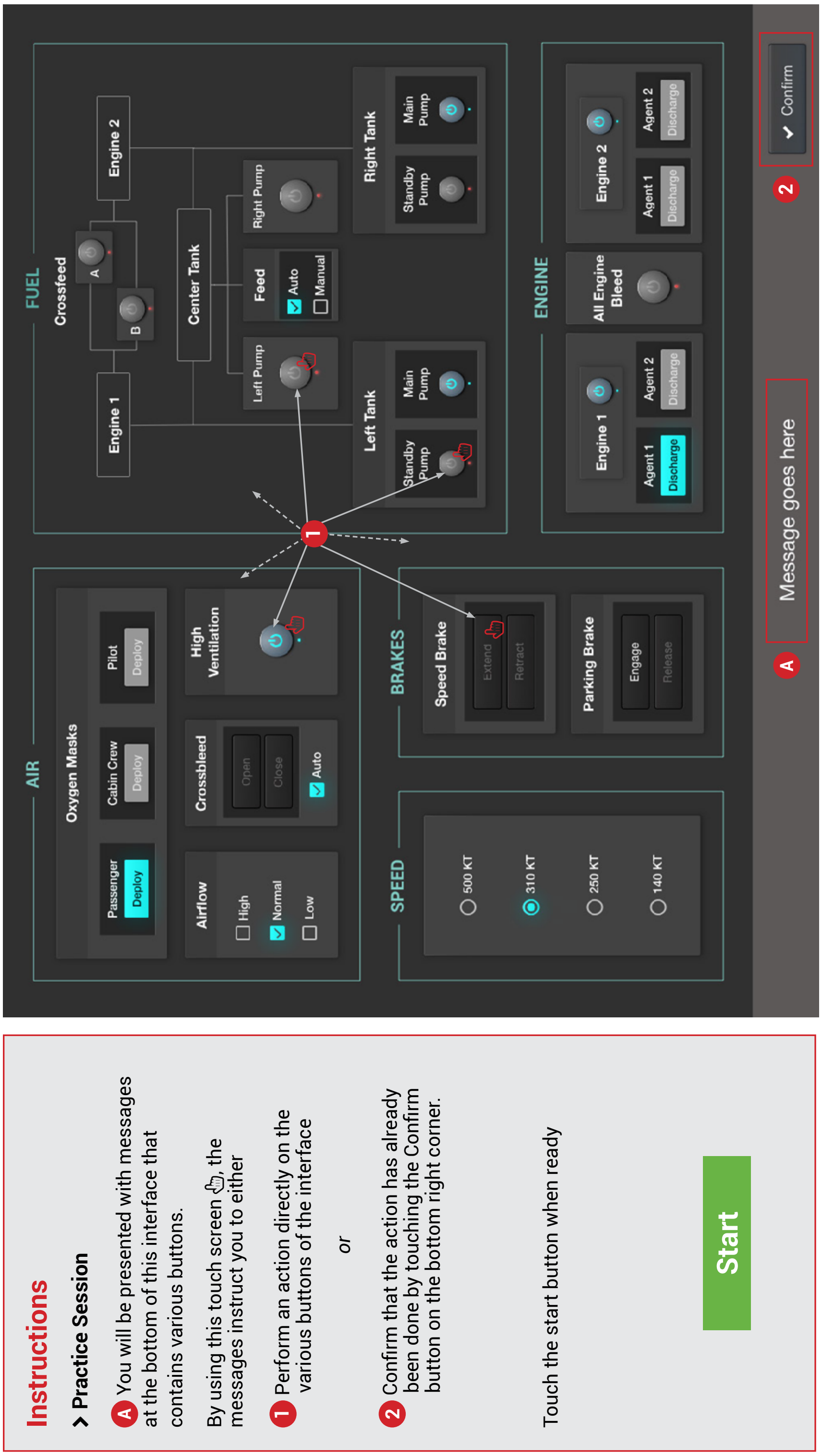


Figure 55. Interface with Parallel Task Instructions

While the participants read the instructions, we installed the SMI[36] eye tracking glasses and calibrated them with the gaze of every participant. The eye tracker was introduced because we were interested to see whether there was an influence of the language factor (MCL and MNL) on time spent on messages and specific areas of interest, on fixations, on revisits, and on the resulting gaze path. Generally speaking, we wanted to see if the language factor influenced the way we read and responded to an injunction or statement to answer a few questions (how many times do we read a message, do we look at the message and locate the button, and then come back to verify what we read, does one language enforce that more than the other? Etc.). Unfortunately, the analysis of the eye tracking data will not be part of this thesis because of time limitations, but would be an object of future investigation.

Participants were then orally informed that they would have a brief practice session. The practice session had 10 randomized stimuli (with respect to language, intention, and zone) in the non-urgent condition where participants had 12000ms to respond. Practice stimuli were not repeated in the target phase (they were specifically elaborated for the practice session).

Afterwards, they were presented with a message informing them that the practice session was over and that they can hit the start button whenever they are ready.

48 equally randomized stimuli then ensued in the non-urgent condition. The background behind the square interface was light gray in all of the non-urgent stimuli.

Should the participants not respond after 12000ms have passed, they would be presented with a "time's up" picture (cf. Figure 56) that we also designed using Adobe InDesign, before the next stimulus appears. As mentioned earlier, in case of a correct answer, as well as having responsive buttons, a correct ding sound plays (native to E-prime) in the headphones. In case of an incorrect answer, a wrong answer buzzer sounds[37].

The sound is also accompanied by a wrong answer image (cf. Figure 57) that was also designed using Adobe InDesign.

36. SMI: SensoMotoric Instruments

37. "error.wav" downloaded from Freesound.org available at: https://freesound.org/people/Autistic%20Lucario/sounds/142608/ [Accessed 9 Dec. 2018].

Figure 56. Time's Up Image

Figure 57. Wrong Answer Image

After 48 non-urgent randomized stimuli have been displayed, the task is interrupted and a slide appears informing the participants that they are about to start the second phase with the following words: "You will now perform the same task but you must respond as urgently as possible as you have less time." Participants then have to go through the urgent phase consisting of 10 practice and 48 randomized stimuli with a red background. At the end of this phase, participants were thanked for their participation and were asked a few questions that we will discuss later in Section 6.4.1. Participants were asked whether some messages were easier than others and why? Did they have any preference? Did they notice two different message formats? Did they have any particular difficulties with the interface, or anything else? Did they find the second phase stressful? They were finally asked if they had any other comments before the experimenter explained briefly the goal of the experiment, and asked if they have further comments.

6.2.7 Participants

- 140 participants took part in this experiment, their ages ranged from 19 till 63, 37 females and 103 males.
- 41 native english speakers and 99 non-native english speakers. Among those, 77 were french native speakers, 2 Arabic, 1 Chinese, 4 German, 2 Italian, 1 Kannada, 2 Portuguese, 1 Romanian, 1 Russian, 1 Serbian, 7 Spanish native speakers.
- 3 non-native participants have scored A2 on the english placement, 3 have B1, 19 have B2, 29 have C1, and 45 have a C2 placement level.
- Since a major part of our participants come from in and around the aviation industry (Airbus employees/pilots/student pilots from ENAC[38] school etc.) their familiarity with the Airbus controlled language needed to be assesed to see whether familiarity is a factor that influences our hypotheses. 41 participants had no familiarity with Airbus Controlled Language. 29 were beginners or had seen or heard of the language/or a similar language before, but do not necessarily know the rules. 39 had intermediate familiarity with the language or had worked with this or a similar language and know some rules. 11 were expertly familiar with the language because they worked or work on the language on an almost daily basis; those participants were predominantely flight warning and human factors linguistic experts, as well as Airbus test pilots.

38. Ecole Nationale de L'Aviation Civile. We would like to thank Julie Saint-Lot for her help with the organization of experimentation sessions with student pilots and private license pilots.

- 37 participants have some sort of piloting experience, ranging from student pilots with a few hours of flight (11), to private pilots and flight instructors (19), to test and fighter pilots (5), etc. Some of the participants belonged in several categories, for example they could still be student pilots and yet have a private pilot licence, etc.

6.3 Research Questions and Hypotheses

We will start by listing our hypotheses and research questions and then proceed to show the results and analysis regarding those questions.

Our metrics and dependent variables were reaction times in ms. for correct answers, and an accuracy average with a score ranging between 0 for errors or non-answers, and 1 for correct answer.

The list of independent variables that we will evaluate are:

- Language (MCL-MNL)
- Urgency (4000ms, 12000ms)
- Intention (injunction, statement)
- Complexity of zone (air, speed, brakes, fuel, engine)

Extraneous and participant variables:

- English placement level (Basic Intermediate, Proficient, Mastery, Native)
- Familiarity with Airbus CL (None, Beginner, Intermediate, Expert)
- Pilot Experience vs no Pilot experience.

Hypotheses:

1. MNL messages produce shorter reaction times than MCL ones in the urgent and non-urgent conditions.
2. MNL messages produce less errors (are more accurate) than MCL ones in the urgent and non-urgent conditions.

Research questions:

3. Did the language factor have an effect on reaction times or accuracy in more complex zones?

4. Given that the MCL format could have a double interpretation, a statement or an injunction, we would like to objectively measure which interpretation is more accurately understood by comparing accuracy averages for both conditions. Was the accuracy average superior for MCL_Statement or MCL_Injunction?
5. Did the language factor play a different role for different levels of English placement (Basic Intermediate, Proficient, Mastery, Natives) regarding accuracy?
6. Did the language factor play a different role for different levels of familiarity with Airbus Controlled language regarding accuracy?

6.4 Results and Analysis

To statistically analyze our results, we used SPSS[39] Statistics after cleaning and preparing the data in Microsoft Excel. In the first hypothesis' analysis, we will give more statistical details and show the methodology using tables and relevant values. We will not be providing this level of detail in the ensuing analyses (hypothesis and research question 2 till 6) but will summarize the statistical end results directly. We will then recapitulate and discuss all the results in the ensuing discussion section.

1. **MNL messages produce shorter reaction times than MCL ones in the urgent and non-urgent conditions.**

Because of the large number of participants that we managed to recruit in this experiment, we are going to use parametric tests that have proved to be robust to the normality assumption with regards to reaction times and accuracy data.

Firstly, we need to verify that the following ANOVA assumptions are true. These assumptions let us see if we are allowed to use a general linear model and the F statistic (used by ANOVA) on our data:

a. **Independent observations** (or, more precisely, independent (a variable is not dependent on the result of another) and identically distributed variables).

39. Ibm. SPSS Statistics - Overview. [online] Available at: https://www.ibm.com/products/spss-statistics [Accessed 11 Dec. 2018].

b. **Normal distribution:** The test variables follow a **multivariate normal distribution** in the population. However, this assumption is not needed if the sample size >= 25, which is the case in our sample. Additionally, as Figure 58 shows, the histograms look plausible and the data does not show abnormal skew.

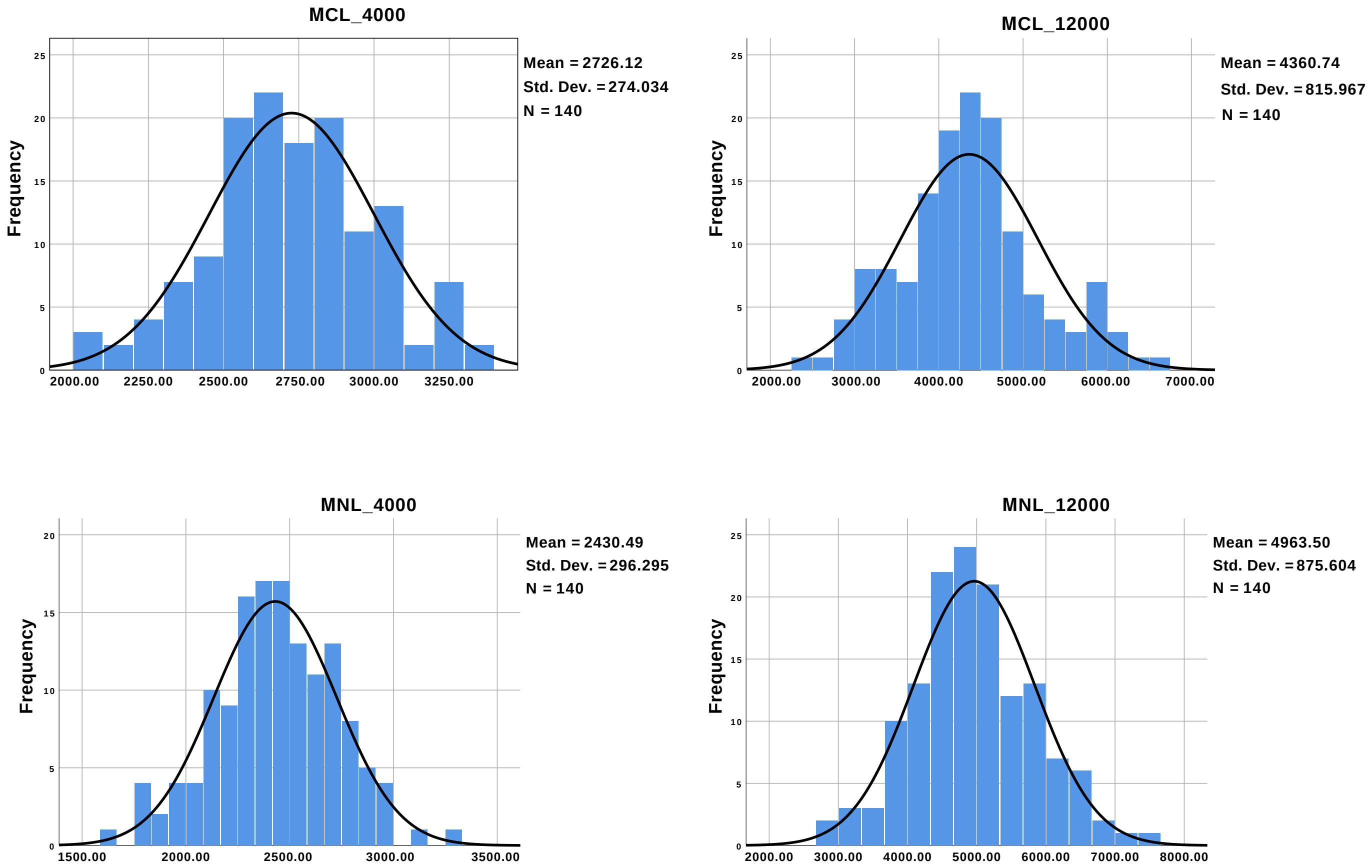


Figure 58. Data Distribution Histogram Showing Normality Skew for the Different Language and Urgency Conditions

c. **Sphericity**. This means that the population variances of all possible different scores are equal. Sphericity here will be tested with Mauchly's test to see whether the variances are significantly unequal, in which case we use corrections such as Greenhouse-Geisser's (cf. Figure 59).

Assumptions 1 and 2 are not rejected. Assumption 3 will be tested for using Mauchly's test of Sphericity and potentially corrected if needs be.

We performed a within-subjects general linear model analysis (cf. Table 9) to test for significance between the different factors. First, we look at the descriptive statistics (cf. Table 7) to verify the means and standard deviation, then Mauchly's test (cf. Table 8) to verify if the data are significantly non-spherical:

	Mean	Std. Deviation	N
MCL_4000	2726.1246	274.03352	140
MNL_4000	2430.4928	296.29470	140
MCL_12000	4360.7417	815.96687	140
MNL_12000	4963.5029	875.60448	140

Table 7. Means of Reaction Times in Different Language and Urgency Conditions

Within Subjects Effect	Mauchly's W	Sig.
Language Urgency	.146	.000

Table 8. Mauchly's Test of Sphericity Measure: Language and Urgency

As we can see in Table 8, the data are significantly non-spherical. Therefore, as Figure 59 suggests, we must use the Greenhouse-Geisser correction as the p-value is inferior to 0.75.

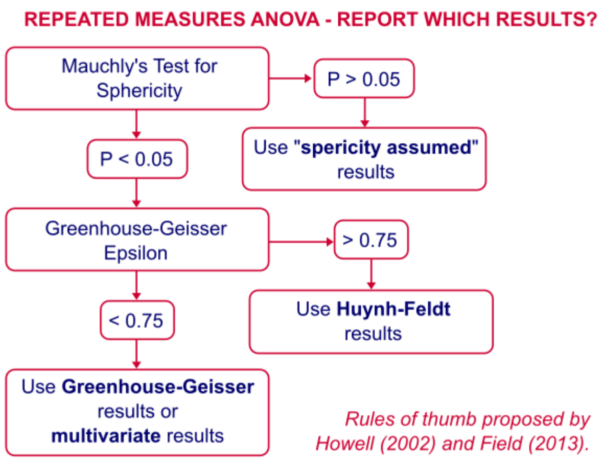


Figure 59. Flow Chart for Sphericity Correction Taken from SPSS Tutorial (itself adapted from Howel 2002 and Field 2013)[40].

40. Spss Tutorials. SPSS Repeated Measures ANOVA. [online] Available at: https://www.spss-tutorials.com/spss-repeated-measures-anova-example-2/comment-page-1/ [Accessed 2 Dec. 2018].

As we can see in Table 9, the difference between means is statistically significant with an effect size calculated with partial regression coefficient Eta Squared (η^2). We apply the following formula by taking the Greenhouse-Geisser corrected values from Table 9 below:

F(degree of freedom(df), degree of Freedom_Error(df)) = F statistic(F), Significance (p):

$$F(1.755, 243.949) = 1085.708, \ p = 0.000, \ \eta^2 = 0.887$$

Source		df	F	Sig.	Partial Eta Squared
Language Urgency	Sphericity Assumed	3	1085.708	.000	.887
	Greenhouse-Geisser	1.755	1085.708	.000	.887
	Huynh-Feldt	1.776	1085.708	.000	.887
	Lower-bound	1.000	1085.708	.000	.887
Error (Language Urgency)	Sphericity Assumed	417			
	Greenhouse-Geisser	243.946			
	Huynh-Feldt	246.818			
	Lower-bound	139.000			

Table 9. ANOVA Results of Within-Subjects Effects concerning Reaction Times in the Different Language and Urgency Conditions

We can see the results graphically on the bar chart below (cf. Figure 60) with error bars at the 95% confidence interval.

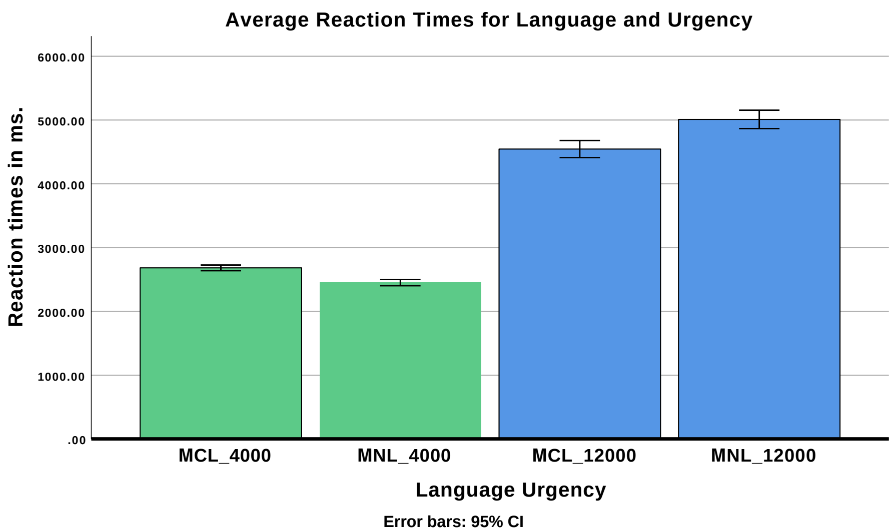


Figure 60. Bar Chart of Mean Reaction Times in the Different Language and Urgency Conditions

Therefore, we proceeded to do post hoc tests to see if the language means in every condition of urgency were significantly different. The pairwise comparison Table 10 shows that in each condition of urgency (MCL_4000 and MNL_4000) and (MCL_12000 and MNL_12000) the two-language means were significantly different.

(I) Language Urgency	(J) Language Urgency	Sig.
MCL_4000	MCL_12000	.000
	MNL_4000	.000
	MNL_12000	.000
MCL_12000	MCL_4000	.000
	MNL_4000	.000
	MNL_12000	.000
MNL_4000	MCL_4000	.000
	MCL_12000	.000
	MNL_12000	.000
MNL_12000	MCL_4000	.000
	MCL_12000	.000
	MNL_4000	.000

Table 10. Pairwise Comparisons for Reaction Times in Each Urgency Condition

In order to verify if there is an interaction between the conditions or language (MCL-MNL) and urgency (4000ms-12000ms), we proceeded to do a two by two factorial analysis of variance.

As the sphericity assumption only needs to be verified when there are more than two measurements, and since there are only two levels of repeated measures here, we'd only need to ascertain a homogeneity of variance, using Levene's test for example. However, we can safely disregard this assumption here since we have equal sample sizes.

Source		df	F	Sig.	Partial Eta Squarred
Language	Sphericity Assumed	1	38.193	.000	.216
	Greenhouse-Geisser	1.000	38.193	.000	.216
	Huynh-Feldt	1.000	38.193	.000	.216
	Lower-bound	1.000	38.193	.000	.216
Error (Language)	Sphericity Assumed	139			
	Greenhouse-Geisser	139.000			
	Huynh-Feldt	139.000			
	Lower-bound	139.000			
Urgency	Sphericity Assumed	1	1477.415	.000	.914
	Greenhouse-Geisser	1.000	1477.415	.000	.914
	Huynh-Feldt	1.000	1477.415	.000	.914
	Lower-bound	1.000	1477.415	.000	.914
Error (Urgency)	Sphericity Assumed	139			
	Greenhouse-Geisser	139.000			
	Huynh-Feldt	139.000			
	Lower-bound	139.000			
Language* Urgency	Sphericity Assumed	1	310.184	.000	.691
	Greenhouse-Geisser	1.000	310.184	.000	.691
	Huynh-Feldt	1.000	310.184	.000	.691
	Lower-bound	1.000	310.184	.000	.691
Error (Language* Urgency)	Sphericity Assumed	139			
	Greenhouse-Geisser	139.000			
	Huynh-Feldt	139.000			
		139.000			

Table 11. Tests of Within-Subjects Effects Tests of ANOVA Within-Subjects Effects of Reaction Times for Language Conditions and Interactions

As we can see from the 5th row of Table 11 and the line chart below (cf. Figure 61), there is a significant interaction between the two conditions: $F(1, 139) = 310.184$, $p = 0.000$, with an effect size for the language factor $\eta^2 = 0.216$, and that of the urgency factor $\eta^2 = 0.914$.

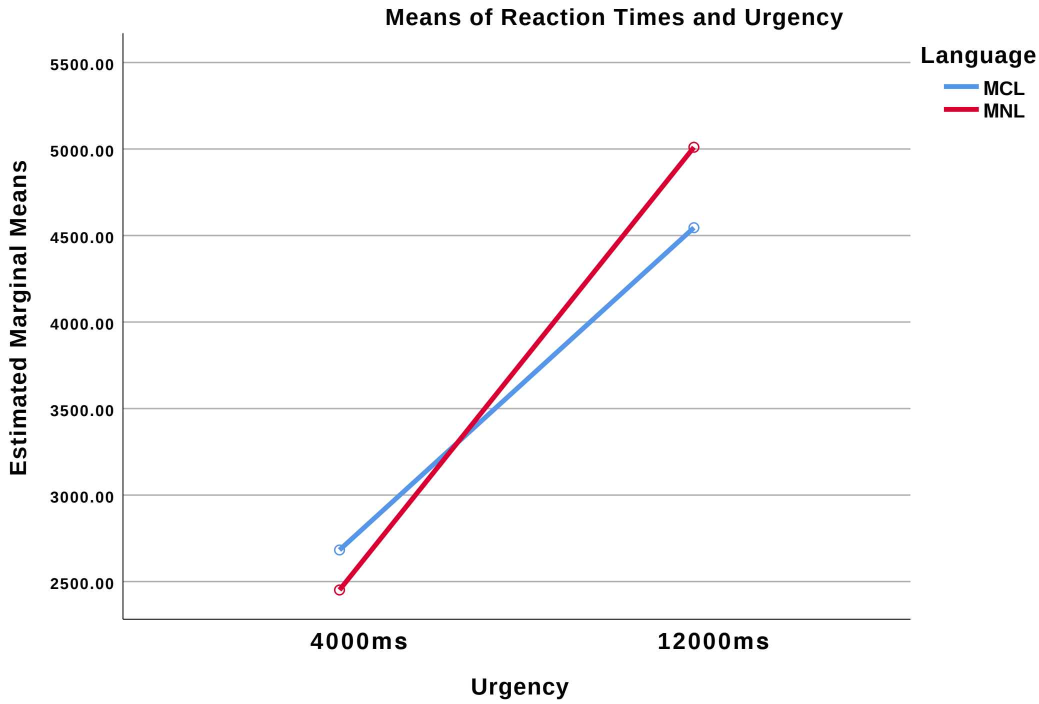


Figure 61. Line Chart of Reaction Time Interactions Between Language and Urgency

This means that the language effect (one language being more efficient than the other) did not behave in the same way in both urgency conditions.

Therefore, as we see in the means Table 12 the MNL_4000 average reaction times = 2430.5 ms is significantly shorter than the MCL_4000 average = 2726.13 ms, and conversely, the MNL_12000 reaction time average = 4963.5 ms is significantly longer than the MCL_12000 average = 4360.74 ms.

MCL_4000	> MNL_4000	MCL_12000	< MNL_12000
2726.12ms	> 2430.49ms	4360.74ms	< 4963.50ms

Table 12. Reaction Time Means of Language in Different Urgency Conditions

In other words, in the urgent condition the more natural controlled language had significantly faster reaction times but in the less urgent condition the more coded controlled language had significantly faster reaction times, therefore, we have a statistically significant interaction between the two conditions. We therefore reject the null hypothesis regarding the urgent condition as MNL has produced faster reaction times than MCL, and the difference in means is statistically significant as we predicted. However, this is not the case in the non-urgent condition where MCL produced faster reaction times contrary to what we hypothesized.

These results seemed unusual but not altogether surprising. When participants were pressed for time as they had 3 x less time to fulfill the task (4000ms), the more natural controlled language structures helped them perform faster. We conclude that the more natural controlled language enabled participants to *process* information faster and consequently *react* faster in the more urgent condition where participants were stressed because of the lack of time, and to a certain extent the red background color. These results are in line with the results from our previous experiments, i.e. MNL is more efficient when time, or a lack thereof, is a factor. This was not the case in the non-urgent condition where participants had 3 x more time to respond (12000ms) accompanied with a gray background color. In this condition, the more coded controlled language helped participants react faster.

This result may be explained by the fact that the participants in the non-urgent phase were not instructed to answer fast, and were for the most part unaware that they had, albeit ample, but nevertheless limited time to respond. Since the MNL messages always had more words than the MCL, it is unsurprising that when participants have all the time they need to perhaps re-read and double check their more (MNL) or less (MCL) long message, they would respond faster when they have less to read (MCL) and nothing urgent or stressful to respond to. What is even more interesting to evaluate however, is how accurate those responses were, and does the same logic apply to accuracy of responses in the relationship between language and urgency?

2. **MNL messages produce less errors (are more accurate) than MCL ones in the urgent and non-urgent conditions.**

This brings us to our second hypothesis regarding accuracy. We will start by observing the descriptive statistics in Table 13:

	N	Minimum	Maximum	Mean	Std. Deviation
MCL_4000	140	.29	1.00	.6996	.15650
MNL_4000	140	.46	1.00	.8470	.11244
MCL_12000	140	.58	1.00	.9178	.08715
MNL_12000	140	.50	1.00	.9342	.07491

Table 13. Means of Accuracy in the Different Language and Urgency Conditions

The means for accuracy were calculated by making an average score ranging from 0 for an error and no answer and 1 for correct answer. As we can see in Table 13, the accuracy mean of the MNL condition in the urgent condition is superior 0.84 (therefore more accurate) than the MCL condition 0.69. It is also the case in the non-urgent condition, MNL is more accurate than MCL: $0.93 > 0.91$. We will now perform an analysis of variance to determine whether this observed difference in the means is statistically significant.

As before, we firstly verified the ANOVA assumptions and proceeded in the same manner. All the assumptions were held. Note that ANOVA assumptions were always tested for and corrected when needed, but in order to avoid redundancy we will not mention them for every analysis.

A one factor with 4 levels ANOVA shows that the observed difference in means is significant:

$F(2.342, 325.485) = 228.326$, $p = 0.000$

A post hoc analysis shows that the difference in means between MCL_4000 and MNL_4000 is significant, $p = 0.000$, and the difference in means between MCL_12000 and MNL_12000 is significant, $p = 0.027$. The MNL is significantly more accurate than its MCL counterpart in both urgency conditions. We can therefore reject the null hypothesis and confidently validate our hypothesis for a true observed difference in means.

We can see that graphically in the bar chart (cf. Figure 62):

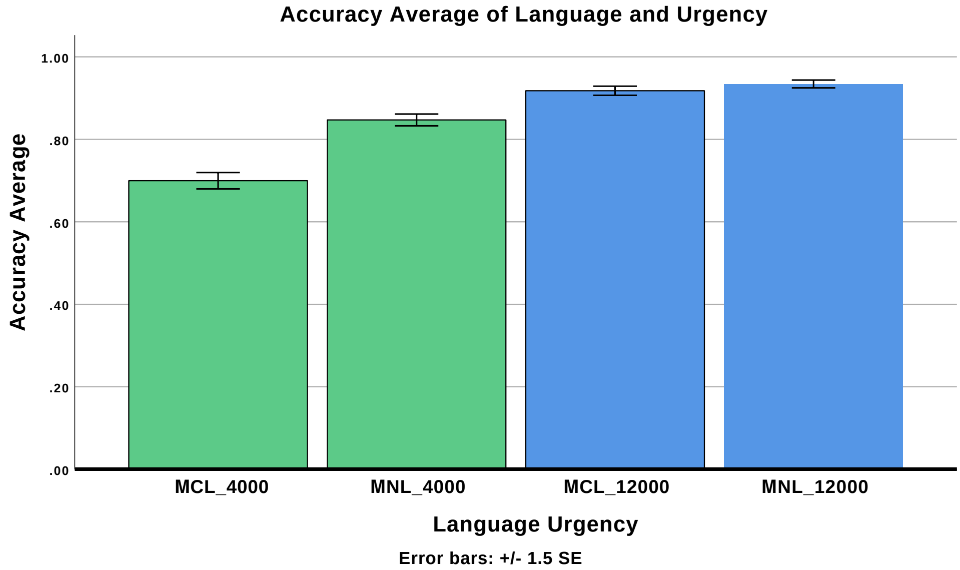


Figure 62. Bar Chart of Accuracy for Different Language and Urgency Conditions

We then performed an 2x2 factors ANOVA to verify if there is a significant interaction between the two languages in the two urgency conditions. The ANOVA showed a significant interaction between Language * Urgency: $F(1, 139) = 103.465$, $p = 0.000$

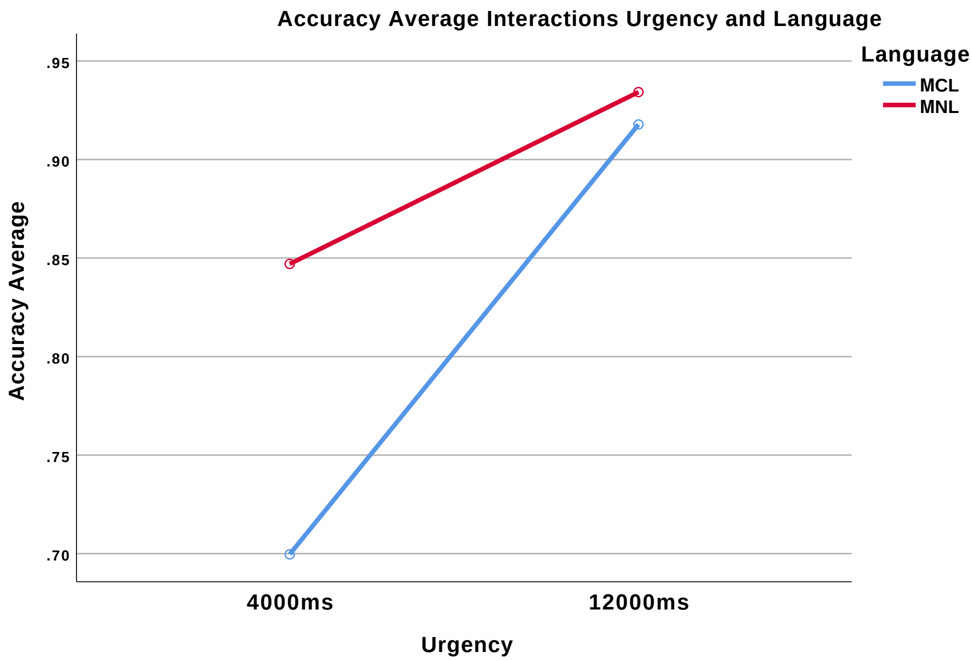


Figure 63. Line Chart of Accuracy Interactions Between Language and Urgency

As illustrated in Figure 63, the linear models representing the two languages' average accuracy in the two urgency conditions are far from parallel. While the MNL is significantly more accurate than the MCL in both conditions, the interval between the difference in means of the two languages shrinks in the non-urgent condition. Be that as it may, the MNL is still significantly more accurate even when the participants had 3 x more time to respond, unlike reaction times, which actually show better efficiency for the MCL condition in the non-urgent condition. Therefore, even though the MNL messages had more words which presumably made participants react slower in the seemingly unlimited time, the MNL messages proved to significantly guarantee more accuracy no matter the urgency (cf. Figure 64). The interaction we observe (the MCL gaining accuracy with more time) is most probably due to the fact that more time will naturally produce better accuracy in both languages, because participants will have more time to process the information and respond. However, there is an inherent limit to how much better accuracy could get in an experiment such as this (which could typically only be remedied by more practice and training). Therefore, since MNL started off with a decent accuracy score in the urgent condition, the expected improvement provided by 3 x more time to respond is not as conspicuous as it is with the MCL which started off with a relatively low accuracy score and improved significantly with more time given.

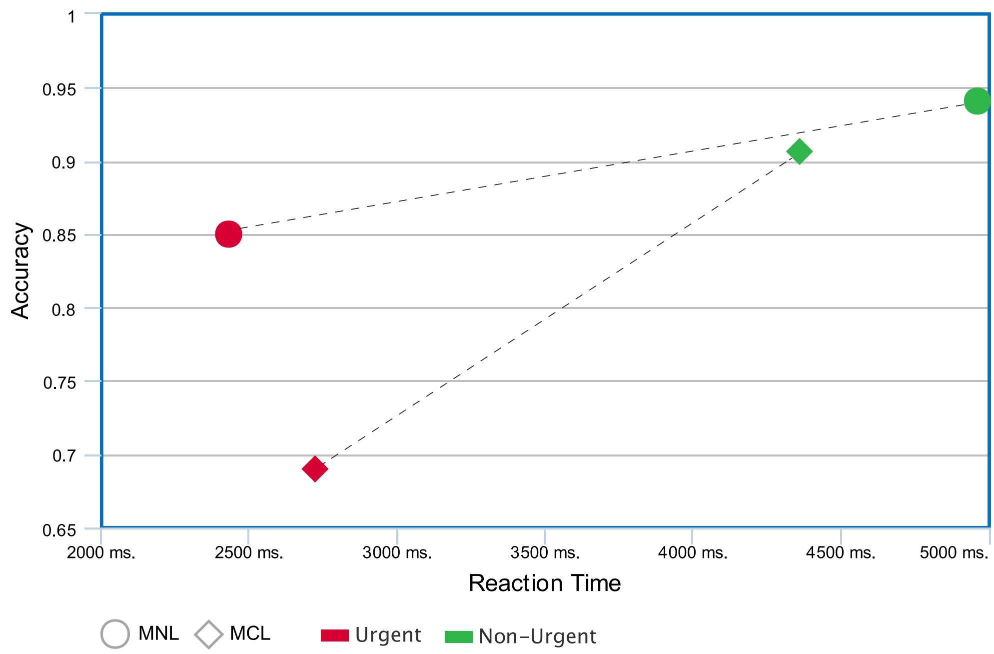


Figure 64. Scatter Plot Showing Average Accuracy as a Function of Reaction Times in the Different Language and Urgency Conditions

Figure 64 shows how MNL (the circle shape) is consistently (and significantly) more accurate than MCL (the diamond shape) in both urgency conditions (red for urgent and green for non-urgent). We can also see that MNL has significantly faster reaction times in the urgent condition but not in the non-urgent condition where MCL has significantly faster reaction times (because of the significant interaction between the two conditions of urgency).

We also worked on three different measures for accuracy by counting the numbers of errors per participant in the different conditions, the number of non-answers per participant in the different conditions of urgency, and the number of both errors and non-answers combined for every participant in the different urgency conditions.

On the 808 total errors, 307 errors were made in the MCL_4000 condition (average of 2.2 per participant), almost double the number of errors in the MNL_4000 condition, 114 (average of 0.8 per participant). And in the non-urgent condition the MCL_12000 had 217 errors (average of 1.6 per participant) and the MNL_12000 had 170 errors (average of 1.2 per participant) (cf. Table 14).

MCL_4000	MNL_4000	MCL_12000	MNL_12000	Total Errors
307	114	217	170	808

Table 14. Number of Errors in the Different Language and Urgency Conditions

When we performed an analysis of variance regarding the average number of errors for different participants in the different conditions, we observed that the differences in means (cf. Table 15) were significant: F(2.756, 383.105)=27.738, $p = 0.00$, and the post hoc tests show that the pairwise comparisons are equally significant $p = 0.00$ for MCL_4000 and MNL_4000, and $p = 0.025$ for MCL_12000 and MNL_12000. The bar chart in Figure 65 highlights the differences graphically.

	Mean	Std. Deviation	N
MCL_4000	2.1929	1.93006	140
MNL_4000	.8143	1.11617	140
MCL_12000	1.5500	1.78855	140
MNL_12000	1.2143	1.52573	140

Table 15. Means of Error Averages in Different Language and Urgency Conditions

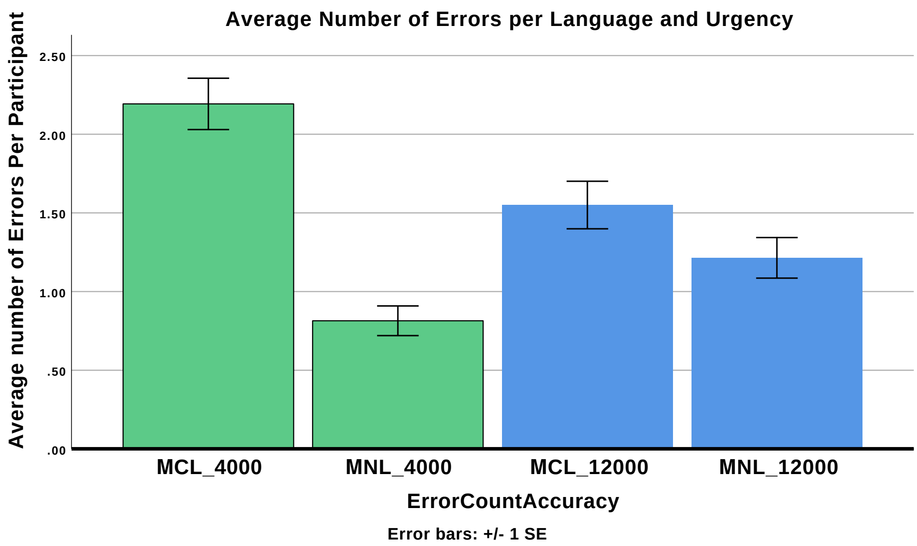


Figure 65. Bar Chart Showing Average Number of Errors in Different Language and Urgency Conditions

The non-answer average was also interesting to analyze, especially for the urgent condition as we would assume that if one has less words and less elements to read and to process, then one could respond faster and within the time limit. On the 1229 total non-answers, 706 were for the MCL_4000 (average of 5.07 per participant) condition and 403 for the MNL_4000 (average of 2.9 per participant), and 64 for the MCL_12000 (average of 0.46 per participant) and 56 for the MNL_12000 (average of 0.403 per participant) (cf. Table 16 and Table 17).

MCL_4000	MNL_4000	MCL_12000	MNL_12000	Total Non-Answers
706	403	64	56	1229

Table 16. Number of Non-Answers in the Different Language and Urgency Conditions

	N	Minimum	Maximum	Mean	Std. Deviation
MCL_4000	140	.00	16.00	5.0791	3.06468
MNL_4000	140	.00	13.00	2.8993	2.45627
MCL_12000	140	.00	6.00	.4604	.95755
MNL_12000	140	.00	3.00	.4029	.67801

Table 17. Means of Non-Answer Averages in Different Language and Urgency Conditions

When we performed an analysis of variance regarding the average number of non-answers for different participants in the different conditions, we observed that the differences in means were significant: $F(2.052, 283.126)= 240.804$, $p = 0.000$. We could graphically observe those differences in the bar chart in Figure 66.

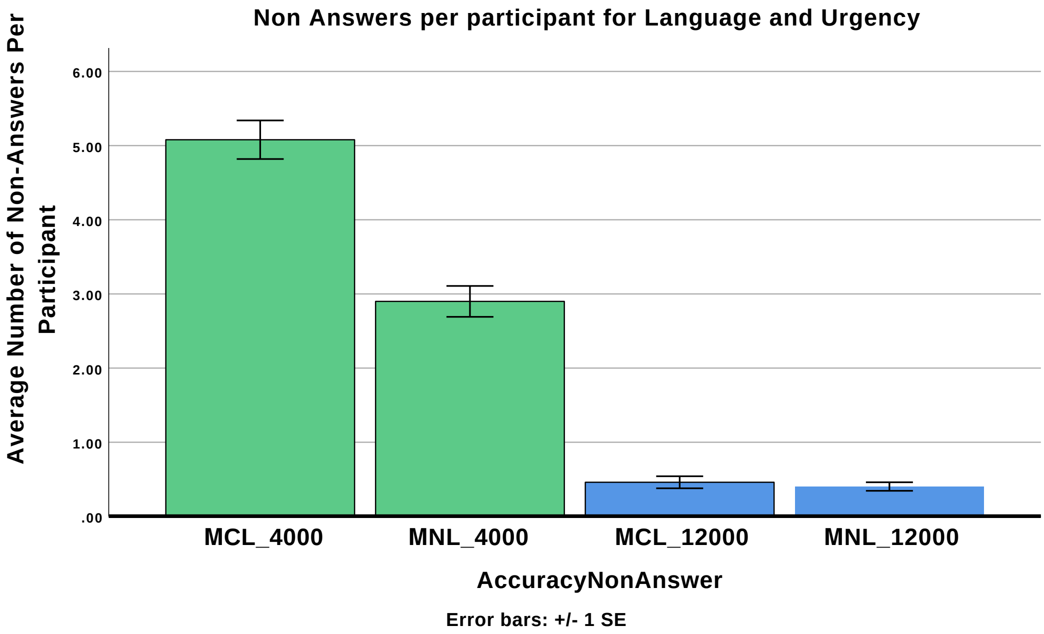


Figure 66. Bar Chart Showing Average Number of Non-Answers in Different Language and Urgency Conditions

The post-hoc tests showed that the difference in means between MCL_4000 and MNL_4000 was significant, $p = 0.000$, however the difference between MCL_12000 and MNL_12000 was not significant. This is an interesting result as it goes with the logic prevailing so far in these analyses. The more time the participants have to answer, the less conspicuous the difference between the language processing is. This is why the average of non-answers is not significant in the non-urgent condition which is not surprising given how close the means are to one another (64 non-answers for MCL vs. 56 non-answers for MNL). Non-answers are noteworthy because they could stand for participants either not having enough time to respond or not understanding what they must do, so preferred not to answer (although that was never in any instructions from our part). Additionally, if we take the example of the urgent condition, participants had 4000ms to answer for messages containing either MCL and

MNL. Since language was the only altered variable in a given condition of urgency, we could safely conclude that the coded format is chiefly responsible for the lack of responses. Moreover, the urgent condition of 48 stimuli always chronologically occurred as a second phase, i.e when participants had already gone half way through the experiment and sufficiently learnt how to perform the task and are sufficiently familiar with the interface and the various buttons. This allows us to exclude the explanation that the participants did not understand what needed to be done, but rather they could not interpret the meaning of the coded format in the MCL messages in time to accurately respond.

What is even more interesting, is the glaring discrepancy between the MCL_4000 and MNL_4000 non-answer average. This result is not anticipated because of the misconception that fewer words (as is the case in the MCL condition) will unavoidably lead to faster reading and faster processing and response. This shows that the natural language structure, while being inevitably more verbose, by its syntax alone, provides a structure that facilitates information processing. The fact that we have almost double the amount of non-answers in the MCL condition, in the stressful limited time phase, proves that even though the structure is deemed more simplified and easier to process, it actually is more complex to decode by the brain.

Predictably, when we calculated the averages of both non-answers and errors, we obtained similar results. On the total of 2037 errors and non-answers, 1013 belonged in the MCL_4000 (average of 7.24 per participant) and 281 in the MNL_4000 (average of 3.7 per participant), 517 in MCL_1200 (average of 2 per participant) and 226 in the MNL_12000 (average of 1.6 per participant) (cf. Table 18 and Table 19).

MCL_4000	**MNL_4000**	**MCL_12000**	**MNL_12000**	**Total Errors & Non-Answers**
1013	281	517	226	2037

Table 18. Number of Errors and Non-Answers in the Different Language and Urgency Conditions

	N	Minimum	Maximum	Mean	Std. Deviation
MCL_4000	140	.00	17.00	7.2357	3.75435
MNL_4000	140	.00	13.00	3.6929	2.67656
MCL_12000	140	.00	10.00	2.0071	2.08970
MNL_12000	140	.00	12.00	1.6143	1.80162

Table 19. Means of Errors and Non-Answer Averages in Different Language and Urgency Conditions

When we performed an analysis of variance regarding the average number of errors and non-answers for different participants in the different conditions, we observed that the differences in means were significant: $F(2.336, 324.647) = 229.232$, $p = 0.000$. We could graphically observe those differences in the bar chart in Figure 67.

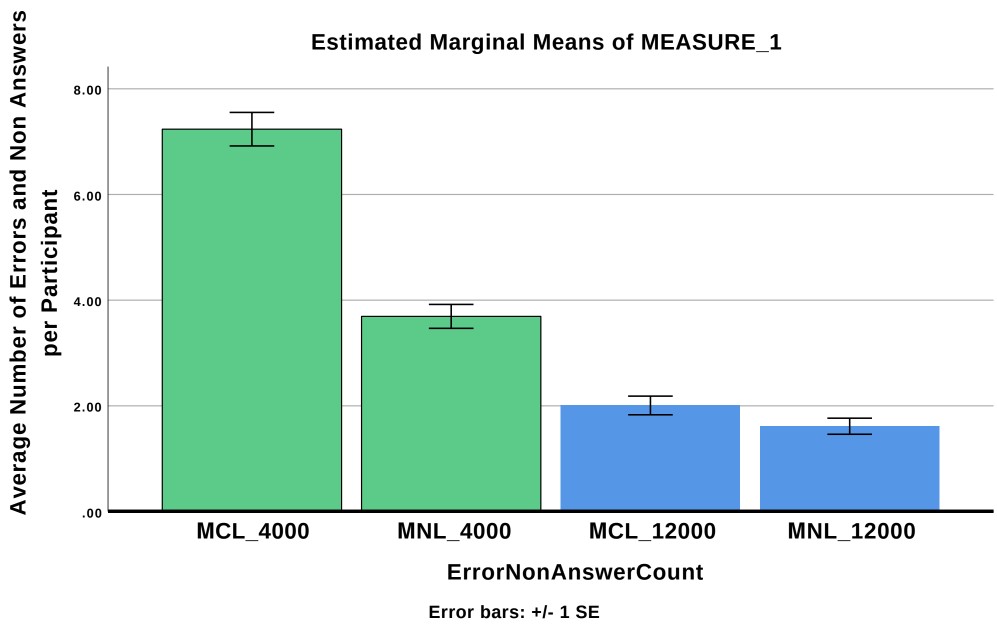


Figure 67. Bar Chart Showing Average Number of Errors and Non-Answers in Different Language and Urgency Conditions

The post hoc tests show that the means of MCL_4000 and MNL_4000 were significantly different $p = 0.00$ and the means of MCL_12000 and MNL_12000 were also significantly different $p = 0.00$.

To recapitulate on the results regarding the two main hypotheses:

- The MNL has proven to be significantly faster to process (faster reaction times) in the urgent condition, but not in the non-urgent condition.
- The MNL has proven to be significantly more accurate than the MCL in all urgency conditions.
- Errors and non-answer averages confirm this effect and highlight the shortages and the main strength of each language type.

3. Did the language factor have an effect on reaction times or accuracy in more complex zones?

This brings us back to our first research question. As we mentioned earlier, some zones were more complex than others as they contained more buttons and possible interactions. For instance, the fuel zone included 3 different tanks (left, right, and center) that each had two pumps (left and right) in two different engines (1 and 2). It also had two buttons of "crossfeed" A and B that were often confused with "crossbleed" buttons in the air zone. The latter had an "Airflow High" button that was confused with "High Ventilation" button. The engine zone had two engines that each had two fire agents and the combination of engines and agents were also perceived as difficult by some participants. The speed zone was by far the easiest as it as was straightforward and included only a simple interaction of choosing the right speed level. Additionally, the buttons in the speed zone did not have two possible positions such as on/off buttons. We will expand on these perceived difficulties in the participant debrief in Section 6.4.1.

In order to verify if some zones were truly more complex than others, we calculated the general accuracy average in different zones (cf. Table 20).

	N	Minimum	Maximum	Mean	Std. Deviation
Air	140	.550000	1.0000000	.8435	.103854293
Brakes	140	.500000	1.0000000	.9482	.132411326
Engine	140	.393939	1.0000000	.8489	.115804016
Fuel	140	.454545	1.0000000	.8114	.110256343
Speed	140	.833333	1.0000000	.9988	.014085904

Table 20. Means of Accuracy in Different Interface Zones

As we can see from the means in Table 20 and bar graph in Figure 68, the speed zone has almost maximum accuracy 0.998, whereas Fuel zone is least accurate 0.811. After performing an analysis of variance and subsequent post hoc analysis, we observed that these differences were all significant towards one another ($p < 0.01$) apart from Air and Engine zones whose means were not significantly different ($p = 0.573$). Therefore, if we consider that the harder a zone is the more errors participants would make, ergo the accuracy average decreases, then we can safely conclude that the true difference in means entails that the Fuel zone was the hardest, followed by the Engine and Air zones in second place, and Brakes and Speed being the easiest with a high accuracy score.

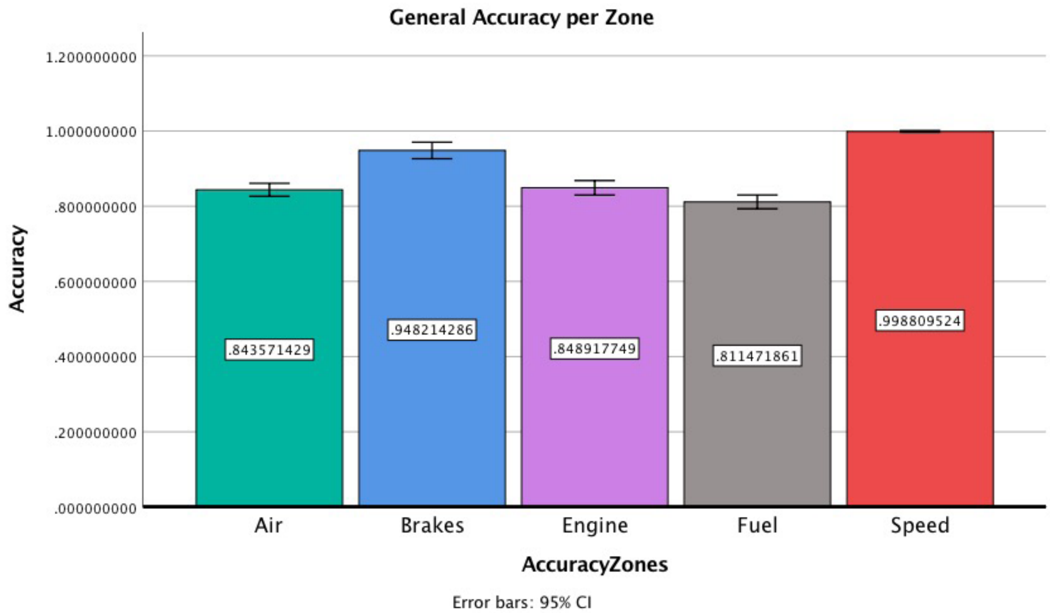


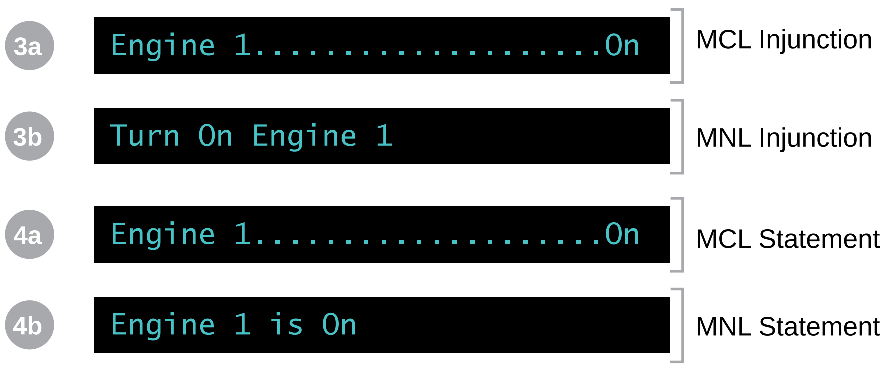
Figure 68. Bar Chart of Accuracy in Different Interface Zones

Furthermore, it is interesting for us to see whether one of the two languages (MCL or MNL) in each of the urgency phases had an effect on reaction times and accuracy in more complex zones. In previous experiments, we concluded that the more natural controlled language had a more marked effect in more complex or difficult situations. Therefore, we wanted to compare the effect of language on easy zones like the Speed zone and more complex zones like the Fuel zone.

Our analyses show that the language factor did not have a different effect in different zones. The general effect observed in the previous sections is also observed across different zones of different levels of complexity. In other words, the MNL was also significantly more accurate than MCL across different urgency conditions, and MNL had significantly faster reaction times than MCL in the urgent condition but not in the non-urgent condition across all zones alike.

4. **Given that the MCL format could have a double interpretation, a statement or an injunction, we would like to objectively measure which interpretation is more accurately understood by comparing accuracy averages for both conditions. Was the accuracy average superior for MCL_Statement or MCL_Injunction?**

As we mentioned earlier in this chapter, the theme…..rheme format is particular to the injunctive messages in the original controlled language format, but as it might be subject to different possible interpretations in some cases ("switch off the engine." vs. "the engine is switched off."), we would like to evaluate through this experiment, which of the two interpretations does the format more naturally evoke, injunction or statement, 'turn off' or 'is off'. Therefore, we compared the accuracy of responses when participants responded to MCL messages in the statements and MCL messages in the injunctions, i.e., when the theme…..rheme format was intended to give an order to perform an action on a button (3a) or when it intended to be informative and require a simple confirmation by touching the confirm button on the screen (4a):



A paired sample t-test was conducted to compare the accuracy of MCL responses in the injunctive condition and in the statement condition. There was a significant difference in the scores for Injunction ($M = 0.78$, $SD = 0.13$) and Statement ($M = 0.82$, $SD = 0.1$) conditions; $t(139) = -4.046$, $p = 0.000$ (cf. Figure 69).

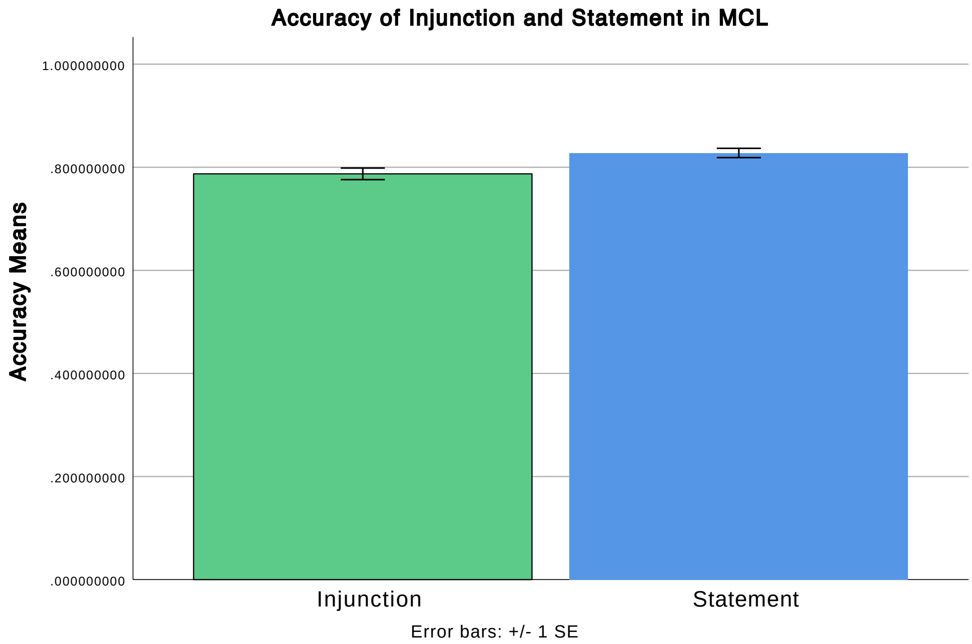


Figure 69. Bar Chart Showing Accuracy of Comprehension of Injunctions and Statements in MCL Condition

We can conclude that the theme.....rheme format is more accurately (therefore more frequently and better) understood as a statement, and significantly so. This is perhaps not so surprising, as in most cases the participle standing in the position of the rheme is a description of a state of a button: off-on-high-low etc. which is more difficult to be understood as an injunction, and more easily understood as a description. This is currently remedied by learning the format, standardization, and training which is a method that has proven itself through the years. But one wonders whether it is not wiser to use the format for its more adequately understood intention, in this case, a statement. Another possibility that must be submitted to further tests would be to substitute the rheme with a verbal form when the intention is injunctive. For example, Engine 1.....turn off. That being said, we should note, however, that half of the messages in this experiment had an action verb rheme such as "set", "deploy", "engage", "discharge", etc. For example: Oxygen Mask.....Deploy. We observe the same significant effects for reaction times and accuracy in favor of the MNL syntax ("Deploy Oxygen Masks") as we do for when the rheme consists of a state of a button such as "on" or "off".

These results are consistent with Spaggiari (2002) findings in the example "right engine pump.....start". When asked, a large majority of Airbus pilot participants who have been trained to understand the format as injunction, said that they understood the format as an injunction to start the right engine pump. However, non-Airbus pilots and non-pilots were split down the middle between both interpretations which led the author to conclude that the structure by itself is not transparent for non-trained users. And this notwithstanding the fact that the verb "start" is semantically more infused with the injunctive mode (as it is a verb in the imperative mode) than words that describe a state such as "off" or "on" or "250 KT", which could in turn lead to an increased bias towards the statement interpretation of the format.

Conversely, we did the same analysis for the MNL messages to see whether intention played a role on accuracy, and we did not find any significant difference in the scores for Injunction ($M = 0.89$, $SD = 0.95$) and Statement ($M = 0.88$, $SD = 0.86$) conditions; $t(139) = 2.287$, $p = 0.2$.

Since the accuracy means are almost equal (injunction and statement were equally accurately understood in the MNL condition), we can conclude that contrary to the MCL messages, the MNL ones accurately convey their intended meaning. This means that MNL is not better or more frequently understood as an injunction or statement, but is understood equally and correctly across both of them. By using a more natural controlled language, the absence of different possible interpretations entails a more coherent understanding of the original intent.

5. **Did the language factor play a different role for different levels of English placement (Basic Intermediate-Proficient-Mastery-Natives) regarding accuracy?**

Paired sample t-tests show that MNL was significantly more accurate than MCL for all English placement levels ($p = 0.000$). Contrary to previous experiments that showed that MNL helped accuracy for lower levels of English placement while the effect did not hold for natives, in this experiment all placement levels benefitted equally significantly from the MNL condition. Therefore, MNL is understood more accurately than MCL by a range of different English comprehension levels. This might be because the differences between the two language formats are not as subtle as they were in the previous experiments (what we added in the MNL conditions in the information category experiment was mostly grammatical articles, and only subtle syntactical reformulations). This experiment offers considerable modifications because of the ellipses in the MCL original format on the one side, and because of the addition of agency (in injunctions) and explicitness (in statements) in the natural language format. These modifications could explain the results being relevant to different levels of English as the improvement is more substantial.

6. **Did the language factor play a different role for different levels of familiarity with Airbus Controlled language regarding accuracy?**

Paired sample t-tests show that once again MNL is significantly more accurate regardless of different levels of familiarity with Airbus CL ($p = 0.000$). While the difference in means of different familiarity groups was always significant, we would like to point out that for the expert population (test pilots/ human factors and Flight warning experts/other Airbus staff that work on the language on an almost daily basis) there was an even larger difference in means than in other groups (cf. Table 21).

MCL_ None	MNL_ None	MCL_ Beginners	MNL_ Beginners	MCL_ Inter-mediate	MNL_ Inter-mediate	MCL_ Experts	MNL_ Experts
0.802	0.877	0.80	0.88	0.82	0.89	0.80	0.92

Table 21. Means of Accuracy of Different Language Conditions Concerning Familiarity with Airbus CL

As we can see there is a 12-point difference (0.80-0.92) in accuracy for the expert population (as opposed to 7.5 (0.802-0.877) for no familiarity-8 (0.80-0.88) for beginner familiarity- 7 (0.82-0.89) for intermediate familiarity). This is unexpected as experts (population that includes pilots) are already used to the way the controlled language is and are familiar with its rules. One would assume that the new more natural controlled language would be more difficult for them (thus be less accurate) because of the habituation process that occurs with the use of the original controlled language.

It is possible however, that because in this case, the MCL was used in two different ways (injunction and statement) whereas they are used to the format being exclusively for injunctions; the confusion that the task created by having injunctions and statements made it more difficult than usual for them to respond accurately, and the more natural language being more explicit helped them respond better. Another reason could simply be that the explicitness of the more natural language for an expert population helps information processing. Whatever the reason, this remains an interesting observation to report, as the argument of moving towards a more natural language being uncertain to an already trained population does not hold as strongly in light of these results.

6.4.1 Subjective Research

6.4.1.1 Questionnaire

As we mentioned in Section 6.2.6, participants answered a questionnaire before they started the main task:

Outside of Airbus context, what do those sentences mean to you?
Please explain on the lines below.

Engine 1 ....On

____

Speed ....250 KT

____

As shown in the objective research (research question 4 regarding accuracy of MCL when it was used for statements or injunctions), we can see in the pie charts in Figure 70 and Figure 71, the majority of participants interpreted more intuitively the messages as statements (60.7% and 62.8% respectively), 25.7% mentioned that the messages could have both interpretations, and only 13.5% and 8.5% respectively interpreted them as injunctions. Therefore, it is safe to say that there is not one overwhelming interpretation for the format. If one were to be privileged, it would/should be the statement interpretation as the majority of participants interpreted the messages more intuitively as statements, and we could conclude that the theme.....rheme format represents statements more naturally than it does injunctions. At best, we could say that it is a divided opinion, one that could be set straight with a more natural format that limits the different possible interpretations, such as "switch on engine 1" for an injunction or "speed is set to 250 KT" for a statement.

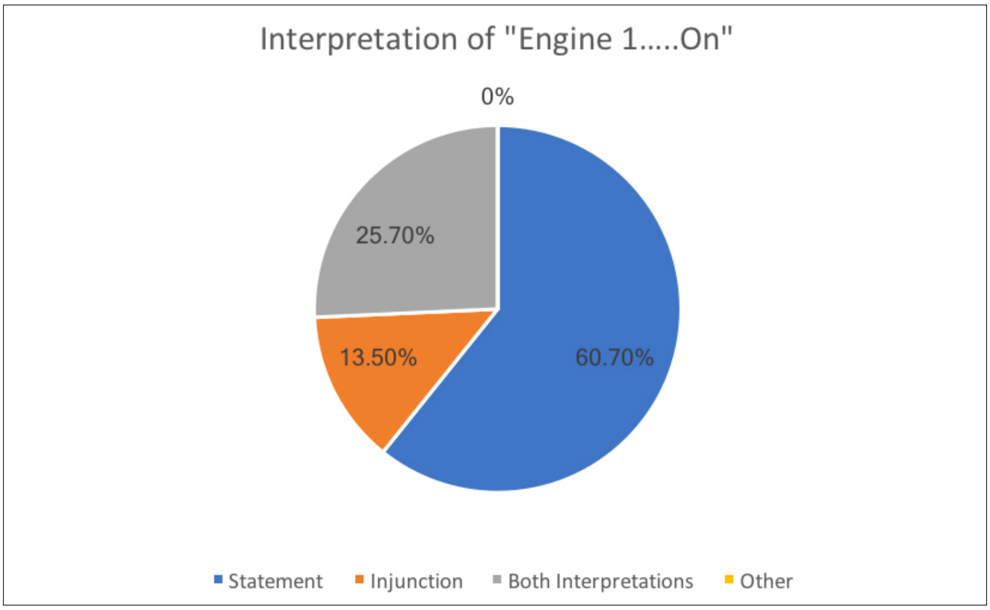


Figure 70. Interpretation of Engine 1.....On

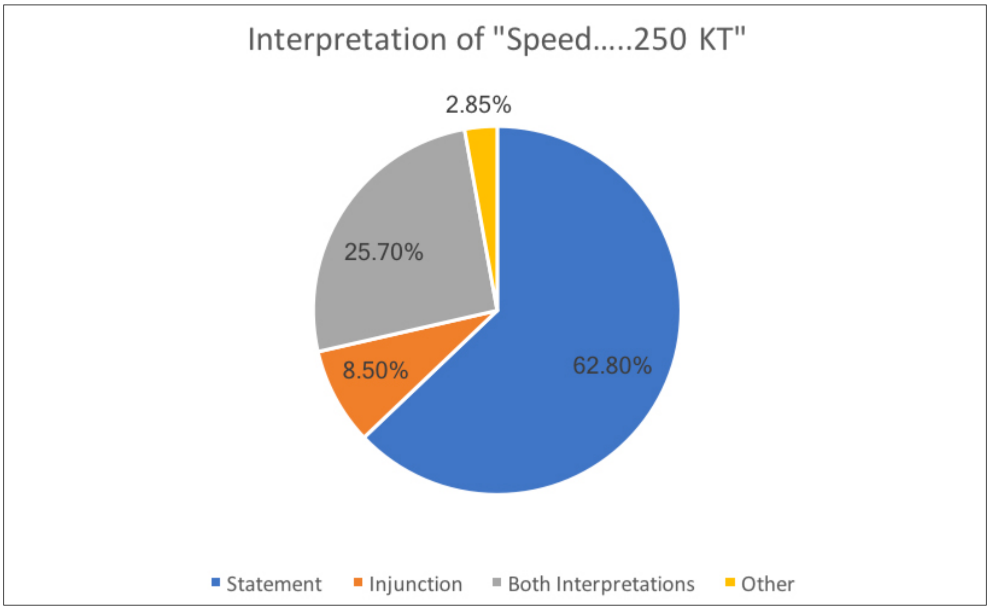


Figure 71. Interpretation of Speed.....250KT

6.4.1.2 Second Phase Stress

Regarding perceived stress, 113 participants mentioned that the second urgent phase (when participants had 4000ms to respond) was stressful and 27 did not perceive it as such (mostly because they did not feel like the stakes were high enough in case of a wrong answer, no consequences on a real flight for example). Out of 113, twenty participants mentioned that the red background color was a helping factor that induced more stress.

We can conclude that an overwhelming majority of participants perceived the second urgent phase as a more stressful task. While we recommend that stress be objectively measured (heart rate/perspiration etc.) in future evaluations, participants answers and results show that the second phase, by the use of time pressure (3 x less time than first phase) and a red background (as a contrast to a neutral gray one in the first phase), did present a higher level of stress because participants had to urgently respond.

6.4.1.3 Preference

Participants were also asked at the end of the experiments whether they had a preference towards one form of messages, the MCL format or the MNL format.

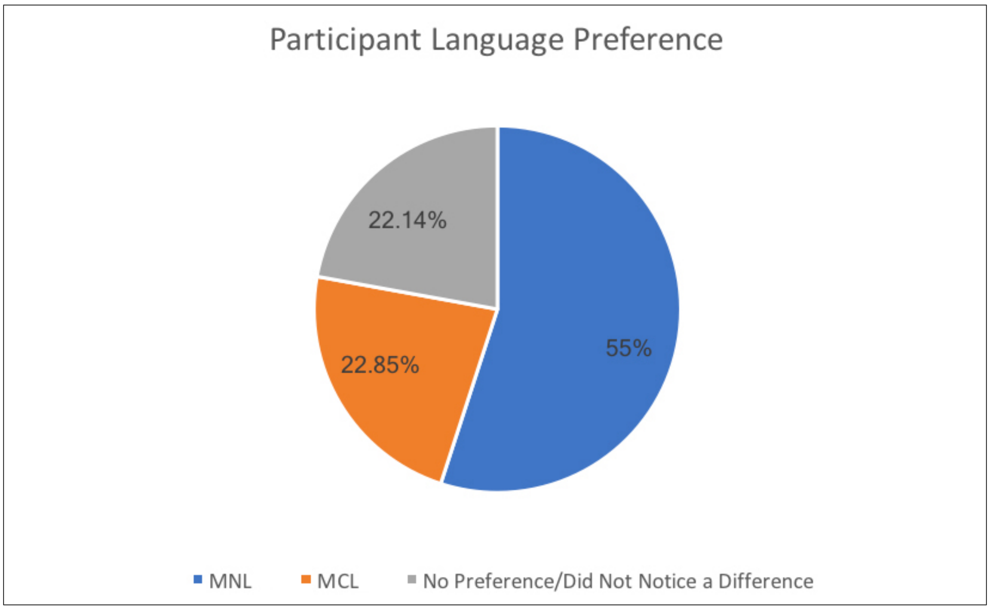


Figure 72. Participant Language Preference

As we can see in the pie chart in Figure 72, 55% of participants mentioned preferring the more natural messages (77 total, of whom 3 were experts, 17 had intermediate knowledge of format, 26 beginners, and 30 had no knowledge of CL), 22.85% of participants mentioned preferring the coded format (32 total, of whom 5 were experts, 13 had intermediate knowledge of format, 10 beginners, and 4 had no knowledge of CL), and 22.14% of participants said they had no preference or that they did not notice the difference between the messages (31 total, of whom 3 were experts, 9 had intermediate knowledge of format, 12 beginners, and 7 had no knowledge of CL). Naturally, as we can see in the set of pie charts in Figure 73, the less you are habituated and acquainted with the format, the more tendency you have to prefer the more natural language structure. This preference could be explained by the intuitiveness of the natural language structures and meanings to the untrained eye which makes it easier to fulfill a task, while familiarity with a certain format (case of experts) creates habituation and a relative affinity with a known tool (controlled language used on a daily basis). Nevertheless, preferring a language does not always equate better performance (as the results for the expert population show in research question 6).

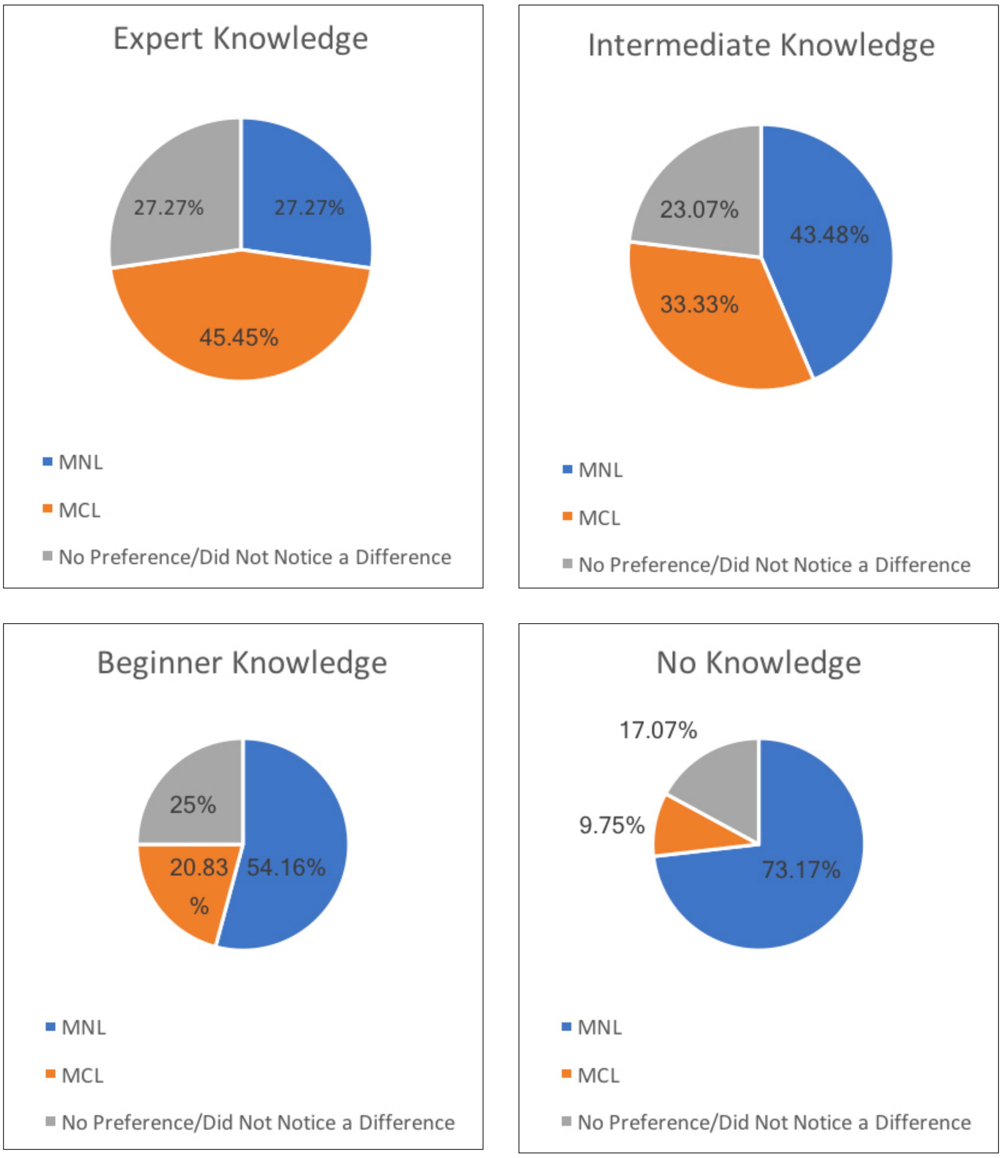


Figure 73. Preferences Linked to Levels of Familiarity with Airbus CL

We would also like to point out that regardless of the participants' familiarity with Airbus controlled language, native speakers of English overwhelmingly preferred the natural language structures, as the structures seemed more intuitive and clearer. This result was paradoxically also observed in the native *expert* population.

Test pilots (pilots who are experts in Airbus aircraft's ways of functioning and cockpit interactions) and to some extent private pilots, particularly appreciated the fact that the MNL formats allowed them to have a superior situational awareness; i.e, it allowed them to memorize and keep in mind the action to be undertaken more than the MCL format that they are used to. They expressed having the feeling of performing faster in the MCL conditions because on the one hand they are used to the format, and on the other because they mechanically and automatically performed what needed to be done, however the MNL allowed them to have a better degree of clarity of what they were supposed to do and what they are about do, and the subsequent state of the aircraft after the fact (as in they kept in mind the action performed). For example, when asked to turn the engine off in the MNL format, they were more aware of what they just performed than when they performed the same action in the theme…..rheme format, which made them execute the action as a machine would, without question or extra thought, which in turn made them less aware of what they just executed.

Another test pilot remarked that as the messages in the cockpit are read on an ECAM monitor and the required action is always done on a different panel or a different monitor, the MCL sometimes requires several visual revisits to the original message before action execution. He went on to say that had the messages been in the MNL format, one would not need to re-read and revisit the original message on the original monitor as much as with the MCL format (which constitutes a loss of time), because the MNL is easier to memorize and makes you more aware of the action to be performed.

We recommend that the effect of the language format on memorization be evaluated in future research to supplement these findings with objective research.

6.4.1.4 Difficulties

- As was confirmed with our objective analyses of accuracy averages, 90.71% of participants mentioned that the fuel zone was difficult and most complex due to its many possible interactions with several pumps and tanks, as we mentioned in Section 6.2.3. It included a multitude of on/off buttons, a tricky visual architecture (confusion between right, left, and center), and two possible interactions with the word "feed" in them. Almost 99% of participants mentioned that the presence of the word "feed" in different zones was particularly confusing to locate and deal with, such as "crossfeed" A and B in the Fuel zone, "Central tank feed" also in the fuel zone, "All Engine Bleed" in the Engine zone, and "Crossbleed" in the Air zone. In the cockpit controlled language, the possible confusion between the words "crossbleed" and "crossfeed" is dealt with by always abbreviating the word "crossfeed" into "XFeed". We can see here that this is an extremely justified rule.

- 12 participants (8.57%) mentioned that they were confused by the "Engine 1 Agent 2.....Off" or "Engine 2 Agent 1" messages in the Engine zone, but not necessarily when the engine number and the agent number matched such as "Engine 1 Agent 1"or "Engine 2 Agent 2".
- 22 participants (15.71%) mentioned having difficulty with the "High Ventilation" on/off button in the Air Zone because they confused it with the "Airflow" button also in the Air Zone, that had "High", "Normal", "Low" as potential selections. Therefore, "High" ventilation and "High" airflow were being confused because the object of the word "High" was semantically very close: "Ventilation" and "Airflow" share the same semantic field of things involving air management.
- 17 participants (12.14%) mentioned that the colors of the layout could have been more intuitive. The blue-green light that signaled the "On" state of the button would have been more easily perceived had it been more obviously green. The reason why we felt we needed to use a blue-green button instead of a more obvious green button was because we had a dark interface background. The blue-green light is what is suggested in Google Material Design UI[41] for dark backgrounds, whereas the more obvious green should be used with a lighter background.

41. Google Material Design UI - Floating Action Buttons set Sketch freebie - Sketch App Sources. [online] Available at: https://www.sketchappsources.com/free-source/597-google-material-design-ui-sketch-app.html [Accessed 9 Dec. 2018].

- Additionally, some difficulties were perceived because of the difference in buttons that light differently, as we mentioned in Section 6.2.1.
- Some participants mentioned that it was confusing to find mentions of the word "Engine" in the Fuel zone (non-clickable diagram put there to orient participants on the use of fuel tanks specific to an engine) as opposed to the real on/off buttons of the engines in the Engine zone.
- The brakes zone, as we mentioned was one of the easiest zones as per accuracy averages, however some participants (especially ones more familiar with aviation) found confusing the fact that when they are asked to retract speed brakes, the button "Extend" speed brakes was inactive, and the same goes for the parking brakes "Engage" and "Release" buttons. This was a mindful decision on our part in the experimental design, since activating the opposite button would have been equally confusing as participants would not have known if it was possible to select (or confirm the state of deactivation of) the needed button when the opposite one was active.
- 7 participants (5%) expressed having some language related issues, such as dealing with apostrophes and the possessive form in short sentences, as in examples we mentioned before when we needed to use phrases like "Engine 1's Agent 2" or "Left Tank's Main Pump".
- A few expert participants who were familiar with the coded format complained that the MNL format included "Shakespearean" sentences, even though the reality of it could not have been farther from the truth (most times, there were only one or two extra words in the MNL sentences, e.g "Turn off Engine 1" vs. "Engine 1…..off"). It is interesting, however, to point out how habituation to a short and coded format could influence perception, especially since those same experts who preferred the MCL performed considerably better in the MNL condition.
- A few participants found confusing some alliterative words such as "Pilot" and "Passengers".
- General difficulties included locating buttons on interface at the beginning of the task, even with the extensive familiarization phase. Locating left and right (left of the screen or left of the Engine zone?). Some participants found the need to press the "confirm" button after performing an action (not when confirming a statement), but we found that need subsided after a few trials.

6.4.1.5 Comments for MNL and MCL

- As participants were asked which language format they preferred and why, they gave their ideas and opinions which we summarized in the following word clouds. We entered in an Excel sheet all the comments and words that participants evoked after performing the experiment. Those comments were then pooled together for all participants and for each language condition. They data was cleaned for negation (for example the words "not easier" or "not clear" were removed and synonyms were unified as much as possible. The data was then submitted on https://www.wordclouds.com/, a website which generates word clouds based on the frequency of words in submitted texts. The bigger the word is, the more frequently participants mentioned it for each language. These word clouds give us a general graphic idea as to what words participants used to describe each language.

Figure 74. MNL Word Cloud Participants' Comments

As the language preference numbers confirm, we can see comments for MNL center around the format being more **clear**, **informative**, and **easy** to **understand** and **memorize** because the format is more familiar and **intuitive** etc. Additionally, participants mentioned that the format **helps** them know more what is happening and what they needed to do, instead of automatically responding (cf. Figure 74).

Figure 75. MCL Word Cloud Participants' Comments

The MCL comments center around the format being more **confusing/ difficult** and **ambiguous**, and the presence of a checklist paradigm might take more time to process and could be harder for beginners, etc.But we also find that by some it was seen as **easier** because it is visually simpler to locate the theme and the rheme, it has **less words**, it takes less time, the **separation** helps, and the **format is familiar** (cf. Figure 75).

6.5 General Discussion

As we have seen, experimental results show that our main hypotheses were validated in the most part. The more natural controlled language proved to be more accurately understood than the more coded controlled language across both conditions of urgency. However, MNL only proved faster to process in the urgent condition but not in the non-urgent condition.

These results are of high interest to us because, contrary to popular belief, in the controlled natural language domain (CNL), more simplification did not lead to a more optimized comprehension. On the contrary, the theoretically most complex language (closer to natural language = more ambiguity, less controllability) has proven by its more natural syntactic structure to be more easily understood than the theoretically most simple (very codified and controlled, farther from natural language structure = reduction of ambiguity).

Even though the MNL often had longer more verbose structures, it also proved to significantly reduce response times in cases of time pressure (which in turn created a more stressful situation and a higher sense of urgency). On the other hand, the MCL had significantly faster reaction times than the more natural format when time pressure and stress were not an issue, and participants had 3 times more time to respond. This is not surprising given the context (no instructions for fast responses) and the fact that MNL had longer structures, which when time is abundant, could take longer to read (and potentially double check) on account of the sentences being longer. This particular result leads us to believe that the more natural controlled language structure is processed faster by the brain when needed and especially under stress, but this effect effervesces with the addition of more time because there is no pressure to respond quickly and more words equal more reading time. However, while it is true that participants performed faster in the MCL non-urgent conditions, they did not respond more accurately. In other words, while participants took significantly longer to respond in the non-urgent MNL condition, they were making less errors than in the MCL non-urgent condition. These results are backed by the analyses of errors and non-answers per participant and urgency phase, separately and together. Errors and non-answers were always significantly higher in the MCL condition across urgency conditions.

As aforementioned, the errors and non-answers in the urgent condition show a large significant discrepancy between the MNL and MCL. Non-answers (or time's up instances) play an important role in showing that MNL is processed faster because there are far fewer non-answers in this condition. Even though there are more words to read, the sentence structure is more easily (significantly less non-answers) and correctly (significantly more accurate) processed. This could be due to the reduction of possible interpretations in the sentence structure, whereas the ellipses in the MCL condition rendered the format harder to decode, understand, and respond to efficiently because of the linguistic economies and lack of explicitness.

Additionally, in some cases, the ellipses render the MCL format more open to different interpretations, thus making processing longer, and consequently led participants to exceed the time limit (or possibly to refrain from answering because unsure of correct response). The separated theme and rheme format did not facilitate the visual search and respond task as its structure is intended to do (locate the theme quickly on the left, and see what needs to be done to it on the right of the elliptical dots), but the complete sentence, which does not include any typographical visual separation, did in fact facilitate the processing of information in the task.

The research questions we explored showed us that more complex zones were not affected by the language factor. In previous experiments, we saw that MNL showed better accuracy and response times in more complex stimuli but did not have the same effect in more simple stimuli (but we should keep in mind that what we considered more complex in previous experiments differs greatly to this type of complexity). In this case, the general observed results (MNL more accurate than MCL in all urgency conditions and has faster reaction times in urgent conditions) were observed across all different zones in the same manner. In other words, the same results were observed in the speed zone (least complex) and the fuel zone (most complex). We could potentially explain this by supposing that even though the zones in this experiment (and interactions associated to them) were significantly distinct in complexity, the least complex of them presented more challenges (and thus was sufficiently complex) than the complexity we evoked in previous experiments (incongruous target stimuli that are close to being considered as congruent, but which fall short of conforming to the prime).

Additionally, we explored whether the MCL had better accuracy when being used as a statement or an injunction. Results showed that the MCL theme…..rheme format is significantly more accurately (therefore more frequently and better) understood as a statement, while the MNL did not have a significantly better accuracy in either injunction or statement conditions as the accuracy averages were very close in both conditions. This shows us that MNL messages, accurately convey their intended meaning, and the scale is not tipped one way or the other, whereas MCL messages are significantly better understood as statements.

As we already mentioned, from a linguistic point of view, this is a logical result as the rheme often describes the state of the button (off-on-high-250kt, etc.). Since the original format is meant to be understood as an injunction, we argue that the format alone does not intuitively convey that intention. This could perhaps be remedied by using action verbs in the place of the rheme, such as “Engine 1…..Turn on” instead of “Engine 1…..on”, or “Speed…..set to 300kt” instead of “Speed…..300Kt”, although a large number of stimuli did include action verbs in the focus position such as “Oxygen Masks….. Deploy” and did not show superiority to the MNL format (Deploy Oxygen Masks) in accuracy or reaction times. Therefore, since the typographically variable format does not particularly show any superiority to natural language (less accurate etc.), we could safely envisage substituting it with the typographically stable sentence format, which itself is loyal to its meaning and intention.

Concerning participant variables, the same significant language related accuracy results were observed for different English language comprehension levels (including the difference between natives and non-natives). These results are consistent with our previous findings (which were at times lacking a sufficient number of native speakers). We can now affirm that the effects could be generalized to natives and non-natives of different English comprehension levels.

We can also affirm that the familiarity with Airbus controlled language did not play any role regarding accuracy, as all populations shared the same significant results observed in the general population. We will however mention that the expert population showed a more striking difference in accuracy between the MNL and MCL (more accurate for MNL format), which is unexpected due to experts being more habituated to the MCL format. This finding is interesting, if nothing else, because it shows that habituation to a format does not necessarily mean one will perform better in said format, and neither does it mean that this same population would be more averse to learning a new more optimal one. We did observe that experts had more of a tendency to prefer the MCL format because of familiarity and habituation, so we could refer to Nielsen and Levy (1994) as this is a clear example of a performance-preference paradox.

The subjective research shows that firstly, the MCL format is more naturally and instinctively understood as a statement and to a much lesser degree as an injunction, at best it could be understood both ways, which in turn raises questions about the adequacy of this choice of format for the injunctive intention.

We also established that the urgent phase was overwhelmingly perceived as such in contrast to the first non-urgent phase.

On the whole, participants preferred the MNL messages (55%) to the MCL (22.85%) ones, while 22.14% of participants had no preference or did not notice a difference in the stimuli. Native speakers overwhelmingly expressed their preference towards the more natural controlled language whereas there were more divergent opinions in the non-native population.

Moreover, the less familiar with the original format the more participants tended to prefer the MNL over the MCL format. However, test pilots and private pilots as a majority expressed the possible benefits of using a more natural controlled language in this task as well as in their usual cockpit interactions.

Reported difficulties included for the most part the Fuel zone with its many buttons and interactions between left and right, main and standby, pumps and tanks etc. The other main confusion was between “Crossfeed”, “Crossbleed”, “All Engine Bleed”, “Center Tank Feed”; and some interface and visual search related issues.

Word clouds obtained from comments made about the two different language formats show that MNL, in concordance with participants’ preference results, was seen as being helpful, more clear and informative, and easy to understand, but also as long and containing a lot of words. The MCL comments center around being more confusing, difficult to decipher, having a checklist paradigm (neither clearly an injunction nor a statement). But it was also seen as visually easier (separation of theme and rheme), taking less time to process because having less words, being a familiar format (for those who were familiar with it).

6.6 Limitations and Perspectives

Due to time limitations, we were not able to complete the eye tracking analyses which would have provided some insight into visual search and respond tasks related to language format and urgency. It would also have shed more light on the cognitive processes involved in task execution.

We also would have liked to propose and use a different interface design which takes into consideration the few adjustments (colors-buttons-zone architecture) proposed by participants, experts, and advisors. The logical next step in this experimentation would be to test the effect of a more natural language with the introduction of the current color-coding scheme as a new variable, which could show interesting combinations of linguistic aspects and semiotic ones.

It also would have been interesting to have more uniform populations of pilots, ones that have the same qualifications, same number of flying hours, flown similar aircraft, in order to better test whether piloting qualification play a role in our analyses, but alas this is not very easy to find. Likewise, we would have liked to have a bigger sample of Airbus test pilots, but this was not made easy by their extremely busy schedules.

6.7 Conclusion and Way Forward

The results in this experiment are satisfactory in the sense that they validate most of our hypotheses, and provide some empirical proof to the question regarding the appropriate level of simplification needed for optimal comprehension, and offers more natural language syntactical structures as a viable alternative for injunctions and statements in different urgency conditions. The results also shed light on interesting topics worth exploring such as the role of language control in memorization and situational awareness, the role of interface design in conjunction with different language formats, the role of typographically variable formats alone versus typographically and semantically variable formats, the role of abbreviations and symbols in different language structures, the role of natural language structures in different categories of information such as questions, conditions, topics, etc.

III

DISCUSSION AND CONCLUSION

7

Global Analysis of Results and Recommendations

7.1 Recapitulation of Over-all Results

These 3 experiments constituted a first attempt at evaluating a controlled language using tightly controlled experimental paradigms inspired by methods from cognitive sciences, and more specifically from psycholinguistics and cognitive psychology.

We tested two data categories: informational statements (or functions' availabilities) and action statements (injunctions, instructions to perform an action) from a highly codified cockpit controlled language destined to guide, advise, and give instructions to pilots in order to help them navigate and operate the aircraft in normal and abnormal situations.

This controlled language was previously simplified and standardized in order to avoid ambiguity, complexity, and incoherence on a syntactic and lexical level. In these experiments, we were particularly concerned with the existing syntactic simplifications and linguistic economies, mainly ones that eliminated function words (experiment 1 and 2) and that enforced structural or elliptical reductions (experiment 3).

Since we questioned the limits of simplification and its role in providing better human comprehension (as its premise suggests) and the role of naturality in a language (or a language closer to an INL, which is theoretically more complex), we put in place experimental paradigms that allowed us to compare similar syntactic structures of the highly codified theoretically simpler language to a more natural and theoretically more complex one.

We built this new more natural language by de-codifying existing sequences and bringing them closer to natural language (by adding function words, explicitness, a syntactic sentence structure instead of typographical separation, etc.).

We took into consideration sentence difficulty and length, as well as time criticality (time induced stress to recreate a sense of urgency).

In the first experiment, we obtained encouraging results concerning reaction times for the more natural controlled language, and confirmed our hypothesis. Responses to MNL stimuli were significantly faster than MCL stimuli even though the more natural stimuli contained more words and allowed for less reading time. Accuracy of responses to stimuli was not affected for either languages. Thus, our accuracy hypothesis in favor of the more natural language was denied, but it could be explained by the fact that the additions to the messages (which consisted in adding a sentence structure with function words) did not offer a significantly conspicuous change that would affect accuracy, whereas it was discrepant enough to be picked up on by subtle, but sufficiently significant, reaction time differences. This suggests that the sentence structure introduced by function words does ease comprehension on a cognitive level.

In experiment 2, when the same stimuli were used but more time was given for their presentation, the significant effects observed in experiment 1 disappeared which suggests that speed, and to a certain extent the stress induced from having to respond urgently, play a role in information processing and comprehension. This led us to introduce the speed variable directly in the experimental protocol of experiment 3 which did not use the same stimuli, as the nature of the task and data type were entirely different. Be that as it may, this allowed us to test for the effect of induced urgency on responses and comprehension.

Syntactic difficulty, on the other hand, did not incur any interactions between the two languages in the 6 difficulty conditions in experiment 1 (one language condition did not show faster reaction times in certain difficulty conditions, but not in others). MNL stimuli were consistently faster than MCL ones in all difficulty conditions.

However, the difficulty of the task itself manifested in the more complex incongruent stimuli did show significantly faster reaction times for the MNL stimuli (in both experiments 1 and 2), while the congruent stimuli did not. Moreover, even though we do not have a significant general effect for reaction times in the results of experiment 2, still the incongruent stimuli show significant effects for the MNL condition over the MCL condition. Therefore, we can conclude that in cases of increased task difficulty the more natural language helps ease comprehension.

In experiment 1, MNL seems to facilitate comprehension for participants in the basic intermediate level placement, and this suggests that weaker English speakers would benefit more greatly from a more natural language than confirmed speakers. While native English speakers performed better on average in the MNL condition, that effect was not statistically significant, and this was more properly verified in experiment 3 with a larger sample of native English speakers.

Experiment 1 and 2 included congruency tasks which are very close to traditional judgment tasks in behavioral experiments. They provided a firmly controlled environment to test our linguistic hypotheses, nonetheless, the downside of using such experiments is that we are limited to evaluating passive comprehension, mainly of specific informative statements. It would be quite difficult to evaluate the comprehension of an order or an instruction using traditional judgment tasks.

That is why we created a human-machine interface that was meticulously constructed to fit our needs of evaluating comprehension of injunctions in real-time performance tasks. It had the benefit of emulating a potential cockpit interface with a few concocted aircraft buttons and functions, in a relatable pseudo-piloting task (as it was on a computer interface and did not contain actual piloting scenarios), all the while remaining within the realm of laboratory condition testing of robust experimental paradigms. Once again, the main variable was the contrasted efficiency of the two language conditions in the performance of the task at hand, in addition to the speed variable, zone complexity, and elliptical format intuitiveness.

Contrary to experiment 1 and 2, experiment 3 showed that MNL was significantly more accurately understood than MCL across both conditions of urgency. However, a significant interaction for speed and reaction times was observed: MNL messages were only significantly faster to process in the urgent condition but not in the non-urgent condition. This is consistent with the results observed in the first two experiments which also suggested that MNL's added value was more discerningly observed in time critical conditions (induced stress from urgent-like situations).

It is interesting to note that even though MNL messages contained more words on account of being more natural and less elliptical, they were more accurately understood and processed faster in urgent conditions.

However, the MCL messages had significantly faster reaction times than the more natural format when time pressure and stress were not an issue (participants had 3 times more time to respond). This could be explained by the fact that participants were not instructed to respond fast and in the non-urgent phase (which was always the first phase of the experiment) had ample time to read, possibly re-read, and reflect on their answers before responding. Since MNL had longer structures and more words, and time was not an issue, more words equaled slower reading time. This particular result leads us to believe that the more natural controlled language structure is processed faster by the brain when needed and especially under stress, but this effect effervesces with the addition of more time because there is no pressure to respond quickly and more words equal more reading time. However, while it is true that participants performed faster in the MCL non-urgent conditions, they nevertheless did not respond more accurately. In other words, while participants took significantly longer to respond in the non-urgent MNL condition, they were also making less errors than in the MCL non-urgent condition.

These results are backed by the analyses of errors and non-answers as they were always significantly higher in the MCL condition across urgency conditions. Non-answers (or time's up instances) play an important role in showing that MNL is processed faster because there are far fewer non-answers in this condition. Even though there are more words to read, the sentence structure is more easily (significantly less non-answers) and correctly (significantly more accurate) processed. This could be due to the reduction of possible interpretations in the sentence structure, whereas the ellipses in the MCL condition rendered the format harder to decode, understand, and respond to efficiently because of the linguistic economies and lack of explicitness. The multiple interpretations could have caused subjects to exceed time limits (more non-answers). We can then conclude that because the sentence structure reduces possible interpretations it also helps visually locate information and respond faster and more accurately. While the separation of the theme and rheme in the MCL messages could seem more efficient for information localization on the screen, it did not translate into significant effects in performance.

Complexity of zones did not play a role or have interactions concerning the two language conditions as MNL was consistently more accurate across all zones of the interface (from least difficult to most difficult). Contrary to experiment 1 and 2, the effect in experiment 3 was observed consistently across difficulty conditions (in experiment 1 and 2 the added difficulty of the incongruent stimuli gave a significant advantage to the MNL messages, when it was absent in less difficult conditions). However, we should note that the added difficulty in the first two experiments and the third are hardly comparable as the task is entirely different.

Other results that caught our attention in experiment 3 consisted in showing how the MCL theme…..rheme format was better understood (as an information statement or an injunction). Results showed that the MCL theme…..rheme format is significantly more accurately (therefore more frequently and better) understood as a statement, while the MNL did not have a significantly better accuracy in either injunction or statement conditions as the accuracy averages were very close in both conditions. This shows us that MNL messages accurately convey their intended meaning, the scale is not tipped one way or the other, whereas MCL messages are significantly better understood as statements. Therefore, since the typographically variable format does not particularly show any superiority to natural language structures (less accurate etc.), we could safely envisage substituting it with the typographically stable sentence format, which itself is loyal to its meaning and intention.

The same effects were observed for all the subjects who had diverse English comprehension levels, including native English speakers, as we now had tested a sufficiently large sample population of natives. These results are consistent with the first two experiments' results.

Also consistent with the results of the first two experiments, was the familiarity with original corpus, which did not produce any specific effects (besides the main significant ones for MNL). Experts, however, tended to perform better in MNL conditions which is not the language structures they are used to working with. This is noteworthy because it shows that habituation to a format does not necessarily mean one will perform better to said format, and neither does it mean that this same population would be more averse to learning a new more optimal one; nevertheless, experts generally leaned towards the MCL format when asked about their preference (performance-preference paradox).

An interesting finding in the subjective research of experiment 3 shows that firstly, the MCL format is more naturally and instinctively understood (by subjects) as a statement and to a much lesser degree as an injunction, at best it could be understood both ways, which in turn raises questions (regarding its integration in future cockpit design) about the adequacy of this choice of format for its original injunctive intention. On the whole, participants preferred the MNL messages (55%) to the MCL (22.85%) ones, while 22.14% of participants had no preference or did not notice a difference in the stimuli. Native speakers overwhelmingly expressed their preference to the more natural controlled language, whereas there were more divergent opinions in the non-native population.

Moreover, the less familiar with the original format the more subjects tended to prefer the MNL over the MCL format (results valid for all 3 experiments). However, both test pilots and private pilots, as a majority, expressed the possible benefits of using a more natural controlled language in the experimental tasks as well as in their usual cockpit interactions. They mentioned that MNL messages help cue them in to the actual actions they need to undertake, and to remember and be more aware of the procedures being done.

The following comparative Table 22 recapitulates the main results of the 3 experiments. In the first part (in gray), we could see the specificities of each experiments such as number of participants, type of corpus data tested, etc. The second part (in blue) shows some of the significant results (highlighted in yellow) that were previously discussed.

<table>
<tr><th></th><th colspan="4">Experiment 1</th><th colspan="4">Experiment 2</th><th colspan="4">Experiment 3</th></tr>
<tr><td rowspan="2">Participants</td><td colspan="2">Native</td><td colspan="2">Non-Native</td><td colspan="2">Native</td><td colspan="2">Non-Native</td><td colspan="2">Native</td><td colspan="2">Non-Native</td></tr>
<tr><td colspan="2">12</td><td colspan="2">60</td><td colspan="2">1</td><td colspan="2">28</td><td colspan="2">41</td><td colspan="2">99</td></tr>
<tr><td>Type of Corpus Data</td><td colspan="4">Information/availability</td><td colspan="4">Information/availability</td><td colspan="4">Injunction/Statement</td></tr>
<tr><td>Type of Task</td><td colspan="4">Congruency
Image/Text
(DMDX)</td><td colspan="4">Congruency
Image/Text
(DMDX)</td><td colspan="4">Performance on
Touchscreen HMI
(ePrime 3)</td></tr>
<tr><td>Syntactic Difficulty of Stimuli</td><td colspan="4">1 → 6</td><td colspan="4">1 → 6</td><td colspan="4">–</td></tr>
<tr><td>Urgency</td><td colspan="4">Urgent
(Time pressure)</td><td colspan="4">Non-Urgent
(No time pressure)</td><td colspan="4">Urgent +
Non-Urgent Phases</td></tr>
<tr><td rowspan="3">Accuracy
General Significance</td><td colspan="2">MCL</td><td colspan="2">MNL</td><td colspan="2">MCL</td><td colspan="2">MNL</td><td colspan="2">MCL</td><td colspan="2">MNL</td></tr>
<tr><td colspan="2" rowspan="2">No</td><td colspan="2" rowspan="2">No</td><td colspan="2" rowspan="2">No</td><td colspan="2" rowspan="2">No</td><td>Urgent</td><td>Non-Urgent</td><td>Urgent</td><td>Non-Urgent</td></tr>
<tr><td>No</td><td>No</td><td>Yes</td><td>Yes</td></tr>
<tr><td rowspan="3">Reaction Time
General Significance</td><td colspan="2">MCL</td><td colspan="2">MNL</td><td colspan="2">MCL</td><td colspan="2">MNL</td><td colspan="2">MCL</td><td colspan="2">MNL</td></tr>
<tr><td colspan="2" rowspan="2">No</td><td colspan="2" rowspan="2">Yes</td><td colspan="2" rowspan="2">No</td><td colspan="2" rowspan="2">No
(avg. Yes)</td><td>Urgent</td><td>Non-Urgent</td><td>Urgent</td><td>Non-Urgent</td></tr>
<tr><td>No</td><td>Yes</td><td>Yes</td><td>No</td></tr>
<tr><td rowspan="3">Reaction Time Significance
English Level</td><td colspan="2">MCL</td><td colspan="2">MNL</td><td colspan="2">MCL</td><td colspan="2">MNL</td><td colspan="2">MCL</td><td colspan="2">MNL</td></tr>
<tr><td>Basic/ Interm.</td><td>Native Speakers</td><td>Basic/ Interm.</td><td>Native Speakers</td><td>Basic/ Interm.</td><td>Native Speakers</td><td>Basic/ Interm.</td><td>Native Speakers</td><td>Basic/ Interm.</td><td>Native Speakers</td><td>Basic/ Interm.</td><td>Native Speakers</td></tr>
<tr><td>No</td><td>No</td><td>Yes</td><td>No
(avg. Yes)</td><td>No</td><td>No</td><td>No</td><td>No</td><td>No</td><td>No</td><td>Yes</td><td>Yes</td></tr>
<tr><td rowspan="2">Reaction Time Significance
Syntactic Difficulty</td><td colspan="2">MCL</td><td colspan="2">MNL</td><td colspan="2">MCL</td><td colspan="2">MNL</td><td colspan="2">MCL</td><td colspan="2">MNL</td></tr>
<tr><td colspan="2">No</td><td colspan="2">Yes</td><td colspan="2">No</td><td colspan="2">No</td><td colspan="2">–</td><td colspan="2">–</td></tr>
<tr><td rowspan="2">Reaction Time Significance
Incongruence Difficulty</td><td colspan="2">MCL</td><td colspan="2">MNL</td><td colspan="2">MCL</td><td colspan="2">MNL</td><td colspan="2">MCL</td><td colspan="2">MNL</td></tr>
<tr><td colspan="2">No</td><td colspan="2">Yes</td><td colspan="2">No</td><td colspan="2">Yes</td><td colspan="2">–</td><td colspan="2">–</td></tr>
<tr><td rowspan="2">Reaction Time Significance
Zone Difficulty</td><td colspan="2">MCL</td><td colspan="2">MNL</td><td colspan="2">MCL</td><td colspan="2">MNL</td><td colspan="2">MCL</td><td colspan="2">MNL</td></tr>
<tr><td colspan="2">–</td><td colspan="2">–</td><td colspan="2">–</td><td colspan="2">–</td><td colspan="2">No</td><td colspan="2">Yes</td></tr>
<tr><td rowspan="3">Accuracy Significance
Comprehension of Format</td><td colspan="2">MCL</td><td colspan="2">MNL</td><td colspan="2">MCL</td><td colspan="2">MNL</td><td colspan="2">MCL</td><td colspan="2">MNL</td></tr>
<tr><td colspan="2" rowspan="2">–</td><td colspan="2" rowspan="2">–</td><td colspan="2" rowspan="2">–</td><td colspan="2" rowspan="2">–</td><td>Statement</td><td>Injunction</td><td>Statement</td><td>Injunction</td></tr>
<tr><td>Yes</td><td>No</td><td>No
(equal means)</td><td>No
(equal means)</td></tr>
</table>

Table 22. Comparative Table Recapitulating Main Results of the 3 Experiments

7.2 Application to CNLs

7.2.1 Implication for the Airbus Cockpit CNL

On the whole, with these experiments we were hoping to shed more light on the limits of simplification and to what extent the naturality component in a controlled language is inherently ambiguous or paradoxically clear. In this case study of a highly coded controlled language, we observed that going towards what could be deemed potentially ambiguous (closeness to natural language structure) and redundant syntactic elements, in reality improves comprehension and performance on the whole, and facilitates information processing. These experiments were also a first attempt at providing empirical proof for linguistic hypotheses of text simplification in controlled languages, using robust psycholinguistic protocols and newly available experimentation technology.

Oversimplification, in the sense of drifting away from naturality of language and its potential ambiguities, did not lead to better cognitive and behavioral results for this controlled language. Hence, we can conclude that going towards a more natural structure by using full syntactic sequences and respecting the adequate speech acts will lead to faster and more accurate comprehension, and will subsequently decrease training time as the structures have proven to be more intuitive for a broad range of people.

7.2.2 Implication for the CNL domain

However, these results do not necessarily apply to all controlled languages (or even to all the coded, less natural ones) as it depends on the domain in which they are used, for what purposes, for which speech acts, and for the different types of transmitted data. A lot of comprehension-oriented controlled languages already use natural language structures or have a high score on the PENS naturalness dimension. However, the rules that define these languages are often untested for user comprehension and the limits of simplification not properly defined. In reference to the comprehension of instructional texts Nickl (2018) writes: *"one of the key factors for achieving greater understanding of how comprehension and comprehensibility function was the shift in focus away from surface phenomena in texts towards an investigation of the reader's/listener's interpretation process. In academia, this shift in focus was the product of cognitive theories, which proved that text interpretation is not a passive process of information-decoding, but rather an active process of meaning-construction. This made it* *evident that anyone wishing to improve textual comprehensibility must take into account the text's prospective readers. Comprehensibility was transformed from being merely an aspect of a text into a description of the relation between a given text and that text's target audience. This entailed/means? that texts can no longer be characterized as comprehensible in themselves."*

Therefore, the answer to the question of the proper extent of simplification lies not only in the usage and purposes given to a controlled language, but also in the systematic evaluation of its rules using adequate techniques for measuring human cognitive skills and behavioral responses. For only by having empirical proof of the efficacy of a controlled language could we be certain of the adequacy of the rules that make it. Ryan (2018) writes, in reference to an overview of the domain: *"however, there remain a number of issues concerning the efficacy of controlled languages in actual use and the quantitative evaluation of the practical gains they convey."*

7.2.3 Recommendations for the Airbus CNL

From the results of our evaluations we established some recommendations concerning the evolution of the Airbus Controlled Language for future aircraft, particularly (but not only) for the informational and injunctive data types. These recommendations are directly derived from results pertaining to both the objective and subjective research, in addition to 1 recommendation from evaluations in the literature (when a recommendation is based on subjective research or literature, it was mentioned in the proposed rule).

This part has been partly redacted from the final version of the thesis for confidentiality reasons. The complete version will be filed along with some other PhD data for internal Airbus use only[42]. We kept this redacted section in order to demonstrate the methodology involved in any CNL evaluation, from beginning to end (starting with clear linguistic hypotheses that are evaluated using cognitive methodology on human performance to obtain significant results from which recommendations would be directly derived).

The following section introduces the titles of the subsections and a few examples of rules from the first subsection (I. Evaluation Rules):

42. It was important to keep this thesis non-confidential on the whole in order to put forth evaluations and results that could be useful for the scientific community, since otherwise, evaluations are somewhat lacking in the domain (even if it meant redacting sensitive material at times).

I. Evaluation Rules

1. Always evaluate new rules against existing ones. The most important recommendation is to evaluate any new additions, compare and contrast them to existing rules or competing new rules, for instance, two different proposals of natural language structures.

2. In the evaluations, isolate the proposed new addition variables before combining them with additional variables, such as using abbreviations or semiotic signs (like color coding).

3. Use proper experimentation protocols (laboratory condition cognitive behavioral tools) that fit the speech act of the data type needed and its usability, before introducing the new language rules in simulator evaluations.

II. Explicitness Rules

III. Difficulty and Stress Condition Rules

IV. Ergonomic Rules

V. Familiarity and Preference Rules

7.3 Way Forward

This PhD research is part of a larger project at Airbus which aims to redesign future disruptive cockpit. As its name suggests, this project offers innovative approaches and solutions to entirely revamp the pilot's work environment namely by proposing more automation and more intuitive/guided designs. These new designs would be especially useful considering aircraft manufacturers' move towards reducing crew numbers in the future. Thus, the new more natural language is a brick in the wall of the disruptive cockpit design concept: a more intuitive interface (bigger screens, newer technology, etc.) goes hand in hand with a more appropriate language: less coded structures are easier to learn and memorize and require less pilot training. Since the system might be more automated there is a greater need to inform the pilot of the state of the aircraft and keep him/her in the loop. More natural language structures are more permissive (not as restrictive as the coded language) and could be a useful tool to help with situational awareness and decision-making. RILEY ET AL. (1999) writes that *"as automation becomes more sophisticated and complex, there is increasing concern about the time and cost of training pilots* *to use it, and the possibility of pilot errors. […] One of the most important factors in usability is the extent to which the underlying functionality and logical operation of a device or system is consistent with the user's mental model. […] we want to make the system work like the pilot thinks, so we don't have to train the pilot to think like the system works. Because actions and targets all behave according to syntactic and semantic rules the pilot already knows, by virtue of knowing how to comply with instrument clearances, no additional training is required to learn a new function or procedure."*

In order to integrate a new more natural language in future cockpits, more evaluations should be done. We will give a few suggestions here:

- A logical next step would be to use the same interface designed for experiment 3 in order to test the effect of using the existing abbreviations instead of the full forms which we used here: "Turn off ENG 1" vs. "Turn off Engine 1".
- Secondly, as we have established that all upper-case lettering was not ideal for optimal legibility, we used lower-case in experiments 1 and 2 and title case (only beginnings of words are capitalized) in experiment 3. It would be interesting to test the effects of capitalization ("ENGINE" or "Engine" or "engine") as well as semiotic aids such as color coding as separate variables in separate experimental designs.
- Thirdly, to ascertain whether the more natural language does enhance memory and awareness as the results suggest, we could design an experimental protocol using the same interface in experiment 3 which would specifically test the effects of the language condition on memory, for example by interrupting the task in order to ask subjects to recall the last action they performed and to repeat the message. Some authors also ask subjects to perform simple or complex mathematical operations before asking them to recall what they read/performed last. This would be a good way to empirically ascertain that the more natural language also aids memory and situational awareness.
- We could also use physiological methods to test the effects of stress on information processing in the different language conditions such as heart rate or blood pressure monitors.
- Eye tracking analysis could be used to monitor participants gaze paths and observe which language requires more revisits for effective comprehension and optimal task achievement.

- Finally, Figure 76 shows the different corpus data types that still need to be evaluated (in blue) such as questions, conditions, titles, etc.

  In green, we have the more ergonomic/semiotic/graphic elements left to evaluate, such as the inclusion of pictograms to facilitate comprehension, organization of linguistic and graphical information on a page.

  In orange, we added the potential evaluation of existing controlled language rules. Rules are already established for the existing data types, and it would be logical to base future evaluations on these rules (and determine which ones could be kept and which ones should evolve) in order to optimize them.

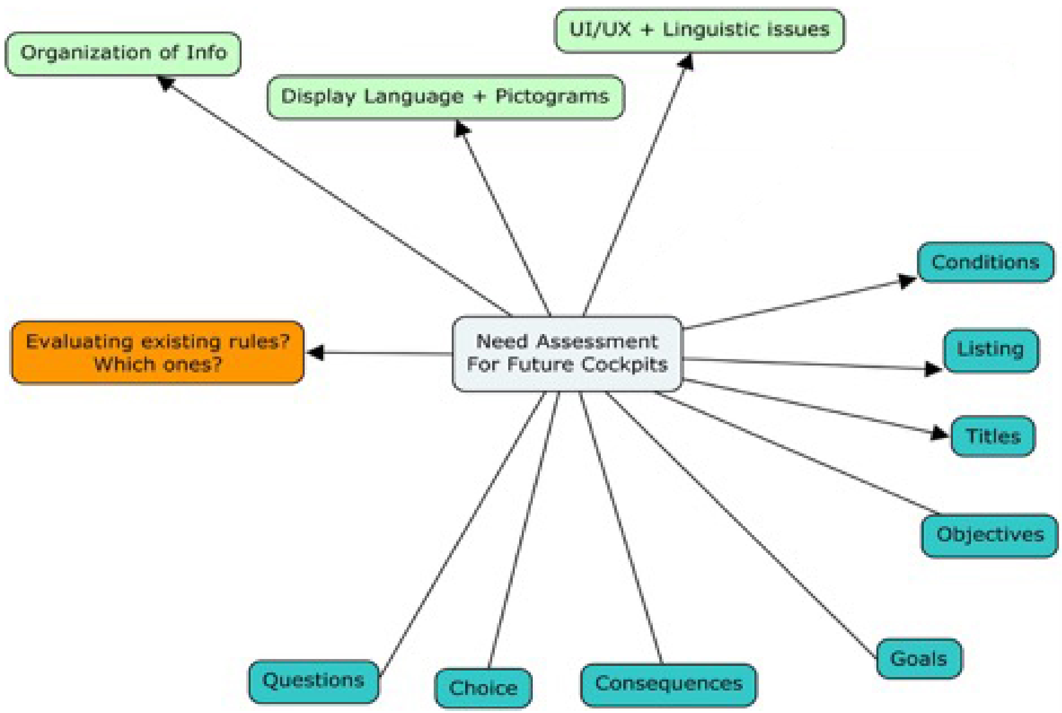


Figure 76. Representation of Future Need Assessment Proposols

8

Conclusion

This PhD research was initiated from an industrial need, that of optimizing the use of language in future Airbus cockpits for maintaining safety and optimizing pilot performance and limiting training needs. In order to fulfil this need, we first had to delve into the existing controlled languages in Airbus and the context in which they are used, the limitations and perspectives of optimization.

Furthermore, we studied, from an academic point of view, the domain of controlled languages and all the research that is associated to it, how and where controlled languages are used, to what end, and do they accomplish their goals in reducing ambiguities in industrial settings with human end users. We discovered that there were surprisingly very little empirical studies done on human comprehension when it came to established CNLs.

Comprehension oriented CNLs tend to be designed to fit industrial needs and specifications to be used by a designated target audience (even if those specifications could prove problematic outside of the domain), rely heavily on general writing style guides, linguistic research guidelines, and recycled simplification rules, which are seldom scientifically tested on human cognitive processing. CNL rules which are often arbitrary (maximum number of words in a sentence, use of function words, use of passive voice, use of modals and pronouns, etc.) are therefore put in place without evidence that they make comprehension easier for end users. A lot of controlled languages destined for industrial use are so well-guarded (for judicial liability reasons and intellectual property) that it is difficult to advance efficiently in the domain, or to do research on the different rules that exist.

What's more, while simplification seems necessary, because uncontrolled languages are dangerously ambiguous in safety critical domains, the rules produced are not guaranteed to achieve better human comprehension (better than natural language, or better than other simplification rules), as the limits are blurred. How much simplification is necessary to achieve better comprehension, is there a threshold after which too much simplification causes more ambiguity? The answer was unclear.

Therefore, our initial industrial need evolved into researching more issues on the linguistic, ergonomic, and psycholinguistic fronts. Natural language or INLs being inherently ambiguous has brought forth the misconception that the more we control or codify a language and take it away from its ambiguity-causing natural elements, the more we have a failsafe way to eradicate misuse; and the obvious way to do that is by enforced training and teaching of the rules of the simplified language.

In order for us to find more answers concerning the limits of simplification, we looked into the times researchers attempted to find proof that certain established controlled languages brought added value for human comprehension. As we showed in Section 3.5, the results were somewhat inconclusive, since reliable significant results were not brought forth in favour of a more simplified language versus a more natural less simplified one.

Additionally, since these evaluations were a first approach to finding empirical proof starting in 1995, the behavioral experimentation methods used were not up to the standards of today's cognitive research techniques. Reading comprehension questionnaires were used to test comprehension of instructional texts (instead of performance, as they are meant to be understood and executed). Response times were not limited nor controlled.

Therefore, our aim is to propose a first approach to evaluating controlled languages using tightly controlled psycholinguistic behavioral protocols in laboratory conditions to test the limits of simplification. Since the main requirement was to optimize the current controlled language by taking into consideration the future design of the cockpit and its interfaces, we based ourselves on this highly codified language (which is reinforced by training and learning) and theoretically least complex (no natural language ambiguities).

This language contains several different categories of data that are meant to guide, warn, and help pilots navigate the aircraft. We identified them and based our experimentation on two main ones, the informational type (informing pilots of occurrences, availabilities etc.), and the action type (instructing pilots to perform an action on a button or lever, screen, etc.). We proceeded to propose more natural language structure equivalents to the existing coded structures by adding missing syntactic and sentential elements, and by respecting the adequate speech acts for each category of information in classical English INL syntax (for example starting with the verb phrase for injunctions).

Results show that oversimplification did not lead to better comprehension and the more natural version produced better results and are more intuitively understood (therefore, will require less training for mastering).

While the existing more coded language (which included several technical limitations due to screen sizes, etc.) has thus far successfully limited ambiguity, it needs to be supplemented by more user training. Using a more natural language for the purposes of the Airbus CL in a redesigned future cockpit is more beneficial for human comprehension and performance.

Other controlled languages have different levels of naturality and simplification that are dependent on their usage, domain, target audience, and general goals. However, the rules that make them should be psycholinguistically evaluated to ensure that the prescriptive and proscriptive rules that make them are as efficient as possible, and that they truly reduce ambiguity and improve human comprehension and performance. Finally, simplification and linguistic economies do not always or automatically equate improvement of comprehension or make for more appropriate conduits of information.

To conclude, we have shown through cognitive scientific methodology, that the Airbus controlled language could benefit from more natural language structures to enhance pilot comprehension and reduce training times. This new more optimized language fits effectively into the future disruptive cockpit concept and its more intuitive designs. We also showed that there is a noticeable lack of controlled language evaluations in the field, as well as adequate methods of evaluating linguistic hypotheses using firm cognitive sciences methodology to satisfy ergonomic needs. We propose that in the future, controlled language rules should be systematically evaluated to demonstrate their efficacy before being applied, especially in safety critical domains.

We also hope for this thesis to provide insight/motivation for using cognitive methodology and behavioral data for testing classical/descriptive linguistic hypotheses, even beyond the domain of controlled languages. Linguistics and cognitive behavioral methodology should come together in a more effective manner to reap the benefits of scientifically verifiable and comparable results, with assumptions based on decades of linguistic theory.

ANNEXES

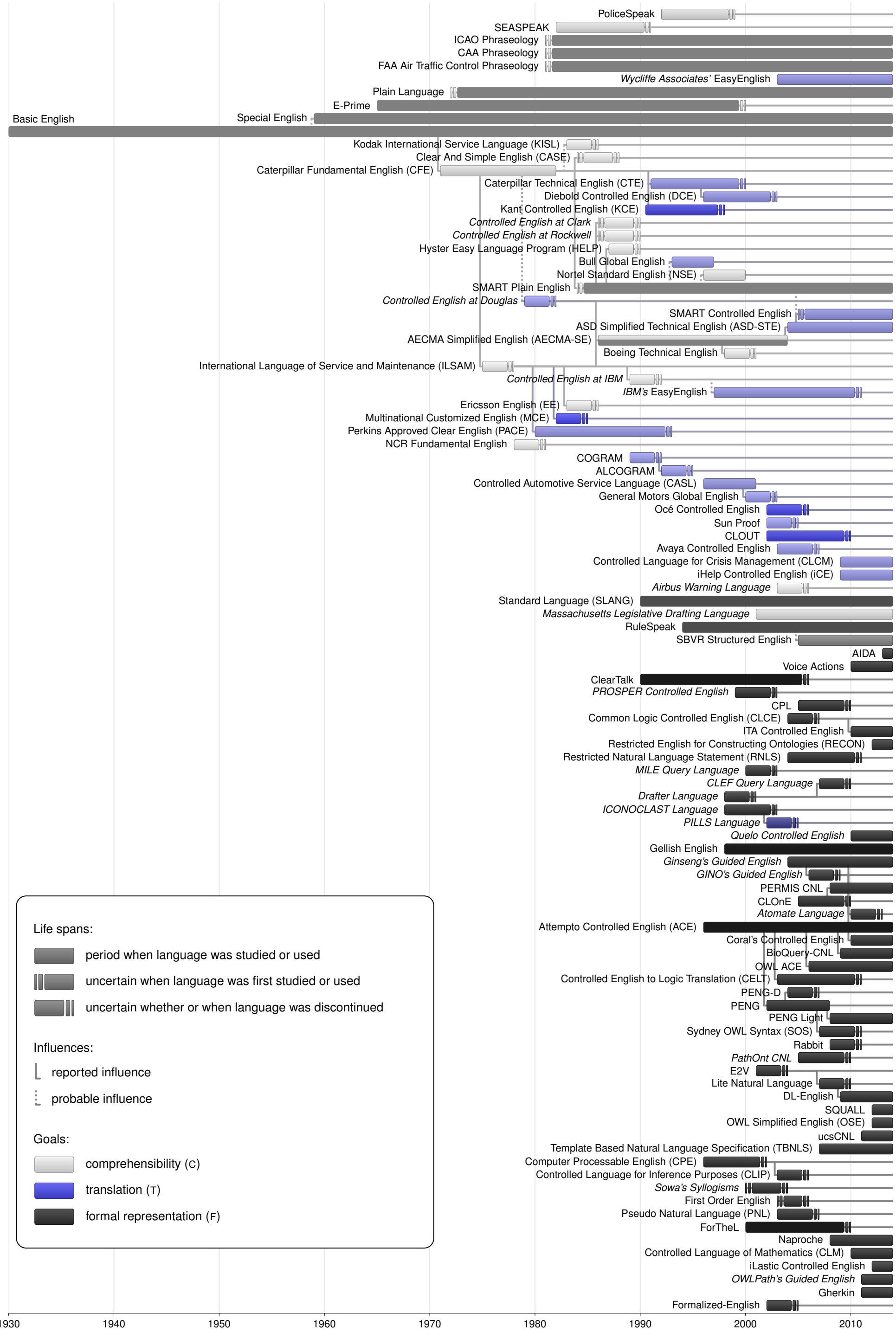
A

Observed PENS Classes and Properties of CNLs (sorted by PENS class)

class	properties	languages
$P^1E^5N^5S^1$	C T W I	IBM's EasyEnglish
	C W S G	Special English
	C W A	E-Prime
	C W G	Plain Language
$P^2E^1N^3S^2$	C S D G	CAA Phraseology, FAA Phraseology, ICAO Phraseology, PoliceSpeak, SEASPEAK
$P^2E^1N^3S^3$	C W D I	Airbus Warning Language
$P^2E^5N^4S^1$	F W A	AIDA
$P^2E^5N^5S^1$	C T W D A I	ALCOGRAM, COGRAM
	C T W D A	CLCM
	C T W D I	ASD-STE, Avaya CE, Bull GE, CTE, CASL, CE at Douglas, DCE, General Motors GE, PACE, Sun Proof
	C T W D	Wycliffe Associates' EasyEnglish
	C T W I	iCE, SMART Controlled English
	C W D I	AECMA-SE, CFE, CASE, CE at Clark, CE at IBM, CE at Rockwell, EE, HELP, ILSAM, KISL, NCR FE
	C W D G	Massachusetts Legislative Drafting Language
	C W I	Boeing Technical English, NSE, SMART Plain English
	C W	Basic English
	T W D I	MCE, Océ Controlled English
	T W A	KCE
	T W I	CLOUT
$P^3E^1N^4S^2$	C F W D I	SLANG
	F S D I	Voice Actions
$P^3E^2N^4S^3$	F W D A	RNLS
$P^3E^3N^3S^3$	F W A	ClearTalk
	F W I	ITA CE
$P^3E^3N^4S^2$	F W I	CPL
$P^3E^4N^4S^2$	C F W I	RuleSpeak, SBVR-SE
$P^4E^1N^4S^3$	F W D A	Drafter Language, MILE Query Language
$P^4E^1N^4S^4$	F W A	Quelo Controlled English
$P^4E^1N^5S^3$	T F D A	PILLS Language
$P^4E^2N^4S^3$	F W D A	Atomate Language
	F W A I	Gellish English
	F W A	GINO's Guided English
	F W I	CELT
$P^4E^3N^4S^3$	F W D A	PROSPER CE
	F W A	ACE
$P^4E^3N^5S^3$	F W D A	ICONOCLAST Language
$P^5E^1N^4S^3$	F W D A	CLEF Query Language
	F W A	Ginseng's Guided English
$P^5E^1N^4S^4$	F W D A	Coral's Controlled English
	F W A	PathOnt CNL
$P^5E^1N^4S^5$	F W A	Sowa's syllogisms
$P^5E^2N^3S^4$	F W D A I	TBNLS
	F W A	OWLPath's Guided English, SQUALL
$P^5E^2N^4S^3$	F W A	CPE, CLIP, OWL ACE, SOS
$P^5E^2N^4S^4$	F W D A	BioQuery-CNL, PERMIS CNL, ucsCNL
	F W A	CLOnE, DL-English, E2V, Lite Natural Language, OSE
	F W G	Rabbit
$P^5E^3N^3S^3$	F W D A	CLM, ForTheL, Naproche CNL
	F W A	CLCE, PNL
$P^5E^3N^4S^3$	F W D A	Gherkin
	F W A G	RECON
	F W A	First Order English, PENG, PENG-D, PENG Light
	F W I	iLastic Controlled English
$P^5E^4N^3S^3$	F W A	FE

B

The Timeline of the Evolution of Controlled English



C

Real Examples from the Corpus of Messages Containing the Word AVAIL

	Original Message	Syntactical Construction
NOT AVAIL (The one thing is not available)	CPNY DTLNK NOT AVAIL	NOUN + NOUN + NOT AVAIL
	APU BAT START NOT AVAIL	NOUN + NOUN + NOUN + NOT AVAIL
	WING A-ICE NOT AVAIL ON APU BLEED	NOUN + NOUN + NOT AVAIL + ON + NOUN + NOUN
	APPR MODE NOT AVAIL	NOUN + NOUN + NOT AVAIL
	SWAP NOT AVAIL	NOUN? + NOT AVAIL
NO + something AVAIL (None of the things are available)	NO VOICE COM AVAIL	NO + NOUN + NOUN + AVAIL
	NO COM AVAIL	NO + NOUN + AVAIL
AVAIL	L TK 17000 KG MIN AVAIL (REMOVE ONE NOUN) X2	ADJ + NOUN + NUM +NOUN + NOUN + NOUN + AVAIL
	STBY INSTRUMENTS AVAIL	NOUN + NOUN + AVAIL
	SATCOM DATALINK AVAIL (ADD 3RD NOUN)	NOUN + NOUN + AVAIL
	SLATS AVAIL	NOUN + AVAIL
	FWD BRK WITH A-SKID AVAIL	ADJ + NOUN + PREP (WITH) + NOUN + AVAIL
	CAPT KCCU KEYBOARD ARROWS AVAIL	NOUN (LOCATION) + NOUN + NOUN + NOUN + AVAIL
	F/O KCCU KEYBOARD ARROWS AVAIL	NOUN (LOCATION) + NOUN + NOUN + NOUN + AVAIL
	CAPT SOFT KEYBOARD AVAIL	NOUN (LOCATION) + ADJ + NOUN + AVAIL
	F/O SOFT KEYBOARD AVAIL	NOUN (LOCATION) + ADJ + NOUN + AVAIL
	FWD BRK AVAIL	ADJ + NOUN + AVAIL
	APPR MODE NOT AVAIL	NOUN + NOUN + NOT AVAIL
	ADS-C DATALINK AVAIL	NOUN + NOUN + AVAIL
	APU AVAIL	NOUN + AVAIL
	AP1/FD AVAIL (SIDESTICKS NOT LOCKED)	NOUN + AVAIL (PRECISION)
	SPEED BRAKES AVAIL	NOUN + NOUN + AVAIL
AVAIL+ condition	LAV & GALLEYS EXTRACT AVAIL IN FLT	NOUN + AND + NOUN + NOUN +AVAIL + IN + NOUN
	AVNCS EXTRACT AVAIL IN FLT (AD D 3RD NOUN) X2	NOUN + NOUN + AVAIL + IN + NOUN
	PACK 1 AVAIL IN FLT	NOUN + NUMBER + AVAIL + IN + NOUN

	PACK 1+2 AVAIL IN FLT	NOUN + NUMBER + AND + NUMBER + AVAIL + IN + NOUN
	CARGO FWD TEMP REGUL AVAIL IN FLT	NOUN + ADJ + NOUN +NOUN + AVAIL + IN + NOUN
	PFD ALPHA PROT SPD AVAIL IN ALL CONF	NOUN + NOUN + NOUN + NOUN + AVAIL + IN + ADJ + NOUN
	BKUP TRAJ AVAIL ON ND(RNG MAX 160 NM) (AD D 3^{RD} NOUN)	NOUN + NOUN + AVAIL + ON + NOUN
	AVNCS EXTRACT AVAIL IN FLT	NOUN + NOUN + AVAIL + IN + NOUN
	NORM BRK AVAIL ON ALL WHEELS	ADJ + NOUN + AVAIL + ON + ADJ + NOUN
	GREEN DOT, S, F, VAPP AVAIL ON FMS	ADJ + NOUN, NOUN, NOUN, NOUN + AVAIL + ON + NOUN
	WING A-ICE NOT AVAIL ON APU BLEED	NOUN + NOUN + NOT AVAIL + ON + NOUN + NOUN
	SMOKE DET AVAIL ON OVHD PNL & FAP ONLY	NOUN + NOUN + AVAIL + ON + NOUN + NOUN + AND + NOUN + ONLY
	FIRE DET AVAIL ON OVHD PNL ONLY	NOUN + NOUN + AVAIL + ON + NOUN + NOUN + ONLY
	APU START AVAIL ON EXT PWR ONLY	NOUN + NOUN + AVAIL + ON + ADJ + NOUN + ONLY
	AFS CTL PNL KNOB AVAIL FOR BUG SETTING	NOUN + NOUN + NOUN + NOUN + AVAIL + FOR + NOUN + NOUN
AVAIL/NOT AVAIL: (has an incoming list)	FMS DATA AVAIL :	NOUN + NOUN + AVAIL:
	AUDIOS NOT AVAIL :	NOUN + NOT + AVAIL

D

Different Possibilities for MNL Message Construction

Original Message	Syntactical Construction	AIRBUS CONTROLLED LANGUAGE	MNL			
L TK 17000 KG MIN AVAIL	ADJ + NOUN + NUM + NOUN + NOUN + AVAIL	LEFT CONTAINER 20 KILOS MINIMUM AVAILABLE	THERE ARE MINIMUM 20 KILOS AVAILABLE IN THE LEFT CONTAINER	THERE ARE 20 KILOS MINIMUM AVAILABLE IN THE LEFT CONTAINER	THE LEFT CONTAINER HAS MINIMUM 20 KILOS AVAILABLE	THE LEFT CONTAINER HAS 20 KILOS MINIMUM AVAILABLE
		GREEN ROOM 100 KILOS MAXIMUM AVAILABLE	THERE ARE MAXIMUM 100 KILOS AVAILABLE IN THE GREEN ROOM	THERE ARE 100 KILOS MAXIMUM AVAILABLE IN THE GREEN ROOM	THE GREEN ROOM CONTAINS 100 KILOS MAXIMUM AVAILABLE	THE GREEN ROOM CONTAINS MAXIMUM 100 KILOS AVAILABLE
		BIG HOUSE 5 ROOMS MAXIMUM AVAILABLE	THERE ARE MAXIMUM 5 ROOMS AVAILABLE IN THE BIG HOUSE	THERE ARE 5 ROOMS MAXIMUM AVAILABLE IN THE BIG HOUSE	THE BIG HOUSE HAS 5 ROOMS MAXIMUM AVAILABLE	THE BIG HOUSE HAS MAXIMUM 5 ROOMS AVAILABLE
		BLUE CAR 2 SEATS MAXIMUM AVAILABLE	THERE ARE MAXIMUM 2 SEATS AVAILABLE IN BLUE CAR	THERE ARE 2 SEATS MAXIMUM AVAILABLE IN BLUE CAR	THE BLUE CAR HAS MAXIMUM 2 SEATS AVAILABLE	THE BLUE CAR HAS 2 SEATS MAXIMUM AVAILABLE
		BIG BUILDING 2 ELEVATORS MAXIMUM AVAILABLE	THERE ARE MAXIMUM 2 ELEVATORS AVAILABLE IN BIG BUILDING	THERE ARE 2 ELEVATORS MAXIMUM AVAILABLE IN BIG BUILDING	THE BIG BUILDING HAS MAXIMUM 2 ELEVATORS AVAILABLE	THE BIG BUILDING HAS 2 ELEVATORS MAXIMUM AVAILABLE
		NEW GALLERY 2 SHOWROOMS MAXIMUM AVAILABLE	THERE ARE MAXIMUM 2 SHOWROOMS AVAILABLE IN NEW GALLERY	THERE ARE 2 SHOWROOMS MAXIMUM AVAILABLE IN NEW GALLERY	THE NEW GALLERY HAS MAXIMUM 2 SHOWROOMS AVAILABLE	THE NEW GALLERY HAS 2 SHOWROOMS MAXIMUM AVAILABLE
L TK 17000 KG MIN AVAIL	ADJ + NOUN + NUM + NOUN + NOUN + NOUN + AVAIL	RIGHT CAGE 50 BANANAS MINIMUM NUMBER AVAILABLE	THERE IS A MINIMUM NUMBER OF 50 BANANAS AVAILABLE IN RIGHT CAGE	-	THE RIGHT CAGE HAS A MINIMUM NUMBER OF 50 BANANAS AVAILABLE	THE RIGHT CAGE HAS A NUMBER OF 50 MINIMUM BANANAS AVAILABLE
		BIG MALL 50 SUPERMARKET GROCERY BAGS AVAILABLE	THERE ARE 50 SUPERMARKET GROCERY BAGS AVAILABLE IN THE BIG MALL	-	THE BIG MALL HAS 50 SUPERMARKET GROCERY BAGS AVAILABLE	-
		CLEAN ROOM 5 WOOD BROOMSTICK HANDLES AVAILABLE	THERE ARE 5 WOOD BROOMSTICK HANDLES AVAILABLE IN THE CLEAN ROOM	-	THE CLEAN ROOM HAS 5 WOOD BROOMSTICKS AVAILABLE	5 WOOD BROOMSTICKS ARE AVAILABLE IN THE CLEAN ROOM
		EMPTY HOUSE 2 METAL BASEMENT DOORS AVAILABLE	THERE ARE 2 METAL BASEMENT DOORS AVAILABLE IN THE EMPTY HOUSE		THE EMPTY HOUSE HAS 2 METAL BASEMENT DOORS AVAILABLE	2 METAL BASEMENT DOORS ARE AVAILABLE IN THE EMPTY HOUSE
		BIG POOL 3 WATER FOUNTAIN HEADS AVAILABLE	THERE ARE 3 WATER FOUNTAIN HEADS AVAILABLE IN BIG POOL		THE BIG POOL HAS 3 WATER FOUNTAIN HEADS AVAILABLE	3 WATER FOUNTAIN HEADS ARE AVAILABLE IN THE BIG POOL
		NEW STORE 5 BICYCLE BELL BRANDS AVAILABLE	THERE ARE 5 BICYCLE BELL BRANDS AVAILABLE IN NEW STORE		THE NEW STORE HAS 5 BICYCLE BELL BRANDS AVAILABLE	5 BICYCLE BELL BRANDS ARE AVAILABLE IN THE NEW STORE

SATCOM DATALINK AVAIL	NOUN + NOUN + AVAIL	CHALK BOARD AVAILABLE	THE CHALK BOARD IS AVAILABLE	THERE'S AN AVAILABLE CHALK BOARD	THE CHALK BOARD IS AVAILABLE	-
		TREE HOUSE AVAILABLE	THE TREE HOUSE IS AVAILABLE	THERE'S AN AVAILABLE TREE HOUSE	THE TREE HOUSE IS AVAILABLE	-
		BUS STOP AVAILABLE	THE BUS STOP IS AVAILABLE	THERE'S AN AVAILABLE BUS STOP	THE BUS STOP IS AVAILABLE	-
		SCHOOL BUS AVAILABLE	THE SCHOOL BUS IS AVAILABLE	THERE'S AN AVAILABLE SCHOOL BUS	THE SCHOOL BUS IS AVAILABLE	-
		FISH TANK AVAILABLE	THE FISH TANK IS AVAILABLE	THERE'S AN AVAILABLE FISH TANK	THE FISH TANK IS AVAILABLE	-
		CITY PARKING AVAILABLE	THE CITY PARKING IS AVAILABLE	THERE'S AN AVAILABLE CITY PARKING	THE CITY PARKING IS AVAILABLE	-
SATCOM DATALINK AVAIL	NOUN + NOUN + NOUN + AVAIL	FLOWER SHOP WINDOW AVAILABLE	THE FLOWER SHOP WINDOW IS AVAILABLE	THE FLOWER SHOP'S WINDOW IS AVAILABLE	THERE'S AN AVAILABLE WINDOW IN THE FLOWER SHOP	
		VILLAGE MOUNTAIN VOLCANO AVAILABLE	THE VILLAGE GROCERY STORE IS AVAILABLE	THE VILLAGE'S GROCERY STORE IS AVAILABLE	THERE'S AN AVAILABLE VILLAGE GROCERY STORE	
		GARDEN LEMON TREE AVAILABLE	THE GARDEN LEMON TREE IS AVAILABLE	THE GARDEN'S LEMON TREE IS AVAILABLE	THERE'S AN AVAILABLE GARDEN LEMON TREE	
		CASINO JAZZ SINGER AVAILABLE	THE CASINO JAZZ SINGER IS AVAILABLE	THE CASINO'S JAZZ SINGER IS AVAILABLE	THERE'S AN AVAILABLE CASINO JAZZ SINGER	
		CRIME SCENE WITNESS AVAILABLE	THE CRIME SCENE WITNESS IS AVAILABLE	THE CRIME SCENE'S WITNESS IS AVAILABLE	THERE'S AN AVAILABLE CRIME SCENE WITNESS	
		TV SCREEN IMAGE IS AVAILABLE	THE TV SCREEN IMAGE IS AVAILABLE	THE TV SCREEN'S IMAGE IS AVAILABLE	THERE'S AN AVAILABLE TV SCREEN IMAGE	
AVNCS EXTRACT AVAIL IN FLT/USABLE/REACHABLE/ ATTAINABLE	NOUN + NOUN + AVAIL + IN + NOUN	LAPTOP BATTERY AVAILABLE IN SHOP	THE BATTERY OF THE LAPTOP IS AVAILABLE IN THE SHOP	THE LAPTOP'S BATTERY IS AVAILABLE IN SHOP	THERE'S AN AVAILABLE LAPTOP BATTERY IN THE SHOP	
		PARK GUARD AVAILABLE IN OFFICE	THE GUARD OF THE PARK IS AVAILABLE IN THE OFFICE	THE PRISON'S GUARD IS AVAILABLE IN OFFICE (?)	THERE'S AN AVAILABLE PRISON GUARD IN (THE) OFFICE	
		DIVORCE ATTORNEY AVAILABLE IN COURTROOM	THE DIVORCE ATTORNEY IS AVAILABLE IN THE COURTROOM	-	THERE'S AN AVAILABLE DIVORCE ATTORNEY IN COURTROOM	
		EMERGENCY EXIT AVAILABLE IN BUILDING	THE EMERGENCY EXIT IS AVAILABLE IN THE BUILDING	-	THERE'S AN EMERGENCY EXIT AVAILABLE IN (THE) BUILDING	
		FISH BOAT AVAILABLE IN STORE	THE FISH BOAT IS AVAILABLE IN THE STORE	-	THERE'S AN AVAILABLE FISH BOAT AVAILABLE IN (THE) STORE	
		CAR SEAT AVAILABLE IN VEHICULE	THE CAR SEAT IS AVAILABLE IN THE VEHICULE	-	THERE'S AN AVAILABLE CAR SEAT IN (THE) VEHICULE	
		ART MUSEUM PAINTINGS AVAILABLE IN GALLERY	THE PAINTINGS IN THE ART MUSEUM ARE AVAILABLE IN GALLERY	THE ART MUSEUM'S PAINTINGS ARE AVAILABLE IN GALLERY	THERE ARE AVAILABLE ART MUSEUM PAINTINGS IN (THE)GALLERY	THERE ARE ART MUSEUM PAINTINGS AVAILABLE IN GALLERY
		SAND CASTLE KIT AVAILABLE IN BAG	THE SAND CASTLE KIT IS AVAILABLE IN THE BAG	THE SAND CASTLE'S KIT IS AVAILABLE IN BAG (?)	THERE'S AN AVAILABLE SAND CASTLE KIT IN (THE) BAG	THERE IS A SAND CASTLE KIT AVAILABLE IN (THE) BAG

PACK 1 AVAIL IN FLT	NOUN + NOUN + NOUN + AVAIL + IN + NOUN	CITY SKYSCRAPERS IMAGES AVAILABLE IN STOCK	THE IMAGES OF THE CITY SKYSCRAPERS ARE AVAILABLE IN STOCK	THE CITY SKYSCRAPERS' IMAGES ARE AVAILABLE IN STOCK	THERE ARE AVAILABLE CITY SCKYSCRAPERS IMAGES IN (THE) BAG	THERE ARE CITY SKYSCRAPERS IMAGES AVAIALBLE IN THE BAG
		LAPTOP SCREEN CLEANER AVAILABLE IN DRAWER	THE CLEANER OF THE LAPTOP SCREEN IS AVAILAVBLE IN DRAWER	THE LAPTOP'S SCREEN CLEANER IS AVAILABLE IN DRAWER	THERE IS AN AVAILABLE LAPTOP SCREEN CLEANER IN (THE) DRAWER	THERE IS A LAPTOP SCREEN CLEANER AVAILABLE IN (THE) DRAWER
		CRYSTAL FLOWER VASE AVAILABLE IN BACK	THE CRYSTAL FLOWER VASE IS AVAILABLE IN THE BACK	-	THERE IS AN AVAILABLE CRYSTAL FLOWER VASE IN THE BACK	THERE IS A CRYSTAL FLOWER VASE AVAILABLE IN THE BACK
		KITCHEN KNIFE BLADE AVAILABLE IN CUPBOARD	THE KNIFE BLADE IN THE KITCHEN IS AVAILABLE IN THE CUPBOARD	THE KITCHEN KNIFE'S BLADE IS AVAILABLE IN THE CUPBOARD	THERE IS AN AVAILABLE KITCHEN KNIFE BLADE IN (THE) CUPBOARD	THERE IS KITCHEN KNIFE BLADE AVAILABLE IN (THE) CUPBOARD

E

The Complete list of MCL and MNL messages proposed for Experiment 1 & 2

CL11	Chalk board available	NOUN + NOUN + AVAIL
CL12	Tree house available	NOUN + NOUN + AVAIL
CL13	Bus stop available	NOUN + NOUN + AVAIL
CL14	Fish tank available	NOUN + NOUN + AVAIL
CL15	City parking available	NOUN + NOUN + AVAIL
CL16	School bus available	NOUN + NOUN + AVAIL
CL21	Mobile car holder available	NOUN + NOUN + NOUN + AVAIL
CL22	Crime scene officer available	NOUN + NOUN + NOUN + AVAIL
CL23	Media room furniture available	NOUN + NOUN + NOUN + AVAIL
CL24	Jewelry store display available	NOUN + NOUN + NOUN + AVAIL
CL25	College art department available	NOUN + NOUN + NOUN + AVAIL
CL26	Restaurant sea view available	NOUN + NOUN + NOUN + AVAIL
CL31	Emergency exit available in building	NOUN + NOUN + AVAIL + IN + NOUN
CL32	Baby seat available in vehicule	NOUN + NOUN + AVAIL + IN + NOUN
CL33	Pizza boxes available in store	NOUN + NOUN + AVAIL + IN + NOUN
CL34	Coffee mugs available in cupboard	NOUN + NOUN + AVAIL + IN + NOUN
CL35	Art museum paintings available in gallery	NOUN + NOUN + AVAIL + IN + NOUN
CL36	Rock singers available in band	NOUN + NOUN + AVAIL + IN + NOUN
CL41	Office writing supplies available in catalogue	NOUN + NOUN + NOUN + AVAIL + IN + NOUN
CL42	Sand castle kit available in e-store	NOUN + NOUN + NOUN + AVAIL + IN + NOUN
CL43	City skyscrapers postcards available in frame	NOUN + NOUN + NOUN + AVAIL + IN + NOUN
CL44	Art museum paintings available in gallery	NOUN + NOUN + NOUN + AVAIL + IN + NOUN
CL45	Bathroom design ideas available in magazine	NOUN + NOUN + NOUN + AVAIL + IN + NOUN
CL46	Kitchen knife holder available in house	NOUN + NOUN + NOUN + AVAIL + IN + NOUN
CL51	Left container 20 kilos maximum available	ADJ + NOUN + NUM + NOUN + NOUN + AVAIL
CL52	White car 2 seats maximum available	ADJ + NOUN + NUM + NOUN + NOUN + AVAIL
CL53	New book 2 pasta recipes available	ADJ + NOUN + NUM + NOUN + NOUN + AVAIL
CL54	Left cage 2 birds minimum available	ADJ + NOUN + NUM + NOUN + NOUN + AVAIL
CL55	Old building 4 defence guards available	ADJ + NOUN + NUM + NOUN + NOUN + AVAIL
CL56	Grey room 4 exercice bikes available	ADJ + NOUN + NUM + NOUN + NOUN + AVAIL
CL61	Yellow hall 2 movie posters minimum available	ADJ + NOUN + NUM + NOUN + NOUN + NOUN + AVAIL
CL62	Roman pool 3 water fountain heads available	ADJ + NOUN + NUM + NOUN + NOUN + NOUN + AVAIL
CL63	Clean neighbourhood 2 house garage doors available	ADJ + NOUN + NUM + NOUN + NOUN + NOUN + AVAIL
CL64	Pretty store 3 paper notebook sizes available	ADJ + NOUN + NUM + NOUN + NOUN + NOUN + AVAIL
CL65	Local bakery 2 vanilla wedding cakes available	ADJ + NOUN + NUM + NOUN + NOUN + NOUN + AVAIL
CL66	Wooden tray 5 door handles minimum available	ADJ + NOUN + NUM + NOUN + NOUN + NOUN + AVAIL
NL11	There is an available chalk board	NOUN + NOUN + AVAIL
NL12	There is an available tree house	NOUN + NOUN + AVAIL
NL13	There is an available bus stop	NOUN + NOUN + AVAIL
NL14	There is an available fish tank	NOUN + NOUN + AVAIL
NL15	There is an available city parking	NOUN + NOUN + AVAIL
NL16	There is an available school bus	NOUN + NOUN + AVAIL
NL21	There is an available mobile car holder	NOUN + NOUN + NOUN + AVAIL
NL22	There is an available crime scene officer	NOUN + NOUN + NOUN + AVAIL
NL23	There is available media room furniture	NOUN + NOUN + NOUN + AVAIL
NL24	There is an available jewelry store display	NOUN + NOUN + NOUN + AVAIL
NL25	There is an available college art department	NOUN + NOUN + NOUN + AVAIL
NL26	There is an available restaurant sea view	NOUN + NOUN + NOUN + AVAIL
NL31	There is an emergency exit available in the building	NOUN + NOUN + AVAIL + IN + NOUN
NL32	There is a baby seat available in the vehicle	NOUN + NOUN + AVAIL + IN + NOUN
NL33	There are pizza boxes available in the store	NOUN + NOUN + AVAIL + IN + NOUN
NL34	There are coffee mugs available in the cupboard	NOUN + NOUN + AVAIL + IN + NOUN
NL35	There is a water hose available in the garden	NOUN + NOUN + AVAIL + IN + NOUN
NL36	There are rock singers available in the band	NOUN + NOUN + AVAIL + IN + NOUN
NL41	There are office writing supplies available in the catalogue	NOUN + NOUN + NOUN + AVAIL + IN + NOUN
NL42	There is a sand castle kit available in the e-store	NOUN + NOUN + NOUN + AVAIL + IN + NOUN
NL43	There are city skyscrapers postcards available in the frame	NOUN + NOUN + NOUN + AVAIL + IN + NOUN
NL44	There are art museum paintings available in the gallery	NOUN + NOUN + NOUN + AVAIL + IN + NOUN
NL45	There are bathroom design ideas available in the magazine	NOUN + NOUN + NOUN + AVAIL + IN + NOUN
NL46	There is a kitchen knife holder available in the house	NOUN + NOUN + NOUN + AVAIL + IN + NOUN
NL51	There are maximum 20 kilos available in the left container	ADJ + NOUN + NUM + NOUN + NOUN + AVAIL
NL52	There are maximum 2 seats available in the white car	ADJ + NOUN + NUM + NOUN + NOUN + AVAIL
NL53	There are 2 pasta recipes available in new book	ADJ + NOUN + NUM + NOUN + NOUN + AVAIL
NL54	There are minimum 2 birds available in left cage	ADJ + NOUN + NUM + NOUN + NOUN + AVAIL
NL55	There are 4 defence guards available in old tower	ADJ + NOUN + NUM + NOUN + NOUN + AVAIL
NL56	There are 4 exercice bikes available in grey room	ADJ + NOUN + NUM + NOUN + NOUN + AVAIL

NL61	There are minimum 2 movie posters available in the yellow hall	ADJ + NOUN + NUM + NOUN + NOUN + NOUN + AVAIL
NL62	There are maximum 3 fountain heads available in the Roman pool	ADJ + NOUN + NUM + NOUN + NOUN + NOUN + AVAIL
NL63	There are 2 house garage doors available in the clean neighbourh	ADJ + NOUN + NUM + NOUN + NOUN + NOUN + AVAIL
NL64	There are 3 paper notebook sizes available in pretty store	ADJ + NOUN + NUM + NOUN + NOUN + NOUN + AVAIL
NL65	There are 2 vanilla wedding cakes available in local bakery	ADJ + NOUN + NUM + NOUN + NOUN + NOUN + AVAIL
NL66	There are minimum 5 door handles available in wooden tray	ADJ + NOUN + NUM + NOUN + NOUN + NOUN + AVAIL
IN CL11	Math teacher available	NOUN + NOUN + AVAIL
IN CL12	Camp fire available	NOUN + NOUN + AVAIL
IN CL13	Car meter available	NOUN + NOUN + AVAIL
IN CL21	Car steering wheel available	NOUN + NOUN + NOUN + AVAIL
IN CL22	Police patrol car available	NOUN + NOUN + NOUN + AVAIL
IN CL23	Children movie snacks available	NOUN + NOUN + NOUN + AVAIL
IN CL31	Elevator door available in entrance	NOUN + NOUN + AVAIL + IN + NOUN
IN CL32	Spare wheel available in car	NOUN + NOUN + AVAIL + IN + NOUN
IN CL33	Frozen fries available in fridge	NOUN + NOUN + AVAIL + IN + NOUN
IN CL41	Women gold jewelry available in picture	NOUN + NOUN + NOUN + AVAIL + IN + NOUN
IN CL42	Children story book available in e-store	NOUN + NOUN + NOUN + AVAIL + IN + NOUN
IN CL43	Paper plane pictures available in frame	NOUN + NOUN + NOUN + AVAIL + IN + NOUN
IN CL51	Green container 30 kilos maximum available	ADJ + NOUN + NUM + NOUN + NOUN + AVAIL
IN CL52	White car 7 seats minimum available	ADJ + NOUN + NUM + NOUN + NOUN + AVAIL
IN CL53	Kitchen area 3 chefs minimum available	ADJ + NOUN + NUM + NOUN + NOUN + AVAIL
IN CL61	Old theater 2 popcorn machines minimum available	ADJ + NOUN + NUM + NOUN + NOUN + NOUN + AVAIL
IN CL62	Roman pool 3 plastic chairs minimum available	ADJ + NOUN + NUM + NOUN + NOUN + NOUN + AVAIL
IN CL63	Big house 2 sports cars minimum available	ADJ + NOUN + NUM + NOUN + NOUN + NOUN + AVAIL
IN NL14	There is an available TV screen	NOUN + NOUN + AVAIL
IN NL15	There is an available food truck	NOUN + NOUN + AVAIL
IN NL16	There is an available school play	NOUN + NOUN + AVAIL
IN NL24	There is an available jewelry store assistant	NOUN + NOUN + NOUN + AVAIL
IN NL25	There is an available sports class teacher	NOUN + NOUN + NOUN + AVAIL
IN NL26	There is an available fish restaurant waiter	NOUN + NOUN + NOUN + AVAIL
IN NL34	There is a fork set available in the cupboard	NOUN + NOUN + AVAIL + IN + NOUN
IN NL35	There is a wood table available in the garden	NOUN + NOUN + AVAIL + IN + NOUN
IN NL36	There is a TV screen available in the background	NOUN + NOUN + AVAIL + IN + NOUN
IN NL44	There is a marble statue display available in the museum	NOUN + NOUN + NOUN + AVAIL + IN + NOUN
IN NL45	There is kitchen design plan available in the picture	NOUN + NOUN + NOUN + AVAIL + IN + NOUN
IN NL46	There is a vegetable chopping board available in the kitchen	NOUN + NOUN + NOUN + AVAIL + IN + NOUN
IN NL54	There are maximum 3 eagles available in the big cage	ADJ + NOUN + NUM + NOUN + NOUN + AVAIL
IN NL55	There are minimum 3 cars available in the old courtyard	ADJ + NOUN + NUM + NOUN + NOUN + AVAIL
IN NL56	There are minimum 3 posters available in the small gym	ADJ + NOUN + NUM + NOUN + NOUN + AVAIL
IN NL64	There are minimum 10 architecture books available in the online s	ADJ + NOUN + NUM + NOUN + NOUN + NOUN + AVAIL
IN NL65	There are maximum 4 apple pies available in the small bakery	ADJ + NOUN + NUM + NOUN + NOUN + NOUN + AVAIL
IN NL66	There are maximum 3 wood boards available in the small space	ADJ + NOUN + NUM + NOUN + NOUN + NOUN + AVAIL
Filler CL11	Ceiling lamp available	NOUN + NOUN + AVAIL
Filler CL12	Pool ladder available	NOUN + NOUN + AVAIL
Filler CL13	Lake house available	NOUN + NOUN + AVAIL
Filler CL21	Plastic coffee cup available	NOUN + NOUN + NOUN + AVAIL
Filler CL22	Chicken soup recipe available	NOUN + NOUN + NOUN + AVAIL
Filler CL23	Plane seat belt available	NOUN + NOUN + NOUN + AVAIL
Filler CL31	Video games available in bookstore	NOUN + NOUN + AVAIL + IN + NOUN
Filler CL32	Mushroom sauce available in kitchen	NOUN + NOUN + AVAIL + IN + NOUN
Filler CL33	Cookie bag available in cupboard	NOUN + NOUN + AVAIL + IN + NOUN
Filler CL41	Metal paper clips available in drawer	NOUN + NOUN + NOUN + AVAIL + IN + NOUN
Filler CL42	Family camping tent available in beach	NOUN + NOUN + NOUN + AVAIL + IN + NOUN
Filler CL43	Christmas holiday decoration available in station	NOUN + NOUN + NOUN + AVAIL + IN + NOUN
Filler CL51	New train 3 restaurants minimum available	ADJ + NOUN + NUM + NOUN + NOUN + AVAIL
Filler CL52	Big cage 4 birds minimum available	ADJ + NOUN + NUM + NOUN + NOUN + AVAIL
Filler CL53	Green field 30 trees minimum available	ADJ + NOUN + NUM + NOUN + NOUN + AVAIL
Filler CL61	Big fridge 2 soda cans minimum available	ADJ + NOUN + NUM + NOUN + NOUN + NOUN + AVAIL
Filler CL62	Small gift 2 ribbon colors maximum available	ADJ + NOUN + NUM + NOUN + NOUN + NOUN + AVAIL
Filler CL63	New shop 3 fish tanks minimum available	ADJ + NOUN + NUM + NOUN + NOUN + NOUN + AVAIL
Filler NL14	There is an available ceiling lamp	NOUN + NOUN + AVAIL
Filler NL15	There is an available pool ladder	NOUN + NOUN + AVAIL
Filler NL16	There is an available lake house	NOUN + NOUN + AVAIL
Filler NL24	There is an available plastic coffee cup	NOUN + NOUN + NOUN + AVAIL
Filler NL25	There is an available chicken soup recipe	NOUN + NOUN + NOUN + AVAIL
Filler NL26	There are available plane seat belts	NOUN + NOUN + NOUN + AVAIL
Filler NL34	There are video games available in the bookstore	NOUN + NOUN + AVAIL + IN + NOUN
Filler NL35	There a mushroom sauce available in the kitchen	NOUN + NOUN + AVAIL + IN + NOUN
Filler NL36	There is a cookie bag available in the cupboard	NOUN + NOUN + AVAIL + IN + NOUN
Filler NL44	There are metal paper clips available in the drawer	NOUN + NOUN + NOUN + AVAIL + IN + NOUN
Filler NL45	There is a family camping tent available in the beach	NOUN + NOUN + NOUN + AVAIL + IN + NOUN
Filler NL46	There is Christmas holiday decoration available in the station	NOUN + NOUN + NOUN + AVAIL + IN + NOUN
Filler NL54	There are minimum 3 restaurants available in the new train	ADJ + NOUN + NUM + NOUN + NOUN + AVAIL
Filler NL55	There are minimum 4 birds available in the big cage	ADJ + NOUN + NUM + NOUN + NOUN + AVAIL
Filler NL56	There are minimum 30 trees available in the green field	ADJ + NOUN + NUM + NOUN + NOUN + AVAIL
Filler NL64	There are minimum 2 soda cans available in the big fridge	ADJ + NOUN + NUM + NOUN + NOUN + NOUN + AVAIL
Filler NL65	There are maximum 2 ribbon colors available in the small gift	ADJ + NOUN + NUM + NOUN + NOUN + NOUN + AVAIL
Filler NL66	There are minimum 3 fish tanks available in the new shop	ADJ + NOUN + NUM + NOUN + NOUN + NOUN + AVAIL

F

A Sample List of Images from Each Difficulty Level in Experiment 1 & 2 and their Corresponding Messages

NOUN + NOUN + AVAIL

There is an available tree house

NOUN + NOUN + NOUN + AVAIL

There is an available restaurant sea view

NOUN + NOUN + AVAIL + IN + NOUN

There is an emergency exit available in the building

NOUN + NOUN + NOUN + AVAIL + IN + NOUN

There are city skyscrapers postcards available in the frame

ADJ + NOUN + NUM + NOUN + NOUN + AVAIL

There are 2 pasta recipes available in new book

ADJ + NOUN + NUM + NOUN + NOUN + NOUN + AVAIL

There are maximum 3 fountain heads available in the Roman pool

G

Forms Used in Experiment 1 & 2

Formulaire de Consentement libre, éclairé et express

Expériences comportementales en psychologie cognitive/psycholinguistique

Airbus Operations S.A.S
Unité CNRS 5263 « CLLE-ERSS»
Maison de la Recherche
5 allées Antonio Machado, 31058 Toulouse cedex 9

Je certifie avoir donné mon accord pour participer à une étude comportementale de psycholinguistique. J'accepte volontairement de participer à cette étude et je comprends que ma participation n'est pas obligatoire et que je peux stopper ma participation à tout moment sans avoir à me justifier ni encourir aucune responsabilité. Mon consentement ne décharge pas les organisateurs de la recherche de leurs responsabilités et je conserve tous mes droits garantis par la loi.

Au cours de cette expérience, j'accepte que soient recueillies des données chronométriques sur mes réponses. Je comprends que les informations recueillies sont strictement confidentielles et à usage exclusif des investigateurs concernés.

J'ai été informé(e) que mon identité n'apparaîtra dans aucun rapport ou publication et que toute information me concernant sera traitée de façon confidentielle. J'accepte que les données enregistrées à l'occasion de cette étude puissent être conservées dans une base de données et faire l'objet d'un traitement informatisé non nominatif par l'Unité CNRS 5263 et Airbus Operations S.A.S. J'ai bien noté que le droit d'accès prévu par la loi « informatique et libertés » s'exerce à tout moment auprès de l'unité CNRS 5263 et Airbus Operations S.A.S.

Nom de l'expérience : NL and CL Processing

Date : ....…………........

Nom du volontaire : …………………………………………………

Signature du volontaire (précédée de la mention « lu et approuvé ») :

Nom de l'expérimentateur responsable: Nataly Jahchan

Signature de l'expérimentateur :

G1 - *General Ethics and Compliance Consent Form*

Participant Sheet (NL and CL Comprehension State-action)

Date and participant number:

Age:

Gender:

Native language:

Familiarity with Airbus Controlled language or any controlled specialized language (None-Beginner-Intermediate-Expert):

English placement:

If you’re a pilot or student pilot please describe your level, qualifications, hours of flight etc:

G2 - *Participant Sheet*

UNIVERSITÉ TOULOUSE Jean Jaurès

UMR 5263
COGNITION, LANGUES, LANGAGE, ERGONOMIE

RECU DE PARTICIPATION

Je soussigné(e) ______________________ certifie avoir reçu 1 clef USB

au titre de ma participation à une expérience en psycholinguistiques,

Sous la direction de Mme Anne CONDAMINES Directrice de Recherche au CNRS

Fait à Toulouse le

Signature du participant

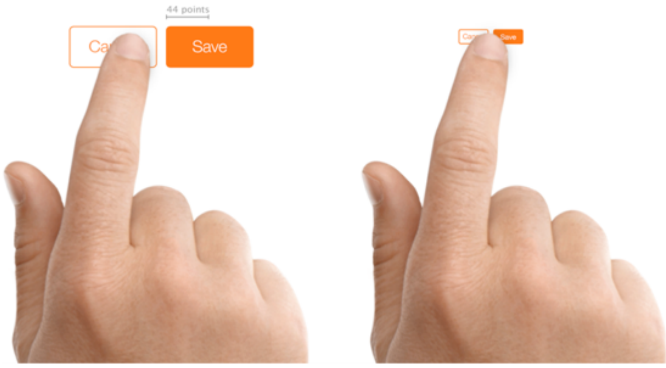
Adresse ou téléphone du participant

G3 - *Participation Receipt*

H

Gestalt Design Principles Used in the Construction of the Interface in Experiment 3

1. The tactile button sizes needed to be at least 44 pixels by 44 pixels so they can be accurately tapped with a finger (Apple standards), or 34px X 34px (Microsoft standards) for mobile touch screens.



Minimum Button Size

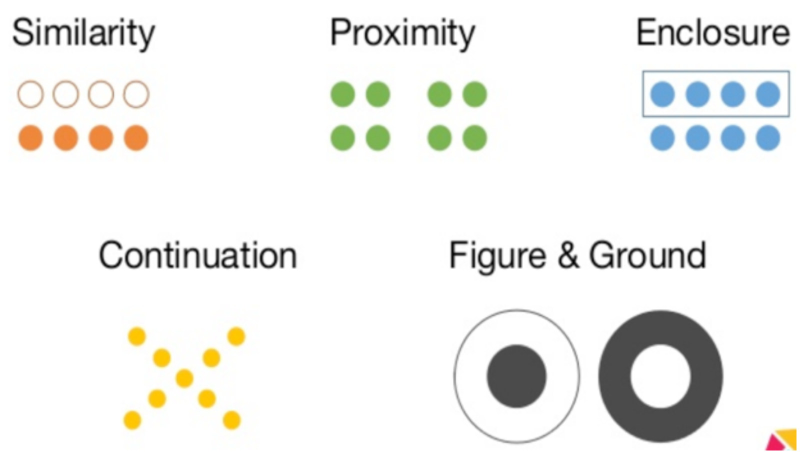
2. MIT Touch Lab study found that averages for finger pads are between 10–14mm and fingertips are 8–10mm, making 10mm x 10mm a good minimum touch target. So, we had to respect padding which is the space between each button to achieve optimum accuracy of the touch area and distinction of buttons.

Average Fingertip pad

3. The use of the Gestalt Design Principles:

> "*Gestalt psychology – an influential theory of perception early in the twentieth century – proposed that perception was determined not by the elemental sensations of light and dark but by laws of similarity, good continuation (analogous to smoothness), closure, symmetry, etc. that grouped such elements within a larger visual context.*" A. Das, in Encyclopedia of Neuroscience, 2009

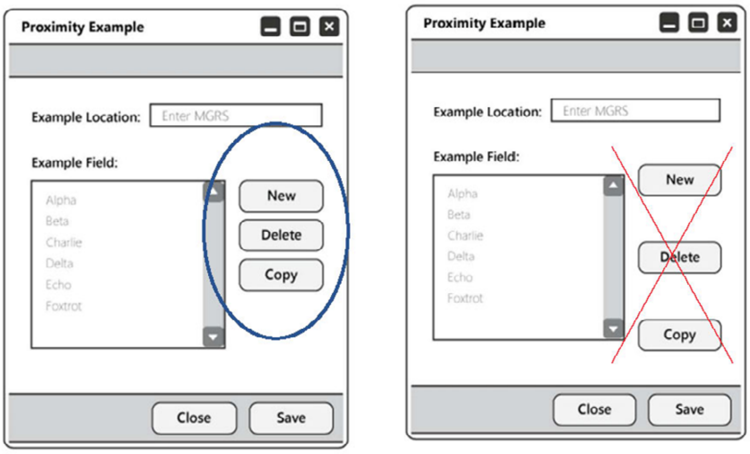
Here is a visual example of the main Gestaltian design principles.



Overview of Gestalt Design Principles[1]

- The first principle used was the principle of proximity. Distance between objects effects our perception of the objects, and objects that are close together are perceived as being grouped.

1. Slideshare.net. Gestalt Principles in UI Design. [online] Available at: https://www.slideshare.net/TDdesign/gestalt-principles-in-ui-design [Accessed 13 Dec. 2018].



Example of good and bad use of the proximity principle[2]

- The second principle used is the principle of similarity. Objects that look similar in shape and style appear to be grouped and share the same function.
- Third principle used is the principle of enclosure. Things that have a boundary around them are seen as grouped

With those guidelines in mind we designed a first version of the interface below which included 4 zones in enclosures and the messages at the bottom.

2. Ezer, N. Improving the User Interface through Gestalt Design Principles.

I

Complete List of Stimuli for Experiment 3

Stimuli	Zone	Language Type
Airflow ....High	AIR	MCL
Airflow ....Normal	AIR	MCL
Airflow ....Low	AIR	MCL
Set Airflow to High	AIR	MNL
Set Airflow to Normal	AIR	MNL
Set Airflow to Low	AIR	MNL
Crossbleed ....Auto	AIR	MCL
Set Crossbleed on Auto	AIR	MNL
Crossbleed ....Open	AIR	MCL
Crossbleed ....Close	AIR	MCL
Open Crossbleed	AIR	MNL
Close Crossbleed	AIR	MNL
High Ventilation ....On	AIR	MCL
High Ventilation ....Off	AIR	MCL
Turn On High Ventilation	AIR	MNL
Turn Off High Ventilation	AIR	MNL
Passenger Oxygen Masks.............Deploy	AIR	MCL
Cabin Crew Oxygen Masks ....Deploy	AIR	MCL
Pilot Oxygen Masks ....Deploy	AIR	MCL
Deploy Passenger Oxygen Masks	AIR	MNL
Deploy Cabin Crew Oxygen Masks	AIR	MNL
Deploy Pilot Oxygen Masks	AIR	MNL
Left Tank Main Pump ....On	FUEL	MCL
Left Tank Main Pump ....Off	FUEL	MCL
Turn On Left Tank's Main Pump	FUEL	MNL
Turn Off Left Tank's Main Pump	FUEL	MNL
Left Tank Standby Pump ....On	FUEL	MCL
Left Tank Standby Pump ....Off	FUEL	MCL
Turn On the Left Tank's Standby Pump	FUEL	MNL
Turn Off the Left Tank's Standby Pump	FUEL	MNL
Right Tank Main Pump ....On	FUEL	MCL
Right Tank Main Pump ....Off	FUEL	MCL
Turn On the Right Tank's Main Pump	FUEL	MNL
Turn Off the Right Tank's Main Pump	FUEL	MNL
Right Tank Standby Pump ....On	FUEL	MCL
Right Tank Standby Pump ....Off	FUEL	MCL
Turn On the Right Tank's Standby Pump	FUEL	MNL
Turn Off the Right Tank's Standby Pump	FUEL	MNL
Center Tank Left Pump ....On	FUEL	MCL

Stimuli	Zone	Language Type
Center Tank Left Pump ....Off	FUEL	MCL
Turn On the Center Tank's Left Pump	FUEL	MNL
Turn Off the Center Tank's Left Pump	FUEL	MNL
Center Tank Right Pump ....On	FUEL	MCL
Center Tank Right Pump ....Off	FUEL	MCL
Turn On the Center Tank's Right Pump	FUEL	MNL
Turn Off the Center Tank's Right Pump	FUEL	MNL
Center Tank Feed ....Manual	FUEL	MCL
Center Tank Feed ....Auto	FUEL	MCL
Set Center Tank Feed on Manual	FUEL	MNL
Set Center Tank Feed on Auto	FUEL	MNL
Crossfeed A ....On	FUEL	MCL
Crossfeed A ....Off	FUEL	MCL
Turn On Crossfeed A	FUEL	MNL
Turn Off Crossfeed A	FUEL	MNL
Crossfeed B ....On	FUEL	MCL
Crossfeed B ....Off	FUEL	MCL
Turn On Crossfeed B	FUEL	MNL
Turn Off Crossfeed B	SPEED	MNL
Speed ....500 KT	SPEED	MCL
Speed ....310 KT	SPEED	MCL
Speed ....250 KT	SPEED	MCL
Speed ....140 KT	SPEED	MCL
Set Speed to 500 KT	SPEED	MNL
Set Speed to 310 KT	SPEED	MNL
Set Speed to 250 KT	SPEED	MNL
Set Speed to 140 KT	SPEED	MNL
Engine 1 ....On	ENGINE	MCL
Engine 1 ....Off	ENGINE	MCL
Turn On Engine 1	ENGINE	MNL
Turn Off Engine 1	ENGINE	MNL
Engine 2 ....On	ENGINE	MCL
Engine 2 ....Off	ENGINE	MCL
Turn On Engine 2	ENGINE	MNL
Turn Off Engine 2	ENGINE	MNL
All Engine Bleed ....On	ENGINE	MCL
All Engine Bleed ....Off	ENGINE	MCL
Turn On All Engine Bleed	ENGINE	MNL
Turn Off All Engine Bleed	ENGINE	MNL
Engine 1 Agent 1 ....Discharge	ENGINE	MCL
Discharge Agent 1 of Engine 1	ENGINE	MNL
Engine 1 Agent 2 ....Discharge	ENGINE	MCL
Discharge Agent 2 of Engine 1	ENGINE	MNL
ENGINE 2 AGENT 1 ....Discharge	ENGINE	MCL

Stimuli	Zone	Language Type
Discharge Agent 1 of Engine 2	ENGINE	MNL
ENGINE 2 AGENT 2 ....Discharge	ENGINE	MCL
Discharge Agent 2 of Engine 2	ENGINE	MNL
Parking Brake ....Engage	BRAKES	MCL
Parking Brake ....Release	BRAKES	MCL
Engage Parking Brake	BRAKES	MNL
Release Parking Brake	BRAKES	MNL
Speed Brake ....Extend	BRAKES	MCL
Speed Brake ....Retract	BRAKES	MCL
Extend Speed Brake	BRAKES	MNL
Retract Speed Brake	BRAKES	MNL

FRENCH SUMMARY

Dans quelle mesure la simplification améliore la compréhension humaine ? Évaluations cognitives pour l'optimisation de la langue contrôlée dans les futurs cockpits des Airbus

Cette thèse de doctorat a été lancée par le département Facteurs Humains et Ergonomie du Design d'Airbus Operations SAS à Toulouse, en France, en collaboration avec le laboratoire CLLE (Cognition, Langues, Langage, Ergonomie) de l'Université Jean Jaurès de Toulouse.

Le laboratoire CLLE, basé à Toulouse, capitale française de l'aérospatial, a développé une base de connaissances dans le domaine des CNLs (Controlled Natural Language) et des corpus spécialisés, liés à l'espace et à l'aviation.

L'objectif principal de cette thèse est d'optimiser un langage contrôlé de cockpit Airbus existant (Airbus Cockpit Controlled Laguage) afin d'intégrer un nouveau langage amélioré dans la conception du futur cockpit. Le langage contrôlé actuel a été soigneusement construit pour éviter les ambiguïtés et la complexité (comme tous les langages contrôlés axés sur la compréhension et est conçu pour aider les pilotes à utiliser l'avion (à l'aide des interfaces écran du poste de pilotage) dans des situations normales et anormales (en cas d'urgence ou de panne). Ainsi, la nécessité d'une communication claire et non ambiguë est vitale dans les domaines critiques pour la sécurité. Ce langage et les règles qui le régissent ont été mis en place à une époque où la flexibilité de la conception n'était pas une option (par exemple, les écrans de petite taille limitent la longueur des mots et des phrases). Alors que nous entrons dans une ère, où une conception de cockpit disruptive, plus flexible pour les futurs appareils est possible, ces restrictions ne sont plus d'actualité : il n'est plus nécessaire que le langage contrôlé soit si codé et compact, ni ne suive des règles de simplification très strictes. Par conséquent, afin d'optimiser le langage existant, nous avons cherché à évaluer les niveaux appropriés de simplification qui permettraient une compréhension plus précise et plus rapide avec une réduction d'heures de formation pour les pilotes.

Pour ce faire, nous nous sommes d'abord intéressés au domaine des CNLs[1] pour avoir une vue d'ensemble des langages contrôlés existants, de leur contexte et de leurs règles. À partir de ces recherches, nous avons tenté de trouver des solutions d'optimisation du langage contrôlé Airbus, et, à l'aune de ces travaux, nous nous sommes également efforcés d'apporter une contribution originale au domaine.

En ce sens, notre travail relève du domaine de la linguistique appliquée. L'AILA[2] (Association Internationale de Linguistique Appliquée) la définit *« comme un domaine interdisciplinaire de recherche et de pratique traitant de problèmes pratiques de langage et de communication pouvant être identifiés, analysés ou résolus en appliquant les théories, méthodes et résultats de linguistique disponibles ou en développant de nouveaux cadres théoriques et méthodologiques en linguistique pour travailler sur ces problèmes ».*

Contrairement aux idées reçues, le domaine de la linguistique appliquée ne propose pas simplement des solutions aux problèmes pratiques issus des théories disponibles, mais pourrait également développer de nouveaux cadres théoriques et outils méthodologiques issus de différents domaines et sources pour traiter les problèmes de langage et de communication. Condamines & Narcy-Combes, 2015 propose le terme « science située »*: « [...] situer la science c'est entrer dans une perspective où la recherche n'est plus appliquée à un projet, mais où elle est une partie de ce projet et où ils se modifient réciproquement au fur et à mesure que le projet avance.* »

Par conséquent, nous commençons au préalable par trouver ou créer de nouvelles solutions à des problèmes concrets de la vie réelle (dans le cas présent, des problèmes industriels à enjeux considérables). Ces solutions ne sont pas de simples applications du savoir-faire linguistique, elles constituent également un moyen d'étudier et de faire progresser les fonctions et évaluations linguistiques dans des domaines relativement sous-développés

1. « CNL », terme interchangeable avec « CL » ou langage contrôlé.
2. AILA. [en ligne] Disponible à l'adresse : https://aila.info [Consulté le 4 décembre 2018].

et inconnus (du moins dans la communauté des sciences linguistiques globale) et où le langage joue un rôle essentiel pour assurer la sécurité (une interprétation erronée pouvant conduire à des conséquences potentiellement catastrophiques).

Pour trouver des solutions, il faut examiner le problème dans son contexte. En ce sens, nous nous intéressons plus particulièrement à la linguistique ergonomique (Condamines, 2018) dans laquelle les modèles / théories / hypothèses linguistiques sont utilisés dans des contextes de travail spécifiés (principalement dans l'industrie) pour atteindre efficacement des objectifs précis. Ces hypothèses et propositions sont issues de productions linguistiques réelles et doivent être évaluées à l'aide de techniques expérimentales et de tests d'acceptabilité.

Dans cette thèse, nous utilisons des outils de psycholinguistique et de psychologie cognitive ainsi que des techniques d'évaluation afin de confirmer ou d'infirmer des hypothèses linguistiques directement liées à un besoin industriel centré sur l'humain.

1. Contexte

1.1 Facteurs humains et ergonomie chez Airbus

Doté de ses divisions *Space*, *Defence* et *Helicopters*, Airbus est un constructeur d'avions commerciaux. À ce jour, il s'agit de la plus grande entreprise aéronautique et spatiale d'Europe et du leader mondial dans son domaine. Airbus conçoit, fabrique et fournit des produits, des services et des solutions aérospatiaux à ses clients à l'échelle mondiale.

Le département Facteurs humains et ergonomie d'Airbus est représenté par une équipe multidisciplinaire composée de linguistes, de psychologues cognitifs, de physiologistes et d'ergonomes cognitifs. Ils organisent les évaluations de l'équipement et des fonctions, des procédures en cours de conception, ainsi que le suivi du processus jusqu'à la phase de certification avec les autorités de l'aviation, telles que l'Agence européenne de la sécurité aérienne (EASA) ou la FAA (Federal Aviation Administration). Ils rédigent des rapports analysant différentes fonctions du point de vue des facteurs humains et offrent des conclusions sur le fonctionnement, la sécurité et les aspects ergonomiques de la fonction étudiée. Ces évaluations sont effectuées sur différentes parties de l'avion concernant les équipements dans le cockpit, la cabine et la maintenance[3].

Le département des facteurs humains mène également des recherches pour trouver et mettre en œuvre de nouvelles solutions et fonctionnalités pour la conception des futurs avions Airbus. De plus, ils formulent des recommandations pour la conception de fonctions, puis fournissent des documents techniques. L'objectif est de s'assurer que l'utilisateur final est pris en compte tout au long du processus de conception.

Le domaine des facteurs humains a une approche spécifique de l'ingénierie et de l'ergonomie car les ergonomes considèrent la question sur le plan humain et prennent en compte les multiples interactions possibles entre l'individu et son environnement physique et cognitif.

L'attention portée aux interactions homme-machine est extrêmement importante car des interfaces mal conçues peuvent entraîner des situations dangereuses et des risques pour la sécurité. La science des facteurs humains est devenue obligatoire dans certains domaines tels que les industries médicales, ceux des transports et de l'aviation. Les questions de sécurité aérienne soulignent l'importance du rôle des facteurs humains dans la validation des équipements et des fonctions qui seraient mis en œuvre sur les avions commerciaux. Au fil des ans, la technologie a progressé à un rythme rapide, les avions sont devenus des machines extrêmement sûres et les accidents mortels d'avions de ligne ont constamment diminué.

Selon le magazine trimestriel *Aero* de Boeing (QTR_02, 2007), *« dans les premiers jours du vol, environ 80% des accidents étaient causés par la machine et 20% par une erreur humaine. Aujourd'hui, cette statistique s'est inversée. Environ 80% des accidents d'avion sont dus à une erreur humaine (pilotes, contrôleurs aériens, mécaniciens, etc.) et 20% sont dus à des pannes de machines (équipements). »* (Cf. *Figure 1*) Même si les accidents d'aviation sont en diminution constante (grâce à une technologie plus avancée), les accidents qui surviennent de nos jours ont 80% de chances d'être causés par des erreurs humaines. C'est pourquoi il est plus que nécessaire de faire appel aux spécialistes des facteurs

3. Le département Facteurs Humains et Ergonomie de la conception dans lequel ce doctorat a été réalisé ne traite que de la conception de cockpit. Les autres départements Facteurs Humains d'Airbus traitent de la conception et de la maintenance des cabines.

humains et de l'ergonomie dans le processus de conception afin d'atténuer ces risques.

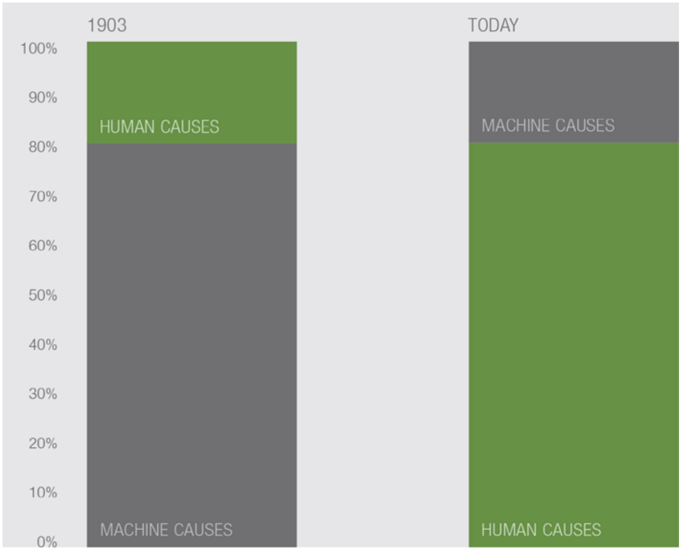


Figure 1. Les causes d'accidents dans l'aviation : 1903-2007

Plus concrètement, l'équipe des facteurs humains collabore étroitement avec les pilotes d'essai et les ingénieurs de vol d'Airbus, souvent sur les simulateurs de vol disponibles des différents avions, dans le but de tester des scénarios de vol sur des équipements et fonctions existants ou nouvellement introduits. Les pilotes d'essai, qui, dans la plupart des cas ont eu une longue carrière au sein de compagnies aériennes ou de l'armée à bord d'avions de combat, sont des experts de cet environnement de travail et des besoins en matière de pilotage. Ils sont indispensables à la certification et aux essais d'avions Airbus destinés à l'ensemble des pilotes de ligne.

1.2 Linguistique dans le domaine des facteurs humains et ergonomie

Les experts en facteurs humains utilisent la science et la connaissance générale des capacités et des limites humaines – avec l'expérience des pilotes de test – afin de déterminer les moyens les plus efficaces de concevoir le cockpit.

Les physiologistes s'assurent que la conception physique du poste de pilotage est adéquate, par exemple si les pilotes peuvent atteindre facilement tous les écrans et tous les boutons, même en cas de turbulence, ou si le champ de vision du pilote n'est pas obstrué par un instrument donné ou par les rayons du soleil. Les ergonomes cognitifs utilisent des techniques d'évaluation pour mesurer l'adéquation de la conception proposée sur les compétences cognitives humaines dans différents scénarios. Entre autres choses, les psychologues cognitifs mesurent les effets de la fatigue ou du stress sur la prise de décision dans les cas de charge de travail élevée.

Les linguistes, quant à eux, utilisent la recherche descriptive linguistique traditionnelle (théories syntaxique, sémantique, pragmatique et terminologique, etc.), la psychologie cognitive et les outils d'évaluation de l'ergonomie (analyses statistiques, questionnaires, etc.), le traitement automatique du langage naturel (pour la technologie des assistants virtuels, par exemple), et des outils psycholinguistiques pour développer des corpus linguistiques spécifiques à chaque domaine (en tenant compte des limites et des spécificités de l'aviation) et conçus pour être centrés sur l'humain (en tenant compte des capacités humaines de compréhension et de perception).

Ils doivent faire face à tous les problèmes de communication et opérationnels pouvant survenir dans un environnement de poste de pilotage. Le cockpit moderne en verre (comportant des affichages d'instrument de vol électroniques / numériques sur des écrans plutôt que des jauges mécaniques classiques) est un environnement de travail assez complexe (une formation et une expertise approfondies sont nécessaires pour pouvoir y être opérationnel) et les informations linguistiques y sont abondantes. (Cf. *Figure 2*).

Figure 2. Airbus A350 Glass Cockpit

Les différents moniteurs contiennent des messages et des balises destinés à être lus et compris par les pilotes. Certains de ces messages font référence à des boutons et des leviers dans le cockpit, qui possèdent eux-mêmes des balises de langage. En outre, les pilotes peuvent entendre des alertes sonores, qui sont essentiellement des messages d'avertissement qui signalent les dangers à venir ou qui leur fournissent des informations supplémentaires de vol.

Enfin, les pilotes aux commandes (*Pilot Flying*) et les pilotes en surveillance (*Pilot Monitoring*), auparavant appelés pilotes et copilotes, communiquent entre eux, avec l'avion lui-même, avec les contrôleurs aériens au sol, avec différents appareils en vol et avec le reste de l'équipage.

Par conséquent, au milieu de cet environnement de travail imprégné sur le plan linguistique, le travail du linguiste doit avant tout ne laisser aucune possibilité à une interprétation erronée du sens voulu ou à des ambiguïtés qui pourraient conduire à des situations potentiellement dangereuses aux enjeux très élevés.

Pour ce faire, les linguistes doivent construire un langage plus restreint (que le langage naturel) dans lequel la syntaxe et le lexique sont contrôlés de manière à réduire la complexité et l'ambiguïté pouvant conduire à une interprétation erronée et à une mauvaise communication. Dans la mesure du possible, ils devraient également élaborer des normes et des réglementations sur la manière dont ce langage contrôlé est utilisé dans différents contextes.

Selon l'organisation de l'aviation civile internationale (ICAO[4]), entre 1976 et 2000, plus de 1 100 passagers et membres d'équipage ont perdu la vie dans des accidents où les problèmes de langue ont joué un rôle déterminant (Mathews, 2004). Les accidents d'aviation ont presque toujours plusieurs facteurs contributifs. Les problèmes linguistiques qui jouent un rôle déterminant dans les accidents sont relativement peu connus ou ne sont pas suffisamment pris en compte dans les rapports d'accident. C'est pourquoi les linguistes et les spécialistes de l'aviation doivent continuer à travailler ensemble pour rendre l'utilisation du langage dans le poste de pilotage aussi intuitive que possible afin d'éviter des situations dangereuses.

4. ICAO. [en ligne] Disponible à l'adresse : https://www.icao.int/Pages/default.aspx [consulté le 10 décembre 2018].

1.3 Langues contrôlées chez Airbus

Actuellement, plusieurs langues contrôlées chez Airbus ont été mis en place et testés pour assurer une compréhension non ambiguë (éviter les **ambiguïtés** (interprétation multiple), les **inexactitudes** (interprétation inexacte), les **incohérences** (terminologie incohérente non standardisée) et les **insuffisances** (emploi d'un terme incorrect dans un contexte spécifique)) afin d'assurer la sécurité de la navigation, les besoins opérationnels et l'adaptabilité de l'interaction homme-machine à différentes situations dans le cockpit, la cabine et la maintenance :

- Langage contrôlé du cockpit *(Cockpit Controlled Language)* utilisé pour ECAM , PFD , MFD moniteurs
- GOLD utilisé pour l'OIS pour les communications en vol et en cabine
- ASD-STE (Anciennement AECMA-SE ou AECMA en abrégé) pour la maintenance au sol des appareils

Depuis le premier vol de l'Airbus A340 en 1991 jusqu'à l'introduction de l'Airbus A380 en 2004, il y eut un processus important de simplification et de normalisation visant à inclure les règles de la nouvelle langue contrôlée pour le cockpit. Cependant, ce langage contrôlé comporte plusieurs limites principalement dues aux :

- Écrans de petite taille (nombre limité de mots et de phrases)
- Caractère hautement codifié (non conforme à la syntaxe du langage naturel, très abrégé, typographiquement variable, inclut un code couleur, etc.), nécessitant donc une formation préalable du pilote afin de maîtriser parfaitement ce langage
- « Concept de famille » et de normalisation de la flotte d'Airbus qui ne devraient donc apporter aucune modification substantielle des interfaces entre deux générations d'avions, même si la nouvelle technologie le permet.

La *figure 3* est un exemple de différents messages trouvés à différents endroits dans l'un des corpus à portée de main (ce n'est pas une réplique exacte[5] d'une alarme).

5. Pour des raisons de confidentialité, les alarmes complètes Airbus ne peuvent pas être publiées. Les lignes de la *Figure 3* sont assemblées à partir de différentes alarmes et sont représentatives des divers types d'informations contenues dans les corpus.

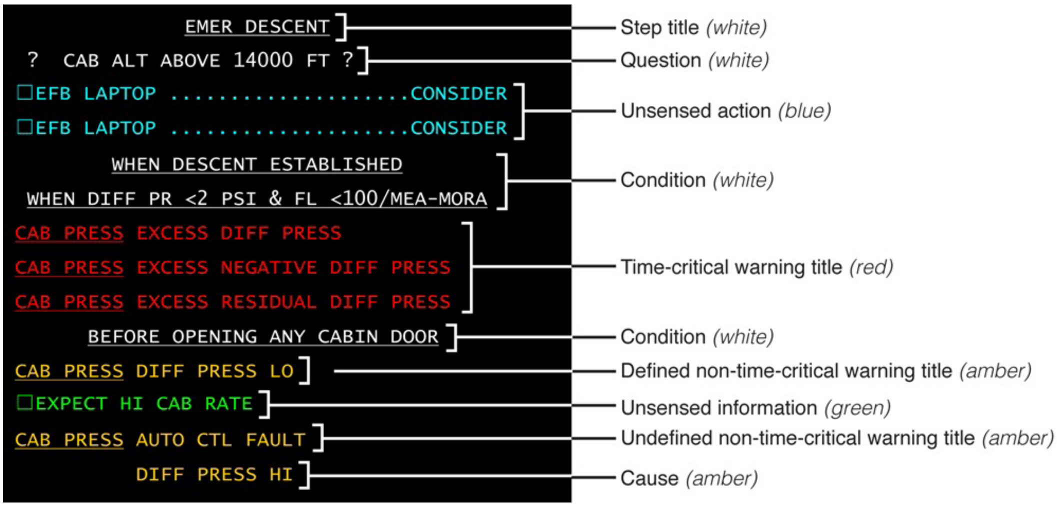


Figure 3. Exemple de différents types de messages dans le Cockpit CL

1.4 Introduction au sujet

Toutefois, étant donné que nous abordons actuellement une configuration de cockpit au caractère disruptif pour les futures générations d'avions, nous pourrions envisager différentes marges de flexibilité : moins de restrictions, des tailles d'écran plus grandes, moins de codage, etc.

En outre, « *depuis plus de dix ans, la communauté aéronautique internationale considère le concept d'opérations à pilote unique (*Single Pilot Operations - SPO*) comme une solution viable pour faire face aux coûts croissants associés au transport aérien commercial. Les progrès récents en communications, navigation, surveillance / technologies de gestion du trafic aérien et avionique (CNS + A) ont permis des niveaux d'automatisation plus élevés, créant une opportunité pour les avions commerciaux de transiter à SPO.* » (Lim et al. 2017). Par conséquent, il existe un besoin encore plus grand d'interfaces optimisées, intuitives et faciles à utiliser, avec le moins d'ambiguïtés linguistiques possible (en réduisant autant que possible toute forme d'interprétation erronée).

En conséquence, afin de tester et d'optimiser la compréhension, la perception et l'utilisation de langues contrôlées dans les cockpits, nous cherchons à mener des expériences comportementales en exploitant de nouveaux outils et de nouvelles recherches dans les domaines des sciences cognitives et des langues contrôlées ainsi qu'en appliquant des hypothèses linguistiques.

Nous allons cibler trois aspects principaux :

- Compréhension plus rapide
- Compréhension plus précise
- Réduction de temps de formation pilotes

1.5 Vers un nouveau langage contrôlé plus naturel (MNL)

La *figure 4* est une infographie qui résume notre question centrale et nos hypothèses. La figure oppose deux entités, le naturel et le contrôlé : d'un côté, en orange, le langage naturel, plus naturaliste, et de l'autre côté, en violet, le langage contrôlé, moins naturaliste. Si nous considérons d'abord le langage naturel sans aucun contrôle, nous tombons dans l'ambiguïté, l'utilisation abusive et l'incompréhension, ce qui ne convient pas à un opérateur humain et peut conduire à des actions erronées. Par conséquent, un contrôle et une simplification sont nécessaires pour éviter toute ambiguïté. Lorsque nous faisons cela, nous créons des règles standardisées qui limitent les ambiguïtés et forment un langage contrôlé.

Cependant, lorsque nous créons un langage contrôlé à partir de règles normalisées, nous devons savoir si ce langage :

- est assez expressif (pouvons-nous dire tout ce que nous devons dire, avec les mots justes ?),
- est-il efficace et utilisable (sommes-nous en mesure de communiquer efficacement certaines informations de manière claire et cohérente afin qu'elles soient utilisées sans effort ?),
- ce langage est-il facilement accessible avec une formation limitée (pouvons-nous l'enseigner facilement, est-il facile à apprendre car il possède des structures plus ou moins familières, et est-il suffisamment intuitif pour ne pas nécessiter la mémorisation de nouveaux codes ?).

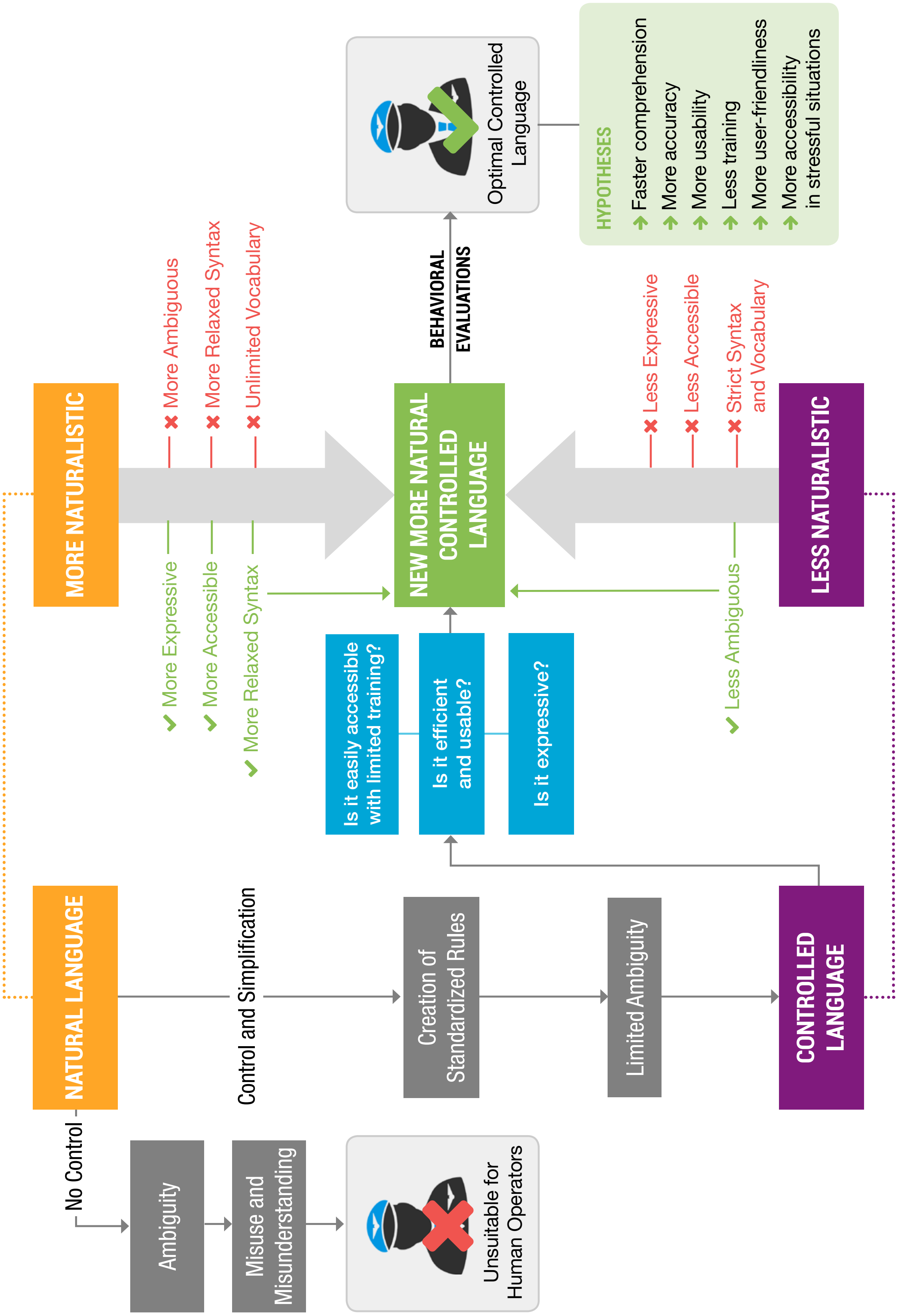


Figure 4. Question fondamentale et hypothèses

Afin de commencer à répondre à ces questions, nous devons trouver un terrain d'entente entre l'entité du langage étroitement contrôlé mais moins naturaliste et celle du langage non contrôlé mais plus naturaliste, car les deux entités présentent des attributs positifs :

- La plus naturaliste, bien que plus ambiguë en raison d'une syntaxe plus détendue et d'un vocabulaire illimité, est plus expressive et plus accessible car plus proche du langage naturel que nous utilisons au quotidien.
- D'autre part, la moins naturaliste, bien que plus restrictive parce qu'elle est moins expressive et accessible, est également moins ambiguë en raison de la syntaxe et du vocabulaire restreints.

Par conséquent, afin de créer un nouveau langage contrôlé plus naturel (langage contrôlé, moins codé, moins restreint mais plus optimisé), nous devons tirer parti des attributs positifs (en vert) des deux entités opposées (la plus naturaliste et la moins naturaliste). Nous le faisons en allant vers un langage plus naturel, en décodifiant petit à petit le langage codé actuel et en le complexifiant (paradoxalement) afin de le rendre plus naturel et plus accessible avec une formation limitée.

Après avoir proposé un nouveau langage contrôlé plus naturel (MNL), qui est essentiellement une version plus naturelle de la syntaxe du langage contrôlé plus codé (MCL) actuel, nous utilisons des méthodes et des outils comportementaux pour évaluer de manière empirique son efficacité en termes de compréhension et de performance. On suppose que ce nouveau langage qui tire parti des attributs positifs des deux entités opposées affiche cette efficacité empirique, permettant une compréhension plus rapide et plus précise, et constituant un langage plus utilisable, disponible avec un minimum de formation. Cela le rendrait plus accessible et convivial, en particulier dans des circonstances plus éprouvantes.

2. Résultats

Les 3 expériences réalisées dans cette thèse ont constitué une première tentative d'évaluation d'un langage contrôlé à l'aide de paradigmes expérimentaux étroitement contrôlés inspirés de méthodes issues des sciences cognitives, et plus spécifiquement de la psycholinguistique et de la psychologie cognitive.

Nous avons testé deux catégories de données : les déclarations d'information (ou disponibilité des fonctions) et les déclarations d'action (injonctions, instructions pour effectuer une action) à partir d'un langage hautement codifié et contrôlé par le poste de pilotage, destiné à guider, conseiller et donner des instructions aux pilotes afin de les aider à naviguer, et faire fonctionner l'appareil dans des situations normales et anormales.

Ce langage contrôlé était auparavant simplifié et standardisé afin d'éviter les ambiguïtés, la complexité et les incohérences au niveau syntaxique et lexical. Dans ces expériences, nous nous sommes particulièrement intéressés aux simplifications syntaxiques et aux économies linguistiques existantes, principalement celles qui éliminaient les mots fonctionnels (expériences 1 et 2) et qui imposaient des réductions structurelles ou elliptiques (expérience 3).

Puisque nous nous sommes interrogés sur les limites de la simplification et son rôle dans l'amélioration de la compréhension humaine (comme le suggère son principe) et sur le rôle de la naturalité dans un langage (ou un langage plus proche d'une INL (*Instance of Natural Language*), qui est théoriquement plus complexe), nous avons mis en place des paradigmes expérimentaux qui nous ont permis de comparer des structures syntaxiques similaires du langage hautement codifié, théoriquement très simple, à un langage plus naturel et théoriquement plus complexe.

Nous avons construit ce nouveau langage plus naturel en « décodifiant » les séquences existantes et en les rapprochant du langage naturel (en ajoutant des mots grammaticaux, en étant explicite, en utilisant une structure de phrase syntaxiquement correcte au lieu d'une séparation typographique, etc.).

Nous avons tenu compte de la difficulté et de la longueur de la phrase, ainsi que de la criticité temporelle (stress induit par un temps limité pour recréer un sentiment d'urgence).

Dans la première expérience, nous avons obtenu des résultats encourageants concernant les temps de réaction du langage contrôlé plus naturel et confirmé notre hypothèse. Les réponses aux stimuli MNL étaient significativement plus rapides que les stimuli MCL même si les stimuli plus naturels contenaient plus de mots et permettaient moins de temps de lecture. L'exactitude des réponses aux stimuli n'a pas été affectée pour les deux langues. Ainsi, notre hypothèse de précision en faveur du langage plus naturel a été rejetée. Cependant, elle pourrait s'expliquer par le fait que les ajouts aux messages (consistant à créer une structure de phrase à l'aide de mots-outils) n'offraient pas un changement significatif qui aurait pu affecter la précision, mais marquaient une divergence assez importante pour mesurer des différences de temps de réaction subtiles, mais suffisamment significatives. Cela suggère que la structure de phrase introduite par les mots facilite la compréhension au niveau cognitif.

Dans l'expérience 2, lorsque les mêmes stimuli ont été utilisés mais que nous avons accordé plus de temps à leur présentation, les effets significatifs observés dans l'expérience 1 ne se sont pas manifestés, ce qui suggère que la vitesse, et, dans une certaine mesure, le stress induit par la nécessité de réagir de manière urgente, jouent un rôle dans le traitement et la compréhension de l'information. Cela nous a conduit à introduire directement la variable de vitesse dans le protocole expérimental de l'expérience 3, qui n'utilisait pas les mêmes stimuli, car la nature de la tâche et le type de données étaient entièrement différents. Quoi qu'il en soit, cela nous a permis de tester l'effet de l'urgence induite sur les réponses et la compréhension.

En revanche, la difficulté syntaxique n'entraînait aucune interaction entre les deux langages dans les 6 conditions de difficulté de l'expérience 1 (une condition de difficulté linguistique n'a pas montré de temps de réaction plus rapide dans certaines conditions de difficulté, que dans d'autres). Les stimuli MNL étaient systématiquement plus rapides que ceux du MCL dans toutes les conditions de difficulté.

Cependant, la difficulté de la tâche elle-même qui se manifeste dans les stimuli disparates plus complexes a montré des temps de réaction nettement plus rapides pour les stimuli MNL (dans les expériences 1 et 2), contrairement aux stimuli concordants. De plus, même si nous n'observons pas de performance globale significative sur les temps de réaction dans les résultats de l'expérience 2, les stimuli non concordants montrent des effets significatifs sur la condition utilisant le MNL en comparaison avec la condition utilisant le MCL. Par conséquent, nous pouvons en conclure que, dans les cas de difficulté accrue de la tâche à traiter, le langage plus naturel aide à faciliter la compréhension.

Dans l'expérience 1, le MNL semble faciliter la compréhension des participants ayant un niveau d'anglais intermédiaire de base, ce qui suggère que les anglophones les plus faibles tireraient avantage d'une langue plus naturelle que les locuteurs confirmés. Bien que les anglophones de langue maternelle aient obtenu de meilleurs résultats en moyenne dans les conditions de MNL, cet effet n'était pas statistiquement significatif, ce qui a été vérifié plus fidèlement lors de l'expérience 3 avec un échantillon plus important d'anglophones de langue maternelle.

Les expériences 1 et 2 incluaient des tâches de congruence très proches des tâches de jugement traditionnelles dans les expériences comportementales. Elles ont fourni un environnement rigoureusement contrôlé pour tester nos hypothèses linguistiques. Néanmoins, l'inconvénient de telles expériences est que nous sommes limités à l'évaluation de la compréhension passive, principalement d'énoncés informatifs spécifiques. Il serait assez difficile d'évaluer la compréhension d'un ordre ou d'une instruction à l'aide de tâches de jugement traditionnelles.

C'est pourquoi nous avons créé une interface homme-machine méticuleusement conçue pour répondre à nos besoins d'évaluation de la compréhension des injonctions dans les tâches de performance en temps réel. Elle présentait l'avantage d'imiter une interface de cockpit potentielle avec quelques boutons et fonctions d'avion concoctés, dans une tâche pouvant être assimilée à du pseudo-pilotage (car il s'agissait d'une interface d'ordinateur et elle ne contenait pas de scénarios de pilotage réels), tout en restant dans le cadre de tests en laboratoire de paradigmes expérimentaux robustes. Encore une fois, la variable principale était l'efficacité contrastée des deux conditions linguistiques dans l'exécution de la tâche à effectuer, en plus de la variable de vitesse, de la complexité de la zone et de l'intuitivité du format elliptique.

Contrairement aux expériences 1 et 2, l'expérience 3 a montré que le MNL était significativement mieux compris que le MCL dans les deux conditions d'urgence. Cependant, une interaction significative en termes de vitesse et de temps de réaction a été observée : les messages en MNL n'en étaient que significativement plus rapides à traiter en cas d'urgence, mais pas dans les cas non urgents. Ceci est cohérent avec les résultats observés dans les deux premières expériences qui suggéraient également que la valeur ajoutée du MNL était plus clairement observée dans des conditions de temps critiques (stress induit par des situations de type urgent).

Il est intéressant de noter que même si les messages en MNL contenaient plus de mots, car ils étaient plus naturels et moins elliptiques, ils étaient mieux compris et traités plus rapidement en cas d'urgence.

Cependant, les messages en MCL ont eu des temps de réaction significativement plus rapides que la version plus naturelle lorsque la pression du temps et le stress étaient absents (les participants avaient trois fois plus de temps pour répondre). Cela pourrait s'expliquer par le fait que les participants n'étaient pas invités à répondre rapidement et que, dans la phase non urgente (qui constituait toujours la première phase de l'expérience), ils avaient amplement le temps de lire, éventuellement de relire et de réfléchir à leurs réponses avant de répondre. Comme le MNL avait des structures plus longues et plus de mots, et que le temps n'était pas un problème, plus de mots équivalaient à un temps de lecture plus long. Ce résultat particulier laisse à penser que la structure du langage contrôlé plus naturelle est traitée plus rapidement par le cerveau lorsque cela est nécessaire et particulièrement sous stress, mais cet effet disparaît avec plus de temps, car il n'y a pas de pression pour réagir rapidement et plus de mots équivaut à plus temps de lecture. Toutefois, s'il est vrai que les participants ont fonctionné plus rapidement dans les conditions non urgentes du MCL, ils n'en ont pas moins répondu avec plus de précision. En d'autres termes, alors que les participants mettaient beaucoup plus de temps à répondre dans les conditions MNL non urgentes, ils commettaient également moins d'erreurs que dans les conditions non urgentes du MCL.

Ces résultats sont corroborés par les analyses des erreurs et des non-réponses, car celles-ci étaient toujours significativement plus élevées en condition MCL lors des situations d'urgence. Les non-réponses (ou les délais d'inactivité) jouent un rôle important en montrant que les temps de réponses dans des tâches en MNL sont plus rapides car il y a beaucoup moins de non-réponses dans cette condition. Bien qu'il y ait plus de mots à lire, la structure de la phrase est traitée plus facilement (nettement moins de non-réponses) et correctement (nettement plus précise). Cela pourrait être dû à la réduction des interprétations possibles dans la structure de la phrase, alors que les ellipses en condition MCL rendaient le format plus difficile à décoder, à comprendre et à traiter efficacement en raison des économies linguistiques et du manque de clarté. Les multiples interprétations auraient pu amener les sujets à dépasser les limites de temps (davantage de non-réponses). Nous pouvons alors en conclure que, parce que la structure de la phrase réduit les interprétations possibles, elle permet également de localiser visuellement les informations et de répondre plus rapidement et avec plus de précision. Bien que la séparation du thème et du rhème dans les messages en MCL puisse sembler plus efficace pour la localisation des informations à l'écran, elle ne s'est pas traduite par des effets significatifs sur les performances.

La complexité des zones n'a joué aucun rôle ni aucune interaction entre les deux conditions linguistiques car le MNL était toujours plus précis dans toutes les zones de l'interface (du moins difficile au plus difficile). Contrairement aux expériences 1 et 2, lors de l'expérience 3 le phénomène a été observé de manière constante dans les conditions de difficulté (dans les expériences 1 et 2, la difficulté supplémentaire des stimuli non concordants donnait un avantage significatif aux messages en MNL, alors qu'ils étaient absents dans des conditions moins difficiles). Cependant, il convient de noter que la difficulté supplémentaire liée aux deux premières expériences et à la troisième est difficilement comparable, car la tâche demandée est totalement différente.

L'expérience 3 a également attiré notre attention sur d'autres résultats, qui consistaient à montrer comment la structure du MCL (sous forme de déclaration d'information ou d'injonction) était mieux comprise. Les résultats ont montré que la structure du MCL est significativement (donc plus fréquemment et mieux) comprise lorsqu'elle est présentée sous forme de déclaration, alors que le MNL ne présentait pas une précision sensiblement meilleure dans les conditions d'injonction ou de déclaration car les moyennes de précision étaient très proches dans les deux conditions. Cela nous montre que les messages en MNL communiquent avec précision le sens voulu, la balance ne penche pas plus d'un côté que de l'autre, alors que les messages en MCL sont nettement mieux compris sous forme de déclarations. Par conséquent, comme le format typographiquement variable ne présente aucune supériorité par rapport aux structures de langage naturel (moins précises, etc.), nous pourrions envisager de le remplacer par le format de phrase typographiquement stable, lui-même fidèle à son sens et à son intention.

Les mêmes effets ont été observés chez tous les sujets ayant différents niveaux de compréhension de l'anglais, y compris ceux dont l'anglais était la langue maternelle, car nous avions à ce stade, testé un échantillon suffisamment important. Ces résultats sont cohérents avec les résultats des deux premières expériences.

La familiarité avec le corpus original, qui n'a produit aucun effet spécifique (hormis les principaux effets significatifs pour le MNL), concorde également avec les résultats des deux premières expériences. Les experts, cependant, avaient tendance à mieux fonctionner dans des conditions de MNL qui ne sont pas les structures linguistiques avec lesquelles ils sont habitués. Ceci est remarquable car il montre que l'accoutumance à un format ne signifie pas nécessairement que l'on obtiendra les meilleurs résultats avec ce format. Cela ne signifie pas non plus que cette même population serait plus opposée à l'apprentissage d'un nouveau format plus optimal ; néanmoins, les experts se sont généralement tournés vers le format MCL lorsqu'on leur a demandé quelle était leur préférence (paradoxe performance-préférence).

Une découverte intéressante dans la recherche subjective de l'expérience 3 montre que, premièrement, le format en MCL est plus naturellement et instinctivement compris (par les sujets) en tant que déclaration et, dans une moindre mesure, comme une injonction. Au mieux, il pourrait être compris dans les deux sens, ce qui à son tour, soulève des questions (concernant son intégration dans la conception future du poste de pilotage) sur l'adéquation de ce choix de format par rapport à son intention d'injonction initiale. Dans l'ensemble, les participants ont préféré les messages MNL (55%) aux messages MCL (22,85%), tandis que 22,14% des participants n'avaient pas de préférence ou n'avaient pas remarqué de différence dans les stimuli. Les locuteurs natifs ont exprimé une préférence écrasante pour une langue contrôlée plus naturelle, alors que les opinions divergeaient davantage parmi la population non native.

En outre, moins le format original est familier, plus les sujets ont tendance à préférer le format en MNL au format en MCL (les résultats sont valables pour les 3 expériences). Cependant, les pilotes d'essai et les pilotes privés, en majorité, ont exprimé les avantages possibles de l'utilisation d'un langage contrôlé plus naturel dans les tâches expérimentales ainsi que dans leurs interactions habituelles dans le poste de pilotage. Ils ont mentionné que les messages MNL les aidaient à se renseigner sur les actions à entreprendre, à se rappeler et à mieux connaître les procédures en cours.

Le tableau comparatif 1 suivant récapitule les principaux résultats des 3 expériences. Dans la première partie (en gris), nous avons pu voir les spécificités de chaque expérience, telles que le nombre de participants, le type de données de corpus testées, etc. La seconde partie (en bleu) montre certains des résultats significatifs (surlignés en jaune) que nous avons décrits précédemment.

	Experiment 1				Experiment 2				Experiment 3			
Participants	Native		Non-Native		Native		Non-Native		Native		Non-Native	
	12		60		1		28		41		99	
Type of Corpus Data	Information/availability				Information/availability				Injunction/Statement			
Type of Task	Congruency Image/Text (DMDX)				Congruency Image/Text (DMDX)				Performance on Touchscreen HMI (ePrime 3)			
Syntactic Difficulty of Stimuli	1 → 6				1 → 6				–			
Urgency	Urgent (Time pressure)				Non-Urgent (No time pressure)				Urgent + Non-Urgent Phases			
Accuracy **General Significance**	MCL		MNL		MCL		MNL		MCL		MNL	
	No		No		No		No		*Urgent*	*Non-Urgent*	*Urgent*	*Non-Urgent*
									No	No	Yes	Yes
Reaction Time **General Significance**	MCL		MNL		MCL		MNL		MCL		MNL	
	No		Yes		No		No (avg. Yes)		*Urgent*	*Non-Urgent*	*Urgent*	*Non-Urgent*
									No	Yes	Yes	No
Reaction Time Significance **English Level**	MCL		MNL		MCL		MNL		MCL		MNL	
	Basic/ Interm.	*Native Speakers*	*Basic/ Interm.*	*Native Speakers*	*Basic/ Interm.*	*Native Speakers*	*Basic/ Interm.*	*Native Speakers*	*Basic/ Interm.*	*Native Speakers*	*Basic/ Interm.*	*Native Speakers*
	No	No	Yes	No (avg. Yes)	No	No	No	No	No	No	Yes	Yes
Reaction Time Significance **Syntactic Difficulty**	MCL		MNL		MCL		MNL		MCL		MNL	
	No		Yes		No		No		–		–	
Reaction Time Significance **Incongruence Difficulty**	MCL		MNL		MCL		MNL		MCL		MNL	
	No		Yes		No		Yes		–		–	
Reaction Time Significance **Zone Difficulty**	MCL		MNL		MCL		MNL		MCL		MNL	
	–		–		–		–		No		Yes	
Accuracy Significance **Comprehension of Format**	MCL		MNL		MCL		MNL		MCL		MNL	
	–		–		–		–		*Statement*	*Injunction*	*Statement*	*Injunction*
									Yes	No	No (equal means)	No (equal means)

Tableau 1. Tableau comparatif récapitulant les principaux résultats des 3 expériences

2.1 Application aux CNLs

2.1.1 Implication pour le Airbus Cockpit CL

Dans l'ensemble, par ces expériences, nous espérions mieux éclairer les limites de la simplification et dans quelle mesure la composante de la naturalité d'un langage contrôlé est par nature ambiguë ou paradoxalement claire. Dans cette étude de cas d'un langage contrôlé hautement codé, nous avons observé qu'aller vers ce qui pourrait être considéré comme potentiellement ambigu (proximité de la structure du langage naturel) et des éléments syntaxiques redondants, améliore en réalité la compréhension et la performance dans son ensemble et facilite le traitement de l'information. Ces expériences constituaient également une première tentative de fournir une preuve empirique des hypothèses linguistiques de simplification de texte dans des langues contrôlées, en utilisant des protocoles psycholinguistiques robustes et une technologie d'expérimentation nouvellement disponible.

La simplification excessive, dans le sens où elle s'éloigne de la naturalité du langage et de ses ambiguïtés potentielles, n'a pas conduit à de meilleurs résultats cognitifs et comportementaux pour ce langage contrôlé. Nous pouvons donc en conclure que le fait d'avoir une structure plus naturelle en utilisant des séquences syntaxiques complètes et en respectant les actes de langage adéquats conduira à une compréhension plus rapide et plus précise, puis réduira le temps d'apprentissage, car les structures se sont révélées plus intuitives pour un grand échantillon de sujets.

2.1.2 Implication pour le domaine CNL

Cependant, ces résultats ne s'appliquent pas nécessairement à toutes les langues contrôlées (ni même à toutes les langues codées, moins naturelles) car ils dépendent du domaine dans lequel ils sont utilisés, à quelles fins, pour quels actes de langage et pour quels types de données transmises. Beaucoup de langages contrôlés axés sur la compréhension utilisent déjà des structures de langage naturel ou ont un score élevé sur la dimension de naturalité du PENS. Cependant, les règles qui définissent ces langages ne sont souvent pas testées pour la compréhension de l'utilisateur et les limites de la simplification ne sont pas correctement définies. En référence à la compréhension des textes pédagogiques, Nickl (2018) écrit : « *L'un des facteurs clés permettant de mieux comprendre le fonctionnement de la compréhension et de la compréhensibilité a été la réorientation des phénomènes de surface dans les textes vers une enquête* *sur le processus d'interprétation du lecteur / auditeur. En milieu universitaire, ce changement de cap résultait des théories cognitives, qui ont prouvé que l'interprétation de textes n'était pas un processus passif de décodage de l'information, mais plutôt un processus actif de construction du sens. Il est donc évident que toute personne souhaitant améliorer la compréhensibilité textuelle doit prendre en compte les lecteurs potentiels du texte. D'un simple aspect d'un texte, la compréhensibilité a été transformée en une description de la relation entre un texte donné et le public cible de ce texte. Cela implique que les textes ne peuvent plus être qualifiés de compréhensibles en eux-mêmes. »*

Par conséquent, la réponse à la question de l'ampleur de la simplification réside non seulement dans l'utilisation et les objectifs d'un langage contrôlé, mais aussi dans l'évaluation systématique de ses règles à l'aide de techniques adéquates de mesure des aptitudes cognitives et des réactions comportementales. Car ce n'est qu'en ayant la preuve empirique de l'efficacité d'un langage contrôlé que nous pourrons être certains de l'adéquation des règles qui le régissent. Ryan (2018) écrit, en se référant à une vue d'ensemble du domaine : « *Cependant, il reste un certain nombre de problèmes concernant l'efficacité des langues contrôlées utilisées et l'évaluation quantitative des avantages pratiques qu'elles procurent* ».

2.1.3 Recommandations pour le Airbus Cockpit CL

À partir des résultats de nos évaluations, nous avons formulé des recommandations concernant l'évolution du langage contrôlé d'Airbus pour les futurs cockpits, en particulier (mais pas uniquement) pour les types de données informationnelles et injonctives. Ces recommandations découlent directement des résultats de la recherche objective et subjective, ainsi que d'une recommandation d'évaluations dans la littérature (si une recommandation est basée sur une recherche subjective ou dans la littérature, elle est mentionnée dans la règle proposée).

Cette partie a été partiellement supprimée de la version finale de la thèse pour des raisons de confidentialité. La version complète sera archivée avec d'autres données de thèse pour l'utilisation interne d'Airbus[6].

6. Il était important de garder cette thèse non confidentielle dans son ensemble afin de proposer des évaluations et des résultats pouvant être utiles à la communauté scientifique, faute de quoi les évaluations font parfois défaut dans le domaine.

3. Pistes à explorer

Cette thèse de doctorat s'inscrit dans le cadre d'un projet plus vaste chez Airbus visant à repenser le futur cockpit disruptif. Comme son nom l'indique, ce projet propose des approches et des solutions innovantes pour réorganiser entièrement l'environnement de travail du pilote, notamment en proposant davantage d'automatisation et des conceptions plus intuitives / guidées. Ainsi, le nouveau langage plus naturel est une brique dans le mur du concept de conception de cockpit disruptif : une interface plus intuitive (écrans plus grands, technologie plus récente, etc.) va de pair avec un langage plus approprié : des structures moins codées sont plus faciles à utiliser, apprendre et mémoriser. De plus, elles exigent moins de formation de la part du pilote. Comme le système pourrait être plus automatisé, il est de plus en plus nécessaire d'informer le pilote de l'état de l'appareil et de le tenir au courant. Des structures de langage plus naturelles sont plus permissives (pas aussi restrictives que le langage codé) et pourraient être un outil utile pour aider à la prise de conscience de la situation et à la prise de décision. Riley et al. (1999) ont écrit qu' *«à mesure que l'automatisation devient de plus en plus sophistiquée et complexe, le temps et le coût de la formation des pilotes à son utilisation et la possibilité d'erreur de pilotage suscitent de plus en plus d'inquiétudes. [...] L'un des facteurs les plus importants de la convivialité est la mesure dans laquelle la fonctionnalité sous-jacente et le fonctionnement logique d'un périphérique ou d'un système sont compatibles avec le modèle mental de l'utilisateur. [...] Nous voulons que le système fonctionne comme pense le pilote, de sorte que nous n'avons pas à le former pour qu'il pense comme le système fonctionne. Étant donné que les actions et les objectifs se comportent tous selon des règles syntaxiques et sémantiques, le pilote sait déjà que, sachant comment se conformer aux autorisations des instruments, aucune formation supplémentaire n'est nécessaire pour apprendre une nouvelle fonction ou procédure.»*

Afin d'intégrer un nouveau langage plus naturel dans les futurs cockpits, il reste encore beaucoup à faire. Nous pourrions par exemple utiliser des méthodes physiologiques pour tester les effets du stress sur le traitement de l'information dans différentes conditions linguistiques telles que les moniteurs de fréquence cardiaque ou de pression artérielle. L'oculométrie pourrait être utilisée pour enregistrer les mouvements oculaires des participants et déterminer quel langage nécessite davantage d'être observé pour obtenir une compréhension efficace et une réalisation optimale de la tâche.

4. Conclusion

Cette recherche de doctorat est née d'un besoin industriel, qui est d'optimiser l'utilisation du langage dans les futurs cockpits Airbus dans le but de maintenir la sécurité et l'optimisation de la performance des pilotes et de limiter les besoins en matière de formation. Pour répondre à ce besoin, nous avons d'abord dû nous intéresser aux langages contrôlés existants chez Airbus et au contexte dans lequel ils sont utilisés, aux limites et aux perspectives de l'optimisation.

De plus, nous avons étudié, d'un point de vue académique, le domaine des langages contrôlés et de toutes les recherches qui y sont associées, comment et où ces derniers sont utilisés, dans quel but et atteignent-ils bel et bien leurs objectifs en réduisant les ambiguïtés dans un cadre industriel avec des utilisateurs finaux humains. Nous avons découvert que, de manière surprenante, très peu d'études empiriques ont été réalisées sur la compréhension humaine en ce qui concerne les CNLs établis.

Axés sur la compréhension et ayant tendance à être conçus pour répondre aux besoins industriels et aux spécifications utilisées par un public cible désigné (même si ces spécifications risquent de poser problème en dehors du domaine), les CNLs reposent largement sur des guides de rédaction généraux, des directives de recherche linguistique et des règles de simplification recyclées, qui sont rarement testées scientifiquement sur le traitement cognitif humain. Les règles des CNLs qui sont souvent arbitraires (nombre maximal de mots dans une phrase, utilisation de mots de fonction, utilisation de la voix passive, utilisation de modaux et de pronoms, etc.) sont donc mises en place sans que cela ne permette de rendre la compréhension plus facile pour l'utilisateur final. Un grand nombre de langages contrôlés destinés à un usage industriel sont si bien gardés (pour des raisons de responsabilité judiciaire et de propriété intellectuelle) qu'il est difficile de progresser efficacement dans le domaine, ou de mener des recherches sur les différentes règles qui existent.

De plus, et bien que la simplification semble nécessaire du fait que les langages non contrôlés sont dangereusement ambigus dans les domaines critiques pour la sécurité, les règles établies ne garantissent pas une meilleure compréhension humaine (meilleure que le langage naturel ou meilleure que d'autres règles de simplification), car les limites sont floues. Jusqu'à quel point la simplification est-elle nécessaire pour parvenir à une meilleure compréhension, existe-t-il un seuil au-delà duquel une simplification excessive entraîne plus d'ambiguïté ? La réponse n'était pas claire.

Par conséquent, notre besoin industriel initial a évolué vers une recherche qui a davantage consisté à questionner les enjeux linguistique, ergonomique, et psycholinguistique. Le langage naturel ou les INLs étant par nature ambigus, ils nous ont laissé entendre à tort que plus nous contrôlons ou codifions un langage et l'éloignons de ses éléments naturels générateurs d'ambiguïté, plus nous disposons alors d'un moyen sûr d'éradiquer les abus ; pour se faire, la façon la plus évidente est de renforcer la formation et l'enseignement des règles du langage simplifié.

Afin de pouvoir trouver plus de réponses concernant les limites de la simplification, nous avons examiné les travaux de recherche qui ont tenté de trouver la preuve que certains langues contrôlées établis apportaient une valeur ajoutée à la compréhension humaine. Les résultats ont été quelque peu non concluants, car des résultats significatifs fiables n'ont pas été mis en avant en faveur d'un langage plus simplifié par rapport à un langage plus naturel, moins simplifié.

De plus, comme ces évaluations constituaient une première approche pour la recherche de preuves empiriques dans les années 90s, les méthodes d'expérimentation comportementale utilisées n'étaient pas à la hauteur des normes des techniques de recherche cognitive d'aujourd'hui. Des questionnaires sur la compréhension en lecture ont été utilisés pour tester la compréhension des textes directives (au lieu de la performance, car ces textes procéduraux sont censés être compris et exécutés). Les temps de réponse n'étaient ni limités ni contrôlés.

Par conséquent, notre objectif est de proposer une première approche pour évaluer les langages contrôlés à l'aide de protocoles comportementaux psycholinguistiques étroitement contrôlés (conditions laboratoires) afin de tester les limites de la simplification. La principale exigence étant d'optimiser le langage contrôlé actuel en tenant compte de la conception future du cockpit et de ses interfaces, nous nous sommes basés sur ce langage hautement codifié (renforcé par la formation et l'apprentissage) et théoriquement moins complexe (pas d'ambiguïtés dues au langage naturel).

Ce langage contient plusieurs catégories de données destinées à guider, avertir et aider les pilotes à piloter l'avion. Nous les avons identifiées et avons basé notre expérimentation sur deux principales catégories : celle de type informationnel (informer les pilotes des occurrences, des disponibilités, etc.) et celle de type action (demander aux pilotes d'exécuter une action sur un bouton ou un levier, un écran, etc.). Nous avons proposé des équivalents de structure de langage plus naturels aux structures codées existantes en ajoutant des éléments syntaxiques et phrastiques manquants et en respectant les actes de langage adéquats pour chaque catégorie d'informations dans la syntaxe de l'INL de l'anglais classique (par exemple, en commençant par le verbe pour injonctions).

Les résultats montrent que la simplification excessive n'a pas conduit à une meilleure compréhension et que la version plus naturelle a donné de meilleurs résultats et est mieux comprise de manière intuitive (par conséquent, elle nécessitera moins de formation pour la maîtrise).

Alors que le langage plus codé existant (comprenant plusieurs limitations techniques dues à la taille de l'écran, etc.) a jusqu'ici limité l'ambiguïté, il doit être complété par une formation plus poussée des utilisateurs. L'utilisation d'un langage plus naturel pour les besoins de l'Airbus CL dans un futur cockpit repensé est plus bénéfique pour la compréhension et les performances humaines.

Les autres langages contrôlés ont des niveaux de naturalité et de simplification différents, qui dépendent de leur utilisation, de leur domaine, de leur public cible et de leurs objectifs généraux. Cependant, les règles qui les régissent doivent faire l'objet d'une évaluation psycholinguistique afin de s'assurer que les règles prescriptives et proscriptives sont aussi efficaces que possible et qu'elles réduisent réellement les ambiguïtés et améliorent la compréhension et les performances humaines. Enfin, la simplification et les économies linguistiques ne signifient pas toujours ou automatiquement une amélioration de la compréhension ou ne conduisent pas à des canaux d'information plus appropriés.

Pour conclure, nous avons montré par une méthodologie scientifique cognitive que le langage contrôlé Airbus pourrait tirer parti de structures de langage plus naturelles pour améliorer la compréhension du pilote et réduire les temps d'entraînement. Ce nouveau langage plus optimisé s'intègre efficacement au futur concept de cockpit disruptif et à ses conceptions plus intuitives. Nous avons également montré qu'il existait un manque notable d'évaluations linguistiques contrôlées sur le terrain, ainsi que de méthodes adéquates d'évaluation d'hypothèses linguistiques utilisant une méthodologie robuste des sciences cognitives pour répondre aux besoins ergonomiques. Nous proposons qu'à l'avenir, les règles de langage contrôlées soient systématiquement évaluées afin de démontrer leur efficacité avant d'être appliquées, en particulier dans les domaines critiques pour la sécurité.

Nous espérons également que cette thèse fournira un aperçu / une motivation pour utiliser une méthodologie cognitive et des données comportementales afin de tester des hypothèses linguistiques descriptives, même au-delà du domaine des langages contrôlés. La linguistique et la méthodologie cognitivo-comportementale devraient être associées de manière plus efficace pour tirer parti des résultats vérifiables et comparables sur le plan scientifique, avec des hypothèses basées sur des décennies de théorie linguistique.

References

Adriaens, G., & Schreurs, D. (1992). From COGRAM to ALCOGRAM: Toward a Controlled English Grammar Checker. In *Proceedings of the 14th conference on Computational linguistics-Volume 2* (pp. 595-601).

Aikawa, T., Schwartz, L., King, R., Corston-Oliver, M., & Lozano, C. (2007). Impact of Controlled Language on Translation Quality and Post-Editing in a Statistical Machine Translation Environment. *Proceedings of the MT Summit XI*, 1-7.

Audley, R. J., & Wallis, C. P. (1965). Cross-Over Effect: A Reply To Arthur. *British Journal of Psychology, 56*(4) (pp. 479-482).

Austin, J. L. (1975). *How To Do Things With Words* (Vol. 88). Oxford University Press.

Barcenilla, J., & Brangier, E. (2000). Propositions pour une intervention en ergonomie des aides textuelles au travail. *Illettrisme et milieu de travail. Paris: La documentation française* (pp. 357-368).

Bernstein, A., & Kaufmann, E. (2006, November). GINO–A Guided Input Natural Language Ontology Editor. In *International Semantic Web Conference* (pp. 144-157).

Bhatia,V.K. (1993). *Analysing Genre*. London, Longman.

Biber, D. (1988). *Variation Across Speech and Writing*. Cambridge: Cambridge University Press.

Bisseret, A. (1983). Physiology for Man Computer Cooperation in Knowledge Processing. In *IFIP Congress* (pp. 113-120).

Bock, K., & Levelt, W. (2002). Language Production. *Psycholinguistics: Critical concepts in psychology, 5*, 405.

Caillies, S., & Sourn–Bissaoui, L. (2013). *Nondecomposable idiom understanding in children: Recursive theory of mind and working memory. Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale, 67*(2), 108.

Cardon, R., & Grabar, N. (2018). Identification of Parallel Sentences in Comparable Monolingual Corpora from Different Registers. In *Proceedings of the Ninth International Workshop on Health Text Mining and Information Analysis* (pp. 83-93).

Carroll, D. (2008). *Psychology of Language*. (Fifth Edition). Thomson Wadsworth.

Chervak, S., Drury, C. and Ouellette, J. (1996). Simplified English for Aircraft Workcards. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 40(5) (pp. 303-307)

Chervak, S. G. (1996). *The Effects of Simplified English on Performance of a Maintenance Procedure*. Doctoral dissertation, State University of New York at Buffalo.

Coltheart, M., Rastle, K., Perry, C., Langdon, R., & Ziegler, J. (2001). DRC: A Dual Route Cascaded Model of Visual Word Recognition and Reading Aloud. *Psychological review, 108*(1), 204.

Condamines, A. (2008). Peut-on prévenir le risque langagier dans la communication écrite? *Langage et société*, (3) (pp. 77-97).

Condamines, A. (2010). Variations in Terminology: Application to the Management of Risks Related to language Use in the Workplace. *Terminology. International Journal of Theoretical and Applied Issues in Specialized Communication, 16*(1) (pp. 30-50).

Condamines, A. (2018). Pour le développement d'une linguistique ergonomique: l'exemple des langues contrôlées. *Le travail humain.*

Condamines, A., Lopez, S., & Warnier, M. (2017). The Linguist and Workplace Norms. *Cahiers de lexicologie*, 2017 (pp. 171-190).

Condamines, A., Narcy-Combes, J.-P. (2015). La linguistique appliquée comme science située, in *Cultures de recherche en linguistique appliquée*, édité par Francis CARTON et al., Paris : Riveneuve éditions (pp. 209-229).

Condamines, A., & Warnier, M. (2014). Linguistic Analysis of Requirements of a Space Project and Their Con-formity with the Recommendations Proposed by a Con-trolled Natural Language. In *International Workshop on Controlled Natural Language* (pp. 33-43). Springer International Publishing.

Condamines, A., & Warnier, M. (2017). Towards the Creation of a CNL Adapted to Requirements Writing by Combining Writing Recommendations and Spontaneous Regularities: Example in a Space Project. In *Language Resources and Evaluation, 51*(1) (pp. 221-247).

Connatser, B. R. (1999). Last rites for readability formulas in technical communication. *Journal of technical writing and communication, 29*(3) (pp. 271-287).

Crystal, D., & Davy, D. (1969). The Language of Legal Documents. *Investigating English Style* (pp. 193-217).

Dashiell, J. F. (1937). Affective value-distances as a determinant of esthetic judgment-times. *The American Journal of Psychology*, 50(1/4) (pp. 57-67).

Delacroix, H. (1924). *Le langage et la pensée*, Paris: Alcan.

Depecker, L. (1996). Terminologie et standardisation. In *Hieronymus Complutensis* 3, (pp. 85–88).

DuBay, W. H. (2004). The Principles of Readability. Impact Information: Costa Mesa, CA.

Eckert, D. (1997). The Use of Simplified English to Improve Task Comprehension for Non-Native English Speaking Aviation Maintenance Technician Students.

Esenova, K. (2017). *Pragmalinguistic Studies in Linguistics*: Monograph. Éditions du JIPTO.

Field, J. (2003). *Psycholinguistics: A resource book for students*. Psychology Press.

Fischler, I. (1977). Semantic Facilitation Without Association in a Lexical Decision Task. *Memory & cognition, 5*(3) (pp. 335-339).

Flesch, R. (1944). How Basic is Basic English?. *Harper's Magazine*, *188*(1126) (pp. 339-343).

Flesch, R. (1979). How to Write Plain English: Let's Start with the Formula. *University of Canterbury*.

Fodor, J. A. (1975). The Language of Thought (Vol. 5). Harvard University Press.

Fuchs, N. E., & Schwitter, R. (1995). Specifying Logic Programs in Controlled Natural Language. In *Proceedings CLNLP 95, ELSNET/COMPULOG-NET/EAGLES Workshop on Computational Logic for Natural Language Processing* Edinburgh: University of Edinburgh.

Fuchs, N. E., & Schwitter, R. (1996). Attempto Controlled English (ace). *arXiv preprint cmp-lg/9603003*.

Garnham, A. (1985). *Psycholinguistics: Central Topics*. Methuen: London.

Ganier, F. (2002). De l'analyse des fonctionnements cognitifs à l'ergonomie des aides opératoires. *Communication au séminaire Phiteco-Costech, Technologies cognitives: technologies numériques, systèmes d'aides et normes, Université de Technologie de Compiègne*, 24.

Gaudin, F. (1993). Pour une socioterminologie: Des problèmes pratiques aux pratiques institutionnelles. *Rouen: Publications de l'Université de Rouen*.

Gaudin, F. (2003). Terminologie et politique linguistique. In *Socioterminologie. Une approche sociolinguistique de la terminologie*. Louvain-la-Neuve : De Boeck Supérieur.

Gaudin, F. (2005). La socioterminologie. *Langages*, (1) (pp. 81-93).

Gavieiro, E. (2001). *Vers un modèle d'élaboration de la terminologie d'une langue contrôlée ; application aux textes d'alarmes en aéronautique pour les futurs postes de pilotage*. Ph.D. thesis, University of Franche Comté, Besançon.

Gernsbacher, M. A. (2013). *Language Comprehension as Structure Building*. Psychology Press.

Giraudo, H., & Hathout, N. (2012). Does The Size of the Series Affect the Visual Recognition of Suffixed Words?. In *Exo-lexical variables in monolingual and bilingual morphological processing*.

Giraudo, H., & Voga, M. (2014). Measuring Morphology: the Tip of The Iceberg? A Retrospective on 10 Years of Morphological Processing. *Cahiers de Grammaire*, 000.

Gledhill, C. J. (2000). *Collocations in science writing* (Vol. 22). Gunter Narr Verlag.

Goyvaerts, P. (1996). Controlled English, Curse or Blessing?–A User's Perspective. In *Proceedings of the 1st International workshop on Controlled Language Applications (CLAW'96)(Leuven)* (pp. 137-42).

Halliday, M. A. (1967). Notes on Transitivity and Theme in English: Part 2. *Journal of linguistics*, 3(2) (pp. 199-244).

Harley, T. A. (2013). *The Psychology of Language: From Data to Theory*. Psychology Press.

Harris, Z. (1968). Mathematical structures of language. *Interscience tracts in pure and applied mathematics*.

Hart, G., Johnson, M., & Dolbear, C. (2008, June). Rabbit: Developing a Control Natural Language for Authoring Ontologies. In *European Semantic Web Conference* (pp. 348-360).

Hart, S. G., & Simpson, C. A. (1976). Effects of Linguistic Redundancy on Synthesized Cockpit Warning Message Comprehension and Concurrent Time Estimation. In *Proceedings of the 12th Annual Conference on Manual Control* (Vol. 170, pp. 309-321).

Heurley, L. (1997). Vers une définition du concept de texte procédural : le point de vue de la psycholinguistique. *Les Cahiers du français contemporain*, 4 (pp. 109-133).

Hinson, D. E. (1988). Simplified English—Is it really simple?. In *Proceedings of the 38th International Technical Communication Conference*.

Holmback, H., Shubert, S., & Spyridakis, J. (1996). Issues in conducting empirical evaluations of controlled languages. In *Proceedings of the first international workshop on Controlled Language applications* (Vol. 166, p. 177).

Huijsen, W.-O. (1998). Controlled Language - An Introduction. In *Proceedings of the Second International Workshop on Controlled Language Applications - CLAW 98*. Sous la dir. de T. Mitamura. Pittsburgh, Pennsylvania: Language Technologies Institute, Carnegie Mellon University (pp. 1-15).

Humbley, J. (2001). Quelques enjeux de la dénomination en terminologie. In *Cahiers de praxématique*, (36) (pp. 93-115).

Jahchan, N., Condamines, A., & Cannesson, E. (2016, July). To What Extent Does Text Simplification Entail a More Optimized Comprehension in Human-Oriented CNLs?. In *International Workshop on Controlled Natural Language* (pp. 69-80). Springer International Publishing.

Kantor, J. R. (1936). *An Objective Psychology of Grammar.* Bloomington: The Principia Press.

Kelvin, W. T. (1883). *Electrical Units of Measurement.* Institution of Civil Engineers.

Kern, R. P. (1985). Modeling users and their use of technical manuals. In *Designing usable texts* (pp. 341-375).

Kittredge, R. I. (2003). Sublanguages and controlled languages. In Ruslan Mitkov, editor, *The Oxford Handbook of Computational Linguistics* (pp. 430-447).

Kuhn, T. (2010). *An Evaluation Framework for Controlled Natural Languages.* In Norbert E. Fuchs, editor, *Proceedings of the Workshop on Controlled Natural Language (CNL 2009)*, Lecture Notes in Computer Science 5972 (pp. 1-20). Springer, 2010.

Kuhn, T. (2014). A Survey and Classification of Controlled Natural Languages. *Computational Linguistics, 40*(1) (pp. 121-170).

Lee, D. (2002). Genres, Registers, Text Types, Domains and Styles: Clarifying the Concepts and Navigating a Path Through the Bnc Jungle. In *Language and Computers 42* (pp. 247-292).

Leroy, G., Helmreich, S., and Cowie, J. R. (2010). The Effects of Linguistic Features and Evaluation Perspective on Perceived Difficulty of Medical Text. In *43rd Hawaii International Conference on System Sciences*, Hawaii.

Lim, Y., Bassien-Capsa, V., Ramasamy, S., Liu, J., & Sabatini, R. (2017). Commercial Airline Single-Pilot Operations: System Design and Pathways to Certification. *IEEE Aerospace and Electronic Systems Magazine, 32*(7) (pp. 4-21).

Lopez, S. (2013). Norme (s) et usage (s) langagiers : le cas des communications pilote-contrôleur en anglais, Doctoral dissertation, Université Toulouse le Mirail-Toulouse II.

Mathews, E. (2004). New Provisions for English Language Proficiency Are Expected to Improve Aviation Safety. *ICAO Journal*, 59 (pp. 4-6).

Mc Laughlin, G. H. (1969). SMOG grading-a new readability formula. *Journal of reading*, 12(8) (pp. 639-646).

McNamara, D., Louwerse, M., McCarthy, P., Graesser, C., (2010). Coh-Metrix: Capturing Linguistic Features of Cohesion, *Discourse Processes*, 47:4 (pp. 292-330).

Mifflin, H (2000). The American Heritage Dictionary of the English Language, fourth edition.

National Information Standards Organization. (2005). Guidelines for the Construction, Format, and Management of Monolingual Controlled Vocabularies (ANSI/NISO Z39. 19-2005).

Nickl, M. (2018). 17 Instructional Texts–Learn How to Get Things Done. *Languages for Special Purposes: An International Handbook*, 321.

Nielsen, J. (1993). *Usability Engineering*. Academic Press, Boston.

Nielsen, J., & Levy, J. (1994). Measuring Usability: Preference vs. Performance. *Communications of the ACM, 37*(4) (pp. 66-75).

O'Brien, S. (2003). Controlling Controlled English: An Analysis of Several Controlled Language Rule Sets. In *Proceedings of EAMT-CLAW*, 3 (pp. 105-114), 33.

O'Brien, S. (2006). *Machine-translatability and post-editing effort: An empirical study using Translog and Choice Network Analysis.* Doctoral dissertation, Dublin City University.

O'Brien, S., & Roturier, J. (2007). How Portable Are Controlled Language Rules? A comparison of Two Empirical MT Studies. In *Proceedings of MT summit XI* (pp. 345-352).

O'Grady, W., Archibald, J., Aronoff, M., & Rees-Miller, J. (2001). *Contemporary Linguistics: An Introduction* (4th ed.). Boston: Bedford/St. Martin's.

Paterson, D. G., & Tinker, M. A. (1946). Readability of newspaper headlines printed in capitals and in lower case. *Journal of Applied Psychology*, 30(2), 161.

Pearson, J. (1998). Terms in context (Vol. 1). John Benjamins Publishing.

PLAIN. (2011). *Federal Plain Language Guidelines*. plainlanguage.gov.

Pollio, H. R. (1964). Some semantic relations among word-associates. *The American journal of psychology, 77*(2) (pp. 249-256).

Pool, J. (2006). Can Controlled Languages Scale to the Web?. In *International Workshop on Controlled Language Applications* 5.

Pratt-Hartmann, I. (2009). Computational Complexity of Controlled Natural Languages. In *Pre-Proceedings of CNL 2009, 5*.

Pratt-Hartmann, I., & Third, A. (2006). More Fragments of Language. *Notre Dame Journal of Formal Logic, 47*(2) (pp. 151-177).

Pulvermüller, F. (2009). Word Processing in the Brain as Revealed by Neurophysiological Imaging. *The Oxford Handbook of Psycholinguistics* (pp. 119-139).

Rankin, W. (2007). *MEDA Investigation Process. Boeing Commercial Aero, 2nd Quarter, 2007*. [online] Available at: http://www.seabasedradar.com/commercial/aeromagazine/articles/qtr_2_07/AERO_Q207_article3.pdf [Accessed 27 Nov. 2018].

Rayner, K. (1978). Eye Movements in Reading and Information Processing. *Psychological bulletin, 85*(3), 618.

Riley, V., DeMers, B., Misiak, C., & Schmalz, B. (1999). The Cockpit Control Language: A Pilot-Centered Avionics Interface. *Air & Space Europe*, 1(1) (pp. 48-50).

Ryan, R., (2018). Controlled Language in G. Budin, J. Humbley & C. Lauren (eds) *Languages for Special Purposes – An International Handbook*, Mouton de Gruyter (pp. 289-306).

Rychtyckyj, N. (2002, October). An Assessment of Machine Translation for Vehicle Assembly Process Planning at Ford Motor Company. In *Conference of the Association for Machine Translation in the Americas* (pp. 207-215).

Rychtyckyj, N. (2005). Ergonomics Analysis for Vehicle Assembly Using Artificial Intelligence. *AI Magazine, 26*(3), 41.

Saussure, F. D. (1995). Cours de linguistique générale (T. de Mauro, Édition critique). *Paris: Payot & Rivages.(Trabalho original publicado em 1916).*

Schade, J., & Schlag, B. (2003). Acceptability of Urban Transport Pricing Strategies. *Transportation Research Part F: Traffic Psychology and Behaviour, 6*(1), 45-61.

Schwitter, R. (2002). English as a Formal Specification Language. In *Database and Expert Systems Applications, 2002. Proceedings. 13th International Workshop* (pp. 228-232). IEEE.

Schwitter, R. (2010). Controlled Natural Languages for Knowledge Representation. In *Proceedings of the 23rd International Conference on Computational Linguistics: Posters* (pp. 1113-1121). Association for Computational Linguistics.

Searle, J. R. (1969). *Speech Acts: An Essay in the Philosophy of Language* (Vol. 626). Cambridge University Press.

Shipley, W. C., Coffin, J. I., & Hadsell, K. C. (1945). Affective distance and other factors determining reaction time in judgments of color preference. *Journal of Experimental Psychology, 35*(3), pp. 206-215.

Shubert, S. K., Spyridakis, J. H., Holmback, H. K., & Coney, M. B. (1995). The Comprehensibility of Simplified English in Procedures. *Journal of technical writing and communication, 25*(4), (pp. 347-369).

Siddharthan, A. (2003). *Syntactic simplification and Text Cohesion*. Ph.D. thesis, University of Cambridge, UK.

Simpson, C. A. (1976). Effects of Linguistic Redundancy on pilot's Comprehension of Synthesized Speech. In *Twelfth Annual Conference on Manual Control NASA TM x-73* (Vol. 170, pp. 294-308).

Slobin, D. (1985). Crosslinguistic Evidence for the Language-Making Capacity. *The crosslinguistic study of language acquisition: Volume 1 & 2.* Hillsdale: Lawrence Erlbaum Associates.

Somers, H. (1998). An Attempt to Use Weighted Cusums to Identify Sublanguages. In *Proceedings of the Joint Conferences on New Methods in Language Processing and Computational Natural Language Learning* (pp. 131–139). Stroudsburg, PA, USA: Association for Computational Linguistics.

Spaggiari L. (2002). *Formalisation des rapports syntaxiques d'une langue contrôlée : application aux messages d'alerte dans les postes de pilotage avion*. Ph.D. thesis, University of Poitiers.

Spaggiari, L., Beaujard, F., & Cannesson, E. (2003). A Controlled Language at Airbus. Proceedings of *EAMT-CLAW03* (pp. 151-159).

Stewart, K. M. (1998). *Effect of AECMA Simplified English On the Comprehension of Aircraft Maintenance Procedures By Non-native English Speakers*. Master's thesis, University of British Columbia.

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of experimental psychology, 18*(6) (pp. 643-662).

Szmrecsányi, B. 2004. On operationalizing Syntactic Complexity. In *Le poids des mots. Proceedings of the 7th International Conference on Textual Data Statistical Analysis. Louvain-la-Neuve* Vol. 2 (pp. 1032-1039).

Tanguy, L., & Tulechki, N. (2009). Sentence Complexity in French: A Corpus-based Approach. In *Intelligent information systems (IIS)* (pp. 131-145).

Temnikova, I., (2012). *Text Complexity and Text Simplification in the Crisis Management Domain*. Ph.D. thesis, University of Wolverhampton

Trauzettel-Klosinski, S., & Dietz, K. (2012). Standardized Assessment of Reading Performance: the New International Reading Speed Texts IReST. *Investigative ophthalmology & visual science, 53*(9) (pp. 5452-5461).

Tricot, A. (2001). Interpréter les liens entre utilisabilité et utilité des documents électroniques. *Les documents électroniques, méthodes, démarches et techniques cognitives. Paris: Europia*.

Valette, M. (2006). *Linguistiques énonciatives et cognitives françaises. Gustave Guillaume, Bernard Pottier, Maurice Toussaint, Antoine Culioli* (p. 320). Honoré Champion.

Van Oosten, P., Tanghe, D., & Hoste, V. (2010). Towards an Improved Methodology for Automated Readability Prediction. In *7th Conference on International Language Resources and Evaluation (LREC 2010)* (pp. 775-782). European Language Resources Association (ELRA).

Warnier, M. (2018). *Contribution de la linguistique de corpus à la constitution de langues contrôlées pour la rédaction technique : l'exemple des exigences de projets spatiaux*. Ph.D. thesis, University of Toulouse - Jean Jaurès.

Wogalter, M. S., Godfrey, S. S., Fontenelle, G. A., Desaulniers, D. R., Rothstein, P. R., & Laughery, K. R. (1987). Effectiveness of warnings. *Human Factors, 29*(5) (pp. 599-612).

Wyner, A., Angelov, K., Barzdins, G., Damljanovic, D., Davis, B., Fuchs, N., ... & Luts, M. (2009). On Controlled Natural Languages: Properties and Prospects. In *International Workshop on Controlled Natural Language* (pp. 281-289).

Ziefle, M. (1998). Effects of Display Resolution on Visual Performance. *Human factors, 40*(4) (pp. 554-568).

Author Index