Towards a More Natural Controlled Language in Future Airbus Cockpits. A Psycho-linguistic Evaluation

Nataly Jahchan[1], Anne Condamines[2], Emmanuelle Cannesson[3], and Hélène Giraudo[4]

[1,3]Airbus Operation SAS
[1,2,4]CLLE-CNRS
{nataly.jahchan,emmanuelle.cannesson}@airbus.com
{anne.condamines,helene.giraudo}@univ-tlse2.fr

Abstract

The main goal of this research is to optimize an existing Airbus Cockpit Controlled Language in order to integrate it in future cockpit design. The current controlled language used aboard Airbus cockpit interfaces was carefully constructed to avoid ambiguity and complexity. In order to optimize the existing language, we set out to evaluate the appropriate levels of simplification that would achieve more accurate and faster comprehension with optimized pilot training time by using psycho-linguistic experimentation and cognitive science tools. We present in this paper a congruency task similar to traditional judgment tasks in behavioral experiments. It provides a firmly controlled environment to test linguistic hypotheses and CNL rules. Results show 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 from a psycho-linguistic point of view.

1 Introduction

The main goal of this research is to optimize an existing Airbus Cockpit Controlled Language in order to integrate it in future cockpit design. The current controlled language used aboard Airbus cockpit interfaces was carefully constructed to avoid ambiguity and complexity (as are all comprehension oriented controlled languages, (Kuhn, 2014; Schwitter, 2010; Kitteridge, 2003) 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. The need for clear and unambiguous communication is vital in safety critical domains. This controlled language and the rules that make it were put in place at a time when design flexibility was limited (for example small screen sizes that restrict word and sentence length (Spaggiari et al., 2003; Jahchan et al., 2016; Jahchan, 2017). This results in a CNL which is non-conforming to natural language syntax, highly abbreviated, typographically variable, and color-coded (Jahchan, 2019). As we are addressing a more flexible disruptive cockpit design for future aircraft, these limitations are no longer immutable constraints, and the future controlled language need not be so coded and compact, or follow very strict simplification rules.

The goal being to take into consideration the disruptive cockpit design (possibly larger screen sizes (less character limitations), newer technology, etc.) which goes hand in hand with an adapted human-oriented controlled language and which is safe, suitable and easily accessible for a human operator.

Therefore, in order to optimize the existing language, we set out to evaluate the appropriate levels of simplification that would achieve more accurate and faster comprehension with optimized pilot training time by using psycho-linguistic experimentation and cognitive science tools. In order to determine the appropriate levels of simplification, one must carefully investigate the problem in context (operational piloting constraints, cockpit design constraints, linguistic ambiguities (syntactic, semantic, and terminological ones). In this sense, we are more particularly dealing with Ergonomic Linguistics (Condamines, 2021) in which linguistic models, theories, and hypotheses are used in specified work contexts (mainly in industry) to achieve precise goals efficiently and serve a real life operational purpose (one of the primary uses of Human-oriented CNLs). These hypotheses and propositions are derived from real language productions and theoretical linguistic theories (for example common CNL construction rules among several languages (O'Brien's, 2003) and should be evaluated using experimental techniques and acceptability tests to acquire empirical evidence to support their efficiency when it comes to comprehension and optimal performance for human operators (target users of CNLs). This concept is closely related to readability and usability. Our own definition of readability for the purposes of this research does not involve the traditional definition, i.e. ease of reading, reading proficiency, or the characteristics that make readers willing to carry on reading (Flesch Kincaid, Smog formula, (Flesch 1979), etc.). 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).

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 and Warnier, 2014; Warnier, 2018) or for text complexity (Tanguy and Tulechki, 2009), and machine translation (O'Brien and Roturier, 2007; Aikawa et al., 2007), or for syntactic transformations and corpus alignment of specialized corpora with existing simplified corpora (Cardon and 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. Consequently, existing CNL research falls short on providing empirical proof 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.

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. Readability, text simplification, and text complexity research have focused on simplifying the language by making it less and less like natural language, and more like an unambiguous set of codes and regulations so that the resulting language veered away from the "natural" dimension. But to what extent is text simplification satisfactory and what are the limits at which it becomes counter-productive? When must natural language structures be respected? We constructed a more natural controlled language (MNL) by basing ourselves on the existing more codified controlled language (MCL) and its operational needs, syntactic and terminological rules) by using research that has been done on readability and text complexity 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 sentence structure we would be limiting the different possible interpretations, therefore avoiding, as much as possible, elliptical ambiguities (C.f Figure 1)

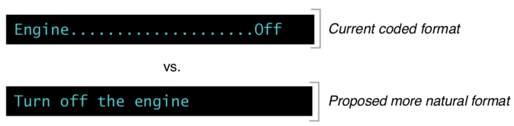

Figure 1, Example of MCL and MNL

Although pilots are trained to understand the meaning of the typographical ellipses (dots separating "engine" and "off" and color coding to mean an action that must be performed, the sentence structure (in the proposed more natural format) provides a fail-safe way of avoiding ambiguity. The sentence "Turn off the engine" adds two more words to the original statement "engine.....off " yet completely eliminates the second possible interpretation (the engine is off). Thus, information is solely contained in the linguistic elements, excluding color and typographical separation. There is only one possible way of interpreting and understanding the second sentence. In this way, we based ourselves on the MCL corpus (operational use and context, goal) and created new more natural structures (MNL) to be evaluated.

2 Method

As a first approach, we used congruency tasks to evaluate passive comprehension. To be able to use congruency tasks (commonly used in cognitive psychology experimentation) 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". These sentences do not require direct action but comprehension and awareness on the pilot's end (c.f Figure 2).

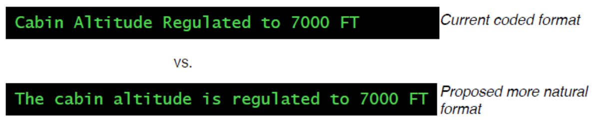

Figure 2, Example of MCL and MNL in an Informational Statement

2.1 Construction of Messages

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 (cf. Jahchan, 2019).

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."*

2.2 Stimuli

We created a new corpus of messages inspired by everyday life situations to test our hypothesis with naïve participants that are not familiar with aeronautical corpus terms. An example of this sentences is "parking spot is available", that emulate the syntax and intentions of our original corpus statements. As a first step, the newly proposed structures were purposefully tested on naïve participants (and not pilots) to avoid expert bias and determine comprehension and performance levels on a more general level. The corpus was divided in 6 difficulty categories that represent syntactical structure of the information availability statements. They went from 1 the easiest structure (noun + nous + 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 Szmrecsanyi (2004). 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. Once the messages were set, we looked for, constructed, or modified existing real life images which accurately portrayed the messages we previously concocted, which have similar syntactic structure and difficulty as messages present in the original corpus (MCL), and for which we created a corresponding MNL version (c.f. Figure 3)

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 3, Example of 6 conditions of difficulty

As messages were different in length, the allotted 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 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).

2.3 Experimental Design and Participant Task

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 Control Language. 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). The levels range from A1 or breakthrough/ beginner to C2 or Mastery/Proficiency.

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"). Participants had 5000 ms to respond. This time lapse was validated by doing several pretests to ascertain the adequate display time for reading the messages. 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. The task consisted of the participants reading a text written in either the More controlled Language (MCL) syntax or the More Natural Language (MNL) syntax (c.f. Figure 4)

The messages appear out of context preceded only by a 3000 ms fixation cross in the middle of the screen. We decreased that value to 150 words per minute (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. The text (the prime) then disappears and a target image appears, an image which could be congruent with the previously read text or incongruent. I.e. if the text 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 language conditions were recorded. We chose sentences that could show an accurate visual description of a situation or scene.

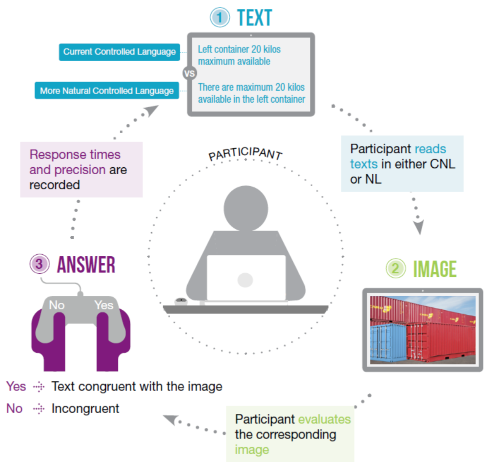

Figure 4, Representation of Task Performance

2.4 Participants

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.

2.5 Experimental Materials and Equipment

DMDX 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 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).

2.6 Variables

The list of independent variables that we will evaluate are:

- Language (MCL-MNL)
- Syntactic Difficulty (1 to 6)
- Type (Congruents-Incongruents) Extraneous and participant variables:
- English placement level (Basic Intermediate, Proficient, Mastery, Native)
- Familiarity with Airbus CL (None, Beginner, Intermediate, Expert)

Dependent variables:

- Reaction time in ms, Accuracy (number of errors)

2.7 Hypotheses and Research Questions:

1. MNL messages produce shorter reaction times than MCL ones in different syntactic difficulty conditions.
2. MNL messages produce less errors (are more accurate) than MCL ones in different syntactic difficulty conditions.
3. Did the language factor play a different role for the different types of congruency responses regarding reaction times?
4. Did the language factor play a different role for different levels of English placement (Basic Intermediate, Mastery, Natives) regarding reaction times?

3 Results and Statistical Analysis

We reported the results below linked to each of the previously mentioned hypotheses. We used non-parametric statistical significance tests such as Wilcox signed rank as the data had a non-normal distribution (Gaussian distribution). These tests help determine whether the independent variables had an effect on reaction time and accuracy of comprehension (dependent variables) by calculating a statistical significance p-value (results are significant if they show a p-value less than 0.05, i.e. implying that it is acceptable to have less than 5% probability of incorrectly rejecting the true null hypothesis). **1. MNL messages produce shorter reaction times than MCL**

ones in different syntactic difficulty conditions. A Shapiro-wilk normality test was run on the reaction times and the results showed that the data is significantly non-normal (p = 2.054e-05) with abnormal skew, therefore we used non-parametric tests to test the main effect such as the Wilcox signed rank test because the same participants took part in both language conditions. Firstly, the general effect was compared regardless of difficulty for both language conditions. 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 syntactic 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 5 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 (confirms Szmrecsanyi (2004) findings). With the hypothesis confirmed, we can also conclude that MNL messages produced consistently faster reaction times than MCL messages in all difficulty conditions.

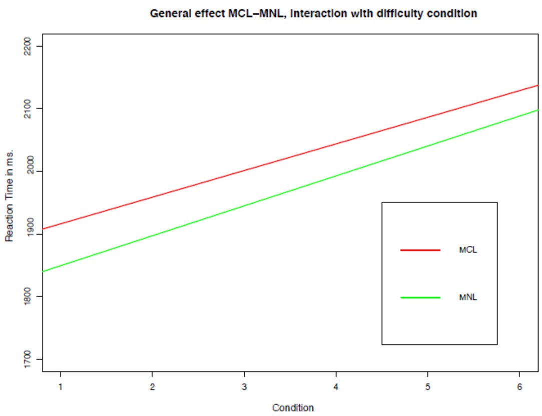

Figure 5, Linear Regression Models for MCL and MNL Difficulty Condition

2. MNL messages produce less errors (are more accurate) than MCL ones in different syntactic difficulty conditions. 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. 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 6 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.

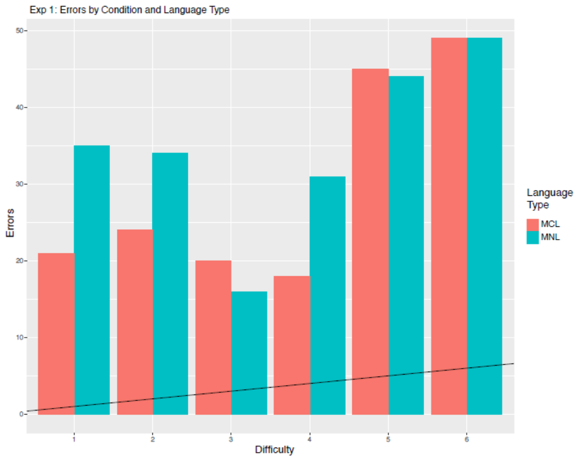

Figure 6, Histogram of errors in MNL and MCL in the 6 difficulty conditions

3. Did the language factor play a different role for the different types of congruency responses regarding reaction times? It was important 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 compared 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

Figure 7, Medians in ms of MCL and MNL reaction times in congruent and incongruent stimuli

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

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 “native” 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. As we can see the only significant result is the basic intermediate level. We can 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 (averages which showed statistical significance). Native speakers often mentioned that they preferred the more natural language, and this is also apparent in their results. 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 8).

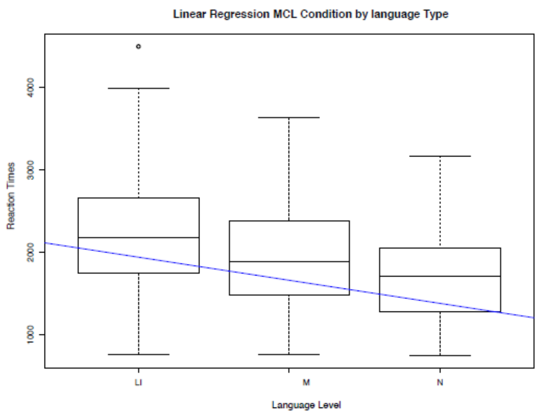

Figure 8, Linear regression of MCL in the different English Placement Levels

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 ms – 190.20 ms) for every English placement level gained. Therefore, reaction times decreased 190.20 ms for every English placement level gained (cf. Figure 39).

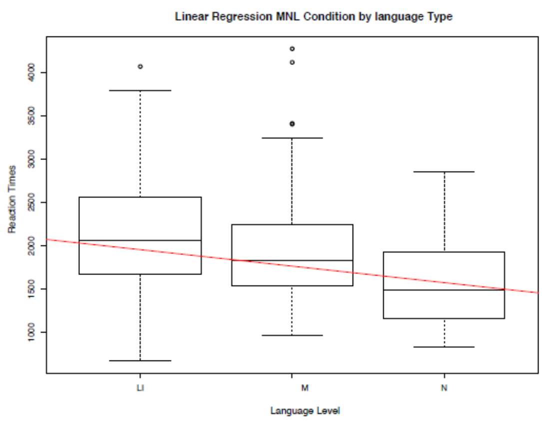

Figure 9, Linear regression of MNL reaction times in the different English Placement Levels

A graphical representation of both of those linear regressions is shown in Figure 10. As we can see, there is no interaction between these 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.

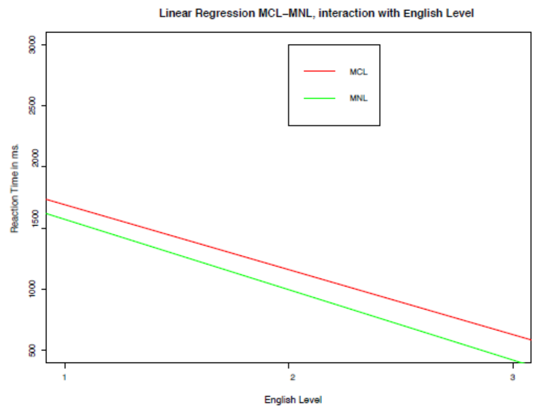

Figure 10, Linear regression for both MCL and MNL reaction times in the different English Placement Levels

4 Discussion

As shown in the results of hypothesis 1, MNL condition shows significantly faster reaction times than MCL condition, and both languages performed equally with regards to accuracy (in hypothesis 2). This could be explained by the fact that the syntactic changes (sentential elements in constative statements) between the two language 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. In the experiment speed of stimuli presentation and to a certain degree the stress it provoked, accentuated the role of the more natural language in information processing. Additionally, 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 syntactic 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. Incongruent stimuli showed significantly faster reaction times for the MNL condition over the incongruent MCL condition, while the congruent stimuli did not. Therefore, in cases of increased difficulty the more natural language helps ease comprehension. Concerning English placement levels (research question 4), MNL seems to facilitate comprehension for participants in the basic intermediate level placement, and this suggests that speakers with weaker levels of English proficiency would benefit more greatly from a more natural language than confirmed speakers, or at least we could say that the effect is more conspicuous. While native English speakers performed better on average in the MNL condition, the effect was not statistically significant and should be the object of further studies with bigger samples of native speakers. We could also conclude that there is no interaction between the reaction time of the two language conditions 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.

5 Conclusion

We presented in this experiment a congruency task similar to traditional judgment tasks in behavioral experiments. It provided a firmly controlled environment to test linguistic hypotheses and CNL rules, 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. In subsequent experiments, the congruency tasks will be replaced by ecological performance tasks for injunctive statements (participants performed the action required and the accuracy and response times are recorded) which include the urgency factor (speed of stimuli, and stress generated by limited response time). We will also be recruiting more native speaker participants to have a more representative panel of the target population (pilots from all around the globe), and ascertain whether the different syntactic language conditions reflect equally on native and non-native English speakers.

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 common misconceptions, results showed that more simplification and linguistic economies and ellipses hardly ever led to better performance (MCL conditions did in no condition show significantly better reaction times or accuracy than MNL conditions). Furthermore, this experiment brought us first elements of empirically tested data which question controlled language construction, and the limits of simplification in general. It showed that what we sometimes mistakenly label as superfluous or empty syntactical elements (such as grammatical words as opposed to lexical words) could go a long way in ensuring better comprehension and faster information processing from a psycho-linguistic point of view.

References

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.

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).

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. 2021. *Towards an ergonomic linguistics: Application to the design of controlled natural languages.*. International Journal of Applied Linguistics, 31(1), 18-30.

Flesch, R. 1979. *How to Write Plain English: Let's Start with the Formula*. University of Canterbury.

ISO. 2016. *Ergonomics of human-system interaction: part 11: usability: definitions and concepts*

(ISO/DIS 9241-11.2:2016). German and English version prEN ISO 9241-11:2016.

Jahchan, N., Condamines, A., Cannesson, E. 2016. *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.

Jahchan, N. 2017. *The importance of using psycholinguistic tools for CNL evaluations*. Actes de Jetou, 99-105.

Jahchan, N. 2019. *To what extent does text simplification improve human comprehension?: cognitive evaluations for the optimization of the Airbus cockpit controlled language for future aircraft*. (Doctoral dissertation, Universite Toulouse le Mirail-Toulouse II).

Kittredge, R. I.. 2003. *Sublanguages and controlled languages*. In Ruslan Mitkov, editor, The Oxford Handbook of Computational Linguistics (pp. 430-447).

Kuhn, T. 2014. *A Survey and Classification of Controlled Natural Languages. Computational Linguistics*. 40(1) (pp. 121-170).

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.

Leroy, G., Helmreich, S., 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 (pp. 1-10). IEEE.

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., 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).

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.

Spaggiari, L., Beaujard, F., Cannesson, E. 2003. A Controlled Language at Airbus. *Proceedings of EAMT-CLAW03 (pp. 151-159).*

Szmrecsanyi, 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).

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)

Warnier, M. . 2018. *Contribution de la linguistique de corpus a la constitution de langues controlee pour la redaction technique : l exemple des exigences de projets spatiaux.*. Ph.D. thesis, University of Toulouse - Jean Jaures.