

# User-Centered Approach to Designing and Evaluating Controlled Languages in the Nuclear Industry

Eléna Martel<sup>1,2</sup>, Amaël Arguel<sup>1</sup>, Anne Condamines<sup>1</sup>, and Julien Kahn<sup>2</sup>

<sup>1</sup>CLLE, University of Toulouse, CNRS, Toulouse, France

<sup>2</sup>EDF R&D, Palaiseau, France

## ABSTRACT

In high reliability organizations such as the nuclear sector, operational safety depends not only on technical expertise but also on the clarity and interpretability of written instructions. To mitigate the risks of misinterpretation, controlled languages (CL) have been proposed. However, CL rules are frequently designed top-down without sufficient involvement of end users. This study presents a user-centered approach to designing and evaluating a CL for procedural documentation in the French nuclear industry. Drawing on ergonomic linguistics and cognitive models of text comprehension, we designed an experiment involving 89 participants (55 nuclear experts and 34 non-experts) to evaluate 10 linguistic rules. Participants completed comprehension tasks on procedural instructions presented in two versions (with and without CL rules) across both nuclear and general (cooking) domains. This cross-domain design allowed us to disentangle domain expertise effects from general linguistic processing abilities and to identify rules that benefit all readers regardless of their background knowledge. Results showed that four rules yielded significant improvements in comprehension, with three producing universal benefits and others revealing expertise-dependent effects. Beyond domain knowledge, experts exhibited structural habituation to complex syntactic patterns, influencing their performance even on neutral content. These findings suggest that expert resilience to complexity is both semantic and syntactic, with implications for CL design.

**Keywords:** Controlled language, Ergonomic linguistics, Expertise, Nuclear industry

## INTRODUCTION

In high-reliability organizations, natural language acts as a hidden danger for operational safety. Condamines (2008) introduced the concept of *language risk* (in French: “*risque langagier*”) to describe the potential gap between a speaker’s communicative intention and the receiver’s interpretation of the message. In contexts where precision and clarity are necessary for operational safety, even minor ambiguities in technical documentation can lead to material failures, safety incidents, or discrepancies between prescribed procedures and real work.

One approach to address language risk is the implementation of *controlled languages* (CL). According to Kuhn (2014), a controlled 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”. One of the most widely used CLs is ASD-STE100 (Simplified Technical English), developed by Aerospace, Security and Defence Industries Association in Europe (ASD, 2007). Despite their potential benefits, traditional CLs face several limitations. First, they are often too restrictive, failing to accommodate the flexibility needed for complex technical descriptions (Warnier and Condamines, 2017). Second, they can be difficult for writers to apply consistently due to the high level of cognitive load required to memorize the set of rules (Renahy, 2015). Third, their development and maintenance require significant resources in time and specialized expertise. Furthermore, many CLs have been designed in a “top-down” approach, driven either by domain experts, who may be desensitized to linguistic ambiguity, or by technical writers who possess limited visibility into operational field level constraints. This top-down approach often results in rules that lack empirical validation and may not address properly the actual sources of comprehension difficulty experienced by end users. To address this limitation, this research adopts an approach in the French nuclear industry that embedded users (nuclear engineers) in both the creation of the CL linguistic rules and its evaluation.

## **THEORETICAL FRAMEWORK**

### **Ergonomic Linguistics**

This research is grounded in the framework of *ergonomic linguistics* named by Condamines (2021). This approach defines itself as an applied linguistics centered on the user, which considers language as a work tool whose effectiveness must be optimized. This can be done by integrated perspectives from corpus linguistics, psycholinguistics and ergonomics.

A central concept borrowed from ergonomics is *usability*, defined in ISO 9241-11 (2018) as the “degree to which a system, product can be used by specified users to achieve specified goals with effectiveness, efficiency, and satisfaction in a specified context of use”. Applied to technical documentation, usability encompasses not only whether a text is grammatically correct or stylistically acceptable, but whether it enables users to accomplish their tasks accurately and efficiently. This user-centered approach shifts the focus from abstract linguistic properties to measurable outcomes in context.

### **From Text to Coherent Representation of Meaning**

To understand how readers process procedural texts and to explain how CL rules can influence comprehension, we drew on the model of text comprehension proposed by Van Dijk and Kintsch (1983). According to this model, comprehension involves constructing mental representations at three distinct levels:

1. *Surface structure*: the exact words, phrases, syntactic structures of the text. This representation is short-lived and quickly decays.

2. *Text base*: which captures a propositional representation of explicit content and local coherence relations (microstructure) as well as higher level thematic organization (macrostructure).
3. *Situation model*: an integrated representation of the described situation, combining text information with prior knowledge and inferences.

### **Working Memory and Text Processing**

The construction of text representation depends on working memory (Van Dijk and Kintsch, 1983), a limited-capacity system responsible for the temporary maintenance and manipulation of information. Baddeley and Hitch (1974) proposed a multicomponent model of working memory comprising three main components: the phonological loop for verbal information, the visuospatial sketchpad for visual and spatial information, and the central executive for attentional control and coordination. Baddeley (2000) later added a fourth component, the episodic buffer which serves as limited capacity temporary storage system that binds information from the subsidiary systems and from long term memory into episodic representations. This component is particularly relevant for text comprehension, as it provides the mechanism by which incoming textual information is integrated with prior knowledge to form coherent situations models.

This limited capacity of working memory has direct implications for text comprehension, thus CL design. Complex syntactic structure, lengthy sentences, and ambiguous expressions all impose extraneous demands on working memory resources. When these demands exceed the available capacity, comprehension suffers. This theoretical perspective justifies rules that limit sentence length, simplify syntactic structures, and reduce ambiguity, all aimed to reducing cognitive load during reading.

### **Expertise, Schemas**

Beyond working memory limitations, comprehension also depends on the knowledge stored in long-term memory. In schema theory, schemata exist “for generalized concepts underlying object, situations, events, actions, and sequences of actions” (Rumelhart and Ortony, 1977). This schemata theory is directly compatible with the situation model approach adopted in discourse comprehension models. Kintsch formalized this link by defining deep comprehension as the construction of a representation in which “the information provided by the text is elaborated from prior knowledge and as integrated with it” (Kintsch, 1994). Zwaan and Radvansky (1998) proposed a clarification: schemata correspond to types (“stereotypical situations”) whereas situation models correspond to tokens (representations of a specific episode), and they characterize schemata as “building blocks” for construction situation models.

From this perspective, domain expertise can be in part characterized as the availability of richer, more strongly organized domain schemata, which can make situation model construction more efficient and more resilient to complex formulation. This is consistent with evidence that readers with strong background knowledge may succeed even when a text leaves more implicit: in Kintsch’s review, high knowledge readers on version of texts that are less

coherent and require more inferential “gap filling”, whereas low knowledge readers benefit from fully explicit versions (Kintsch, 1994). Likewise, in a study using baseball as the knowledge domain, prior knowledge produced a robust advantage across comprehension and recall measures, this advantage did not interact with general reading ability (Recht and Leslie, 1988). These findings imply that comprehension outcomes for procedural instructions may reflect both linguistic formulation and the reader’s domain schemata potentially allowing experts to compensate for linguistic under specification in ways that non-experts cannot.

This theoretical perspective motivates our experimental design, which includes both domain specific (nuclear) and domain neutral (cooking) materials. By controlling for domain knowledge through the cooking corpus while maintaining parallel syntactic structure, we can distinguish between semantic expertise effects and structural habituation effects.

## **METHODOLOGY**

We conducted a comparative, user-centered study combining co-design and validation of controlled language (CL) rules with end users and an experimental evaluation of their impact on sentence-level comprehension. The experiment used a within-participant design: each participant completed comprehension trials on both domain-specific nuclear instructions and cooking instructions, in non-reformulated (NR) and reformulated (R) versions.

### **Participants**

A total of 89 participants took part in the experiment: 55 nuclear experts (nuclear engineers) and 34 non-experts with no nuclear-domain knowledge. Expert participants were recruited through the industrial partner; non-experts were recruited outside the nuclear sector. All participants were fluent French speakers.

To mitigate fatigue effects given the length of the protocol, the experiment was designed in two distinct blocks, each targeting a specific subset of five linguistic rules (Block A for rules 1-5; Block B for rules 6-10). Participants were offered the flexibility to complete the study in one or two sessions, which led to some attrition between the first and second block. Consequently, in the expert’s group, 27 participants completed the full experiment (both blocks), while 28 participants completed only the first block. Similarly, in the non-expert group, 21 participants completed both blocks, whereas 13 participants discontinued after the first block.

Despite this attrition, the presentation order of items was counterbalanced across participants with respect to corpus and version. This design ensured that data obtained from partial completions remained statistically valid and free from systematic order effects for the specific rules tested.

### **Controlled Language Rules and Materials**

We evaluate ten CL rules targeting recurrent sources of ambiguity in procedural writing. These rules addressed various linguistic dimensions

including: sentence length (rule 1: limit sentences to 25 words), polarity (rule 2: use positive formulations), anaphoric reference (rule 3: prefer noun phrases over pronouns, e.g., “the valve” instead of “it”, to avoid ambiguity about what is being referenced), conditional marking (rule 4: use explicit “if/otherwise” structures; rule 5: place conditions before main clauses), injunction expression (rule 6: use infinitive verbs for commands: rule 7: use “must be” for required states), and numerical expressions (rule 8: express quantities in Arabic numerals; rule 9: use mathematical symbols for inequalities; rule 10 use “X or more/less” despite “more of X/ less of X” formulations);

Starting from 20 technical instructions extracted from the operational prescriptive document, we produced a reformulated version for each instruction by applying the relevant rule. This yielded 40 nuclear sentences (20 NR vs 20 R). To disentangle syntactic effects from domain knowledge, we created a “general” (cooking) corpus matched to the nuclear corpus at the level of syntactic structure and rule violations but placed in a neutral cooking context with non-technical vocabulary. The cooking corpus contained 40 sentences (20 NR vs 20 R). Overall, the testing set included 80 sentences (40 nuclear + 40 cooking).

### Validation of Reformulations

To ensure meaning preservation, the reformulated nuclear sentences were reviewed and validated by a panel of three domain professionals (technical writers and/ or nuclear experts). Validation focused on semantic equivalence between NR and R versions. Only validated items were retained for experimental testing.

### Procedure and Task

Participants completed on Qualtrics<sup>1</sup> a sentence comprehension experimental test using a within-participant design. Each trial followed a dual-task sequence. First, a three-digit sequence (e.g., 381) was displayed and participants were instructed to memorize it. This concurrent memory task was intended to occupy working memory, thereby increasing reliance on the linguistic form of the sentence rather than on schematic inference from nuclear domain. The target sentence was then presented, followed by a short statement. Participants evaluated the statement by selecting one of four responses options: *yes*, *no*, *maybe*, *I don't know*. Immediately after their response, they were prompted to recall the three-digit sequence.

The experiment was divided into two blocks: Block A included items targeting rules 1-5 and block B included items targeting rules 6-10. Within each block, the presentation order of items was counterbalanced across participants with respect to corpus (nuclear vs. cooking) and version (NR vs. R). To control for variation in processing time, sentences were presented according to their word count. A short practice block was provided before

---

<sup>1</sup><https://www.qualtrics.com/>

the experimental trials to familiarize participants with the procedure and response format.

This study protocol was approved by the University of Toulouse Research Ethics Committee (approval no. 2025\_1077, May 23, 2025).

## RESULTS

We analysed data using mixed effects models with version (NR vs R.), corpus (nuclear vs cooking), and participant group (expert vs. non-expert) as fixed factors. Two dependent variables were examined: Accuracy (correct responses) and Response Time. However, only Accuracy analysis will be addressed in the present paper. The results are shown in Table 1. Of the ten rules reevaluated, four yielded statistically significant results, which we organize into three categories based on their effect patterns.

**Table 1:** Statistical significance of effects (p-values).

Controlled Language Rule	Main Effects		Two-Way Interactions	
	Version	Participants	Participants x Domain	Version x Domain
R1. Sentence length	.897	.001 (.367)	.001(.162)	.003(.110)
R2. Avoiding negations	.001(.231)	.300	.003(.107)	.059
R3. Noun phrases vs pronouns	.384	.001(.194)	.007	.117
R4. Explicit conditional marking (if/otherwise)	.206	.073	.011	.003(.106)
R5. Condition before subject/verb	.641	.001(.143)	.001(.251)	.394
R6. Infinitive verb for injunctions	.605	.023	.621	.695
R7. “must be” for required state	.202	.023	.400	.444
R8. Quantities in Arabic numerals	.004(.279)	.119	.586	1.000
R9. Mathematical symbols	.220	.895	.948	.206
R10. Only Y or more/Y or less	.001(.266)	.006	.856	.976

*Note.* Bold values indicate  $p < .05$ , values in parentheses are partial eta squared ( $\eta^2p$ )

### Rules With Universal Benefits

Three rules produced significant main effects that benefited all participants regardless of expertise level or corpus domain. Rule 2 (positive polarity: avoiding negations in sentences and conditions) showed a highly significant main effect ( $p < .001$ ). The reformulated versions significantly improved accuracy for both expert and non-expert groups across both corpora.

Rule 8 (expressing quantities in Arabic numerals rather than spelled out words) also showed a significant main effect ( $p = .004$ ). These results suggest that numerical notation provides a clear processing advantage over verbal expression of quantities.

Rule 10 (using standardized expressions “X or more”/“X or less” instead of ambiguous comparative formulations) demonstrated a significant main effect ( $p < .001$ ).

### **Rules With Expertise-Dependent Effects**

Rule 4 (using explicit conditional markers “if/otherwise” instead of implicit conditional structures) revealed a particularly interesting pattern. The simple effect of the rule was not significant, but a significant Domain x Version interaction emerged ( $p = .003$ ). When reading nuclear content, experts showed no improvement from the reformulation, their familiarity with domain context allowed them to correctly interpret even ambiguous conditional structures. Domain knowledge masked the ambiguity that the rule was designed to address. However, when the same syntactic structure appeared in the cooking corpus (where domain knowledge could not compensate), the benefit of conditional marking became significant for all participants. This finding suggests that the absence of an effect on domain specific content does not necessarily indicate that a rule is ineffective rather, it may indicate that expertise provides an alternative resolution pathway.

### **Rules With No Significant Effects**

Six rules did not show significant main effects on Accuracy. Rule 1 (limiting sentence length to 25 words) showed no significant effect on Accuracy ( $p = .897$ ). A significant two-way Version x Domain interaction was observed ( $p = .003$ ). However, post hoc analysis revealed that this interaction did not reflect a benefit of the reformulation. The effect was driven by a difference between specific items pairs (e.g., performance on 1 nuclear NR item vs. 1 cooking R item) rather than by a systematic improvement when applying the rule.

Rule 3 (preferring noun phrases over pronouns) did not show significant effects on any measure. Analysis suggested potential weaknesses in the experimental items for this rule, warranting further investigation with improved materials. This null result reflect a limitation of the sentence level evaluation: each item presented a single sentence in isolation, the pronoun and its antecedent necessarily appeared within the same sentence, eliminating potential competitor referents.

Rules 5 (placing conditions before main clauses), 6 (using infinitive verbs for conjunctions), and 9 (using mathematical symbols for inequalities) similarly showed no significant main effects. These rules may require retesting with refined materials or may genuinely have minimal impact on comprehension at the sentence level.

Rule 7 (using “must be” to express required states) showed no significant effect on any measure. This absence of effect likely reflects the fact that this formulation is already widely used in existing nuclear documentation for expressing required states but not systematically. As a result, experts have developed established schemas for interpreting “must be” constructions, which they transferred to the cooking corpus.

## **DISCUSSION**

### **Interpreting the Pattern of Results**

Our findings provide empirical support for the effectiveness of controlled language rules in improving text comprehension, while also revealing

important nuances related to expertise. The three rules that showed universal benefits (positive polarity, Arabic numerals, and standardized comparative expressions) share a common characteristic: they could reduce processing demands at a fundamental level that is not easily compensated by domain knowledge. Negative constructions require additional cognitive operations to process (Clark and Chase, 1972), spelled-out numbers must be converted to numerical representations, and ambiguous comparatives require inference to determine their precise meaning. These cognitive demands affect all readers, regardless of expertise.

In contrast, rules addressing structural ambiguity (such as conditional marking) revealed existing mental schemas. Domain experts might have developed rich mental models that allow them to resolve ambiguities automatically, drawing on their knowledge of typical procedures, expected sequences, and domain constraints. When reading familiar content, they may not even perceive ambiguities that would confuse less experienced readers. This finding has important implications for CL development: evaluation studies that rely solely on expert feedback may underestimate the comprehension difficulties faced by novices or by experts reading unfamiliar content, or by experts operating under degraded conditions such as time pressure, fatigue, or high cognitive load situations common in real operational contexts.

### **The Nature of Expertise**

Our results highlight an unanticipated dimension of expertise that extends beyond domain-specific knowledge. Experts consistently outperformed non-experts even on the cooking corpus, where domain knowledge should provide no advantage. This pattern suggests that experienced readers of technical documentation have developed a form of structural habituation a familiarity with the complex syntactic structures, nested conditions, and lengthy terminology conventions typical of specialized procedural texts.

According to Kintsch's model of text comprehension (1994), advantages associated with expertise are primarily attributed to richer domain knowledge, which supports the construction of coherent situation model and allows readers to compensate for underspecified or weakly coherent texts. However, our findings suggest that expertise may confer benefits beyond situation model construction. The fact that technical documentation experts outperformed non-experts on an unfamiliar domain indicates that their advantage also operates at the level of processing complex surface structures. In other words, their advantage is not only in what they know about the domain, but also in how they process the linguistic structures characteristic of technical writing. This finding suggests that expert "resilience" to textual complexity is both semantic (related to domain knowledge) and syntactic (related to processing fluency with specific text types). Experts who have developed fluency with existing conventions may perceive simplified formulations as unnecessary, unfamiliar, or even disruptive to their comprehension. Imposing linguistic recommendations, designed solely from a linguistic perspective without user input may therefore encounter "resistance" from the users the CL is meant to serve. This underscores the importance of co-designing CL rules with end

users, ensuring that formulations are perceived as genuine improvements rather than arbitrary formulations.

### **Limitations and Future Directions**

Several limitations of this study should be acknowledged. First, the sentence-level evaluation paradigm, while allowing precise isolation of rule effects, did not capture the full complexity of document use in operational contexts. Linguistic effects may have been diluted, distributed, and embedded within broader reading strategies when documents are used in authentic work situations.

Second, some rules showed no significant effects, which may reflect either genuine lack of impact or weaknesses in the experimental materials. Further refinement of test items and additional testing is needed before concluding that these rules are ineffective in this context.

### **CONCLUSION**

This study demonstrated the value of a user-centered, empirically grounded approach to controlled language development. By involving end users throughout the design process and testing rules through experimentation, we identified both universally effective rules and expertise-dependent effects.

Our key findings can be summarized as follows. First, rules targeting fundamental cognitive processes (positive polarity, numerical notation, standardized comparatives) produced universal comprehension benefits. Second, rules addressing structural ambiguity may have been masked by expertise when tested on domain-specific content, revealing an “expertise wall” that must be accounted for in CL evaluation. Third, expertise encompassed not only domain knowledge but also structural habituation to genre-specific linguistic patterns. These findings have practical implications for CL design in high-reliability organizations and theoretical implications for understanding the relationship between linguistic form, cognitive processing, and professional expertise. Future work will target less experienced readers within the nuclear domain, that are individuals with less established schemas to better assess the benefits of controlled language rules to encourage the reading of technical documents and familiarization with them.

### **ACKNOWLEDGMENT**

This research was funded by the ANRT (Association Nationale de la Recherche et de la Technologie) under CIFRE PhD fellowship (Convention Industrielle de Formation par la Recherche) and by EDF R&D (Electricité de France) and was conducted in partnership between EDF R&D and CLLE (CNRS, University of Toulouse).

### **REFERENCES**

Aerospace and Defence Industries Association of Europe (2007). Simplified Technical English. Specification ASD-STE100. International specification for the preparation of maintenance documentation in a controlled language. Issue 4.

- Baddeley, A. (2000). The episodic buffer: A new component of working memory? *Trends in Cognitive Sciences*, 4(11), pp. 417–423.
- Baddeley, A. D., and Hitch, G. (1974). Working memory. *The psychology of learning and motivation*, 8(8), pp. 47–89.
- Clark, H. H., and Chase, W. G. (1972). On the process of comparing sentences against pictures. *Cognitive Psychology*, 3(3), pp. 472–517.
- Condamines, A. (2008). Can we prevent language risks in written communication? *Language and Society*, 125, pp. 77–97.
- Condamines, A. (2021). Towards an ergonomic linguistics: Application to the design of controlled natural languages. *International Journal of Applied Linguistics*, 31(1), pp. 18–30.
- ISO 9241-11 (2018). Ergonomics of human-system interaction Part 11: Usability: Definitions and concepts. International Organization for Standardization.
- Kintsch, W. (1994). Text comprehension, memory, and learning. *American psychologist*, 49(4), pp. 294–303.
- Kuhn, T. (2014). A survey and classification of controlled natural languages. *Computational Linguistics*, 40(1), pp. 121–170.
- Recht, D. R., and Leslie, L. (1988). Effect of prior knowledge on good and poor readers' memory of text. *Journal of educational psychology*, 80(1), pp. 16–20.
- Renahy, J., A. Vuitton, D., Rath, B., Thomas, I., de Grivel, V and Cardey, S. (2015). Controlled language and information on vaccines: Application to package inserts. *Current Drug Safety*, 10(1), pp. 41–48.
- Rumelhart, D., and Andrew Ortony. 1977. "The Representation of Knowledge in Memory." In *Schooling and the Acquisition of Knowledge*, eds. R. C. Anderson, R. J. Spiro, and W. E. Montague, (1st ed.). Routledge. pp. 99–135. doi:10.4324/9781315271644.
- Van Dijk, T. A., and Kintsch, W. (1983). *Strategies of discourse comprehension*. New York: Academic Press.
- Warnier, M., and Condamines, A. (2017, September). Evaluating the relevance of some rules for writing requirements through an online survey. In *International Requirements Engineering Conference (RE'17)*.
- Zwaan, R. A., and Radvansky, G. A. (1998). Situation models in language comprehension and memory. *Psychological bulletin*, 123(2), pp. 162–185.