Towards a More Semantically Transparent *i** Visual Syntax

Nicolas Genon[1], Patrice Caire[1], Hubert Toussaint[1],
Patrick Heymans[1,2], and Daniel Moody[3]

[1] PReCISE Research Centre, University of Namur, Belgium,
[2] INRIA Lille-Nord Europe, Université Lille 1 – LIFL – CNRS , France
[3] Ozemantics Pty Ltd, Sydney, Australia
{nge,pca,hto,phe}@info.fundp.ac.be
daniel@ozemantics.com.au

Abstract. [Context and motivation] *i** is one of the most popular modelling languages in Requirements Engineering. *i** models are meant to support communication between technical and non-technical stakeholders about the goals of the future system. Recent research has established that the effectiveness of model-mediated communication heavily depends on the visual syntax of the modelling language. A number of flaws in the visual syntax of *i** have been uncovered and possible improvements have been suggested. **[Question/problem]** Producing effective visual notations is a complex task that requires taking into account various interacting quality criteria. In this paper, we focus on one of those criteria: *Semantic Transparency*, that is, the ability of notation symbols to suggest their meaning. **[Principal ideas/results]** Complementarily to previous research, we take an empirical approach. We give a preview of a series of experiments designed to identify a new symbol set for *i** and to evaluate its semantic transparency. **[Contribution]** The reported work is an important milestone on the path towards cognitively effective requirements modelling notations. Although it does not solve all the problems in the *i** notation, it illustrates the usefulness of an empirical approach to visual syntax definition. This approach can later be transposed to other quality criteria and other notations.

Keywords: *i**, Goal-oriented modelling, Empirical evaluation, Physics of Notation, Semantic Transparency

1 Introduction

*i** [1] is one of the most popular modelling languages for Requirements Engineering (RE). It provides conceptual and visual means to express, and reason on, the functional and non-functional goals of a system. Its visual syntax is meant to facilitate communication between technical and non-technical stakeholders. However, this assumption has been challenged recently. Moody *et al.* [2, 3] have evaluated the visual syntax of *i** against the Physics of Notations [4] (PoN). PoN is a theory comprised of nine principles, namely Semiotic Clarity, Perceptual Discriminability, Semantic Transparency, Complexity Management, Cognitive Integration, Visual Expressiveness, Dual Coding, Graphic Economy and Cognitive Fit. A major advantage of those principles is that they are evidence-based: they do not rely on common sense and experience but on theory and empirical evidence from a wide range of fields, including linguistics, cartography, cognitive psychology...

Following these principles is meant to lead to more *cognitively effective* notations, i.e. notations which diagrams can be understood quickly, easily and accurately. In [2, 3], a number of suggestions were made in order to improve the cognitive effectiveness of *i**. Although those were made on the basis of the evidence-based principles of PoN, an open question remained: how and to which extent the principles coming from other disciplines transpose to software engineering, and more particularly to RE.

Moody *et al.*'s analysis uncovered a number of flaws in the visual syntax of *i** and suggested various improvements. One of them was to improve the *Semantic Transparency* of *i**. Semantic Transparency refers to the ability of the symbols of a notation to suggest their meaning. Semantically transparent symbols can be seen as the visual equivalent of onomatopoeia. For example, a stick figure is more semantically transparent than an abstract shape (e.g., a circle) to represent the concept of person. According to the PoN, Semantic Transparency has a major influence on the cognitive effectiveness of a notation.

In [2, 3], Moody *et al.* proposed a set of supposedly more semantically transparent symbols for *i**. Our work aims at evaluating and complementing their proposal with experimental studies. We defined a series of controlled experiments to identify a "super" symbol set for *i** and to assess its semantic transparency. The main difference with previous research lies in the way new symbols were obtained: the authors of the present work did not design a new symbol set by themselves, based on some theory. On the contrary, an experiment was set up where participants were asked to draw what they thought would be the most appropriate symbols given the *i** concepts and their definitions.

We present the plan of this experiment series in Section 2. Then, we describe the three experiments of that series that were already performed: the "production of drawings" experiment (Section 3), the "population stereotype" experiment (Section 4), and the "population prototype" experiment (Section 5). We share the preliminary results obtained for each experiment. Section 6 wraps up the paper and gives an overview of future work.

2 Experiment Plan

This series of controlled experiments consists in identifying a "super" symbol set for *i** and assessing its semantic transparency. By "super symbol set", we mean the symbols that are judged the most semantically transparent by *i** users. The eligible symbols are taken from 4 sources: the original *i** symbol set [1], the symbols proposed in [3], and 2 sets based on the outcome of the present experiments. In this work, we do not evaluate the semantic transparency of the symbol sets *in context*, i.e, by exposing participants to diagrams and letting them perform RE tasks based on these diagrams. Instead, we focus on the symbols on their own. Thereby, we avoid the biases that occur when dealing with diagrams (e.g., bias due to the relative positioning of symbols; bias due to the number of symbols on the diagram; bias due to the complexity of the diagram; shift of attention introduced by the colour and size of some symbols on the diagram, etc.).

Experiment 1 is concerned with the *production of drawings*. The goal of this experiment is to obtain drawings hand-sketched by participants to represent each *i** concept. We used a sign production technique that relies on the following assumption: "[experiment] subjects, when properly instructed, can actually produce signs for referent concepts and will do so in frequencies proportional to stereotype strength" [5]. The *population stereotype* refers to the sign(s) that is/are the most frequently produced by participants to denote the referent concept. The outcome of the experiment are symbols designed by *i** users for *i** users, contrary to the symbols proposed in [2, 3]. Indeed, the latter may suffer from a strong bias of the authors being RE experts, which could result in symbols that are ineffective for novices or non-technical users.

Experiment 2 focused on identifying the stereotypical drawings out of the results of Experiment 1. All the stereotypes resulting from Experiment 2 would constitute our first new set of hand-sketched symbols for *i**. However, the population stereotype is not sufficient by itself because it does not take into account the level of approximation of the idea depicted by the drawing wrt. the referent *i** concept. Actually, the drawing that is the *most frequently* produced to denote a concept, is not necessarily expressing the idea that is the *closest approximation* to that concept. Conversely, the most "evocative" drawing is usually designed by only a small part of the participants, and then not identified by population stereotype. In other words, while population stereotype can be seen as the best *median* drawing, creativity and originality is captured through the *population prototype*.

Population prototype is the purpose of Experiment 3. This experiment looks for the drawing that *best* represents the corresponding *i** concept. We rely on the personal opinion of participants to elect the drawing that is the most semantically transparent for a given referent concept. The outcome of Experiment 3 would be the second new set of hand-sketched symbols for *i**.

Instead of addressing all the *i** concepts, we limited the scope of this work to 13 concepts: actor, agent, role, position, actor boundary, goal, softgoal, task, resource, belief, means-end link, decomposition link and dependency link. The rationale for this choice is to focus on the subset of the *i** language that appears to be most used in practice.

3 Experiment 1: Production of Drawings

The **question** we addressed in the experiment is the following: "What kinds of drawings can novices produce when presented with a set of concepts and their definitions?"

Experiment design. The participants were composed of 104 students (53 females and 51 males) in 1^{st} year Bachelor in Economics and Management from the University of Namur. These students had no previous knowledge of *i** or modelling in general, which is a profile we expect to find in many real-life RE settings, for stakeholders like users, subject matter experts, and managers. The participants were not remunerated for their contribution.

Each participant was provided with a 14-page booklet, a pencil and an eraser. The first page presented a form to collect participants' demographic data. The remaining 13 pages were respectively dedicated to the 13 *i** concepts. A 2-column table was added at the top of each page. The first column provided the name of the *i** concept in French and English. The second column contained the French definition of the concept[1].

A (3" x 3") frame where participants were asked to sketch their drawing was printed in the middle of the page. The sketching instructions were repeated on each page and placed above the frame. A 5-point scale and the corresponding instructions were added at the bottom of the page. The 5 values of the scale were "easy", "fairly easy", "neither easy nor difficult", "fairly difficult" and "difficult".

We deliberately decided not to randomise the presentation order of the concepts because the definitions of part of the *i** concepts rely on the definitions of other concepts, e.g. Agent, Role and Position refer to Actor; Softgoal refers to Goal.

The 104 students were brought together in an auditorium. The average time for completion of this experiment was around 45 minutes. For each *i** concept, participants were asked *(a)* to sketch what they estimate to be the best drawing to represent the name and the definition of this concept. There was no time limit but they were asked to sketch as quickly as possible. The intent was to capture their intuition. We drew their attention on the fact that we would focus on the idea(s) expressed by the drawing, not the quality of the sketching. *(b)* Each time a drawing was produced, the participant had to evaluate the difficulty of the task on the 5-point scale. Participants were also told to respond one page at time and not to go back in the booklet.

Results. We eventually retrieved 1352 drawings (blank and null drawings included[2]). One of the main observations is that participants had much more difficulty sketching drawings for concepts denoting relationships than for the concepts denoting objects or persons. The reasons to this observation still have to be investigated. We also observed that the produced drawings often do not rely on both the name *and* the definition of the concept. Moreover, some participants depicted concepts not through a single symbol but through several symbols interacting in a scene (e.g., the concept of Task can be represented by a stick figure performing some action on an object).

4 Experiment 2: Population Stereotype

The **question** we addressed in this experiment is the following: "Among the presented drawings, what is the stereotypical representation for the selected concepts?" The population stereotype is the best *median* drawing, that is the representation that is most frequently recognised and selected by people to depict the concept.

Experiment design. We applied a judges' ranking method [6]. Concept per concept, three of the authors categorised the drawings obtained from Experiment 1 based on the similarity of ideas that they expressed. Hence, 13 times (because we considered 13 *i** concepts), each author had to split the 104 drawings into piles. All the drawings from a pile depicted the same idea(s) and thus form a category. The three judges were instructed to define as many categories as needed relying on their personal opinion.

[1] The French version was used to avoid bias regarding the English skills of the participants.

[2] These range from 5 to 15% of the drawings, depending on the concept.

The categorisation process was inevitably prone to a certain degree of interpretation and subjectivity. However, it was required for the judges to follow instructions and to perform the work independently. Afterwards, the judges compared their respective categories and agreed on a common set of categories. In this operation, several categories from different judges were merged into one common category. Finally, for each concept, the judges selected, from the category that contained the largest number of drawings, the drawing that best expressed the ideas of the category.

Results. The outcome of this experiment is the set of 13 stereotypical drawings (one per concept) presented in Figure 1. It is noteworthy that, except for relationship concepts, there is no abstract shape in the population stereotypes.

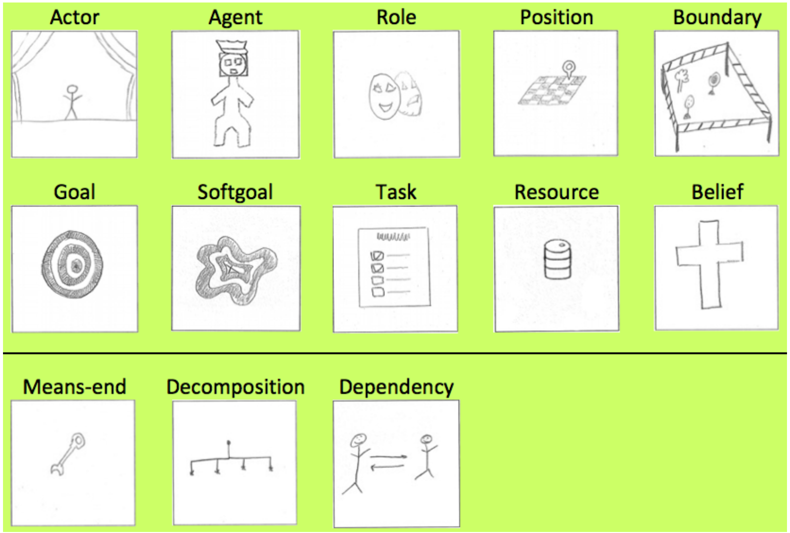

Fig. 1. The population stereotypes for the *i** concepts

5 Experiment 3: Population Prototype

The **question** we addressed in this experiment is the following: “Among the presented drawings, what is the prototypical representation for selected concepts?”

Experiment design. We conducted this experiment on a different sample of population (no overlap) but the participants had the same profile, i.e. they had neither knowledge of *i** nor modelling in general. We opened the experiment to students in 1^{st} year Bachelor in Computer Science or Economics and Management. We welcomed 30 participants (1 female, 29 males). They were not remunerated for their participation. Instructions were to choose one best drawing per concept. As for the drawing production experiment, they were provided with the name and the definition of the concept. The eligible set of drawings was composed of 160 drawings: the 13 stereotypes along with one representative of each category of the 13 *i** concepts. It was a deliberate decision not to expose the participants to all 1352 drawings. This would have been counter-productive: taking them too much time and leading to bias caused by tiredness.

An online questionnaire was set up. The participants were asked to enter their demographic data on the first page and then they navigated through 13 pages, one per *i** concept. Each page displayed the French name and the definition of the concept at the top of the page. The middle of the page was dedicated to instructions for selecting (using radio buttons) the best drawing among the matrix of representatives. The difficulty of the selection task was evaluated on a visual analogue scale (VAS) at the bottom of the page. The order the concepts appearing in the questionnaire as well as the position of the drawings in each matrix were randomised for each participant.

As we built an online questionnaire, we booked a pool hosting 30 computers. To be as compliant as possible with the students' schedules, we ran the experiment from 10.30 AM to 6.00 PM. The 30 students came at their best convenience. The experiment was not constrained by time limit and the average duration was between 5 to 15 minutes.

Results. We have only preliminary results and observations to report. The population prototype obtained for each concept is shown in Figure 2. There was most of the time one indisputable leader – low level of vote dispersal – except for the concept of Dependency Link where we had an *ex-aequo*.

We also noticed that four prototypical drawings matched the stereotype of the concepts: Actor, Goal, Task and Decomposition Link. These drawings can be assumed to have a significant level of semantic transparency: they depict the idea that is the most frequently used by participants and that evokes the referent *i** concept most clearly.

Regarding the difficulty to select the best drawing (measured on the VAS), the value ranged from 30% to 60%. As discussed in the results of Experiment 1, drawings depicting concrete objects or persons seem to be preferred. Except for relationship concepts, there is no abstract shape in the population prototypes.

We observed an adequacy between the nature of the concepts denoting relationships and the nature of their representations: these concepts are depicted with "links".

6 Conclusion and Future Work

In this paper, we gave a preview of a series of empirical studies that aim at improving the semantic transparency of the *i** symbols. So far, we have performed 3 experiments: the first one is concerned with the empirical production of drawings for the *i** concepts by inexperienced subjects. Based on the drawings from Experiment 1, Experiment 2 looked for population stereotypes, i.e., drawings that are *the most frequently* produced to denote a referent concept. Experiment 3 aimed at identifying population prototypes, i.e., drawings depicting the idea that is *the closest approximation* to the semantics of a referent concept. At this stage, we have empirically obtained two new *i** symbols sets: one set is composed of the stereotypical drawings; the second set gathers the best drawings selected in Experiment 3. We also have promising preliminary results that allow us to envision the next steps of our work. We plan to confront our two new symbol sets with two other sets of *i** symbols: the original *i** symbols [1] and the symbols proposed in [3]. The final objective is twofold: to empirically evaluate which of the

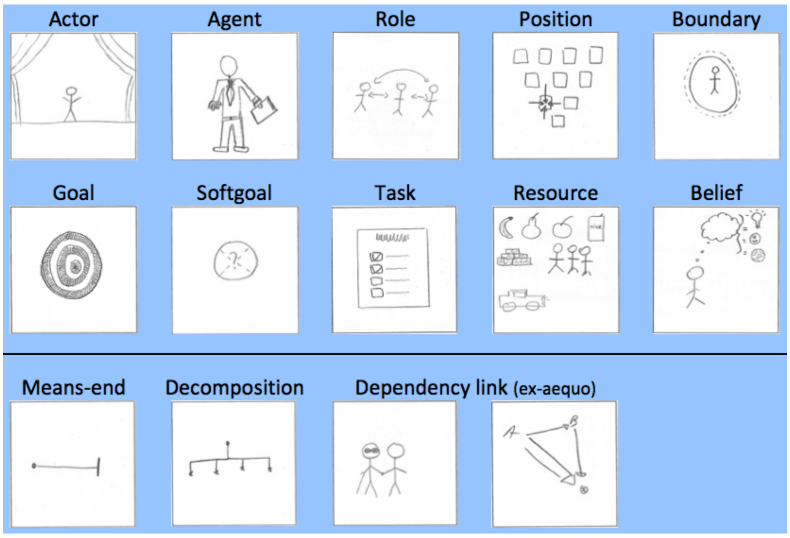

Fig. 2. The population prototypes for the *i** concepts

four symbol sets is the more semantically transparent, and to propose a new *super i**, that could be a combination of all 4 symbol sets.

From a broader perspective, the reported work is an important milestone on the path towards cognitively effective notations in RE and software engineering. Although it does not solve all the problems in the *i** notation [7], it illustrates the usefulness of an empirical approach to visual syntax definition. This approach can later be transposed to other quality criteria and other notations.

Acknowledgement. This work was supported by the Interuniversity Attraction Poles Programme - Belgian State - Belgian Science Policy (MoVES) and the BNB.

References

1. Yu, E.: Towards Modeling and Reasoning Support for Early-Phase Requirements Engineering. In: Proc. of RE'97. (1997) 226–235
2. Moody, D.L., Heymans, P., Matulevičius, R.: Improving the Effectiveness of Visual Representations in Requirements Engineering: An Evaluation of i* Visual Syntax (Best Paper Award). In: Proc. of RE'09. (2009) 171–180
3. Moody, D.L., Heymans, P., Matulevičius, R.: Visual syntax does matter: improving the cognitive effectiveness of the *i** visual notation. Requirements Engineering **15**(2) (2010) 141–175
4. Moody, D.L.: The "Physics" of Notations: Towards a Scientific Basis for Constructing Visual Notations in Software Engineering. TSE **35** (2009) 756–779
5. Howell, W.C., Fuchs, A.H.: Population Stereotype in Code Design. Organizational Behavior and Human Performance **3** (1968) 310–339
6. Jones, S.: Stereotypy in Pictograms of Abstract Concepts. Ergonomics **26** (1983) 605–611
7. Mussbacher, G., Amyot, D., Heymans, P.: Eight Deadly Sins of GRL. In: Proc. of the 5^{th} International *i** Workshop, Trento, Italy, August 2011 (to appear)