Business process modeling: Vocabulary problem and requirements specification

Jonas Bulegon Gassen
Universidade Federal do Rio Grande do Sul, PPGC
Porto Alegre, Brazil
jbgassen@gmail.com

Jan Mendling
Wirtschaftsuniversität Wien
Vienna, Austria
jan.mendling@wu.ac.at

Lucinéia Heloisa Thom
Universidade Federal do Rio Grande do Sul, PPGC
Porto Alegre, Brazil
lucineia@inf.ufrgs.br

José Palazzo M. de Oliveira
Universidade Federal do Rio Grande do Sul, PPGC
Porto Alegre, Brazil
palazzo@inf.ufrgs.br

ABSTRACT

Process models are composed of graphical elements and words. However, words used to name elements during process design have potentially ambiguous meanings, which might result in quality problems. We believe that ontologies might serve as a means to address this problem. This paper discusses aspects related to words used to represent concepts in labels and why ontologies can improve this representation. Also, we analyze how the requirements specifications can influence the terms used during modeling. The discussion regarding ontologies is conceptual. We performed an experiment to analyze empirically the vocabulary problem in the context of process models. In the experiment the selection of terms represented with different levels of explicitness in requirements specifications is evaluated. Our findings suggest that the vocabulary problem occurs in process models. Also, different levels of explicitness affect the labels but are not sufficient to solve the vocabulary problem.

Categories and Subject Descriptors

F.3 [**Logics and meanings of programs**]: Semantics of Programming Languages; F.3.2 [**Semantics of Programming Languages**]: Process models

General Terms

Design and Documentation

Keywords

Business process modeling, vocabulary problem, ontology

1. INTRODUCTION

Business process design is one of the most popular forms of conceptual modeling [28]. It is increasingly employed in academia and commercial organizations [4] and the resulting process models have a crucial role in decision making related to analysis and design of process-aware enterprise systems [5]. Moreover, they are very important to document the business processes executed in organizations [27]. Conceptual models are useful to represent the intention of the modeler or of a group of modelers. The model development consolidates a better communication among the participants. Process improvement and process understanding are also essential aspects concerning process models [12]. In the last years more attention has been given to factors which can improve the quality of business process models, as emphasized in the works by Mendling [19] [20] Recker [27] or Fettke [7]. The SEQUAL framework [14] defines quality based on semiotics and in relation to eight different levels. These levels such as syntactic, semantic and empirical quality are discussed by several authors when analyzing factors of process models quality. In this research, we focus on the present paper on issues that may cause misinterpretation of words used by modelers, which ranges from process design until model interpretation. In turn, the levels of semantics, pragmatics and social quality are of special interest.

This paper analyzes personal and specification factors, which relate to words that modelers chose when describing a process. For personal factors, we test its relevance with reference to the vocabulary problem. For specification factors, we investigate different levels of explicitness of terms in requirements specifications and its influence on modeling. To this end, we conduct an online experiment in order to evaluate the users' behavior during the modeling. Against this background, this paper is structured as follows. Section 2 presents the conceptual framework that supports our discussions. Section 3 summarizes related work. Section 4 describes the experiment design and Section 5 discusses our results. Finally, Section 6 presents the conclusions.

2. BACKGROUND

In this section the contextual framework of our work is discussed. The well-known graphical notations for designing process models, such as Event-driven Process Chains (EPC) or Business Process Model and Notation (BPMN), utilize both abstract graphical elements and text. This combination is described as most suitable for conceptual modeling [21]. Although having graphical (visual channel) and textual (auditory channel) elements, the majority of critical information about the domain is provided by text [18]. We will argue that ontologies have the potential to improve quality. Therefore, we first discuss labeling issues in general and in conceptual modeling. Then, we describe the potential usage of ontologies and hypotheses about expected benefits.

2.1 Labeling issues in general

The usage of natural language can lead to ambiguity in terms of, e.g., synonyms, homonym, hypernyms, or personal disagreements. This entails three major problems:

1. the modeler uses suitable words that are misinterpreted by the reader (reader problem);
2. the modeler uses unsuitable words, which could not be understood by the reader (modeler problem);
3. the modeler uses words that are correct in his point of view but not in the reader's point of view (agreement and social problems).

The general problem is how to know the intention of the modeler when using a word when the real conceptualization is a personal characteristic. Persons have large differences in their spatial and verbal processing abilities [36, 21], thus words employed by each modeler could depend on many details, as past experiences, culture, language, knowledge level, or context. This may also have influence how people understand words used by others.

The personal differences as mentioned above can be related to bounded rationality as defined by Simon [29]. He described decision making as a search process guided by aspiration levels, which are goals that must be reached to a satisfactory level [10]. Alternatives are assessed in an iterative search process until a satisfactory alternative is found. The satisfaction level that triggers stopping can be affected by many factors as available time, cognitive resources, emotional and cultural aspects and so on. Jones [13] divide those factors in two categories: those relating to the nature of the decision maker and the nature of the environment. He notes on their complexity that "Empirical objections to rational choice are so voluminous that they are, in effect, a laundry list of problems" [13]. In this regard, Simon uses the analogy of a pair of scissors, where one blade is the environment and the other, the cognitive limitations of a person. Both are necessary for making a proper cut.

2.2 Labeling issues in conceptual modeling

In this paper we investigate the words used by modelers to create labels in business process models. Applying bounded rationality to this work, the modeler reads the requirements and decides when he needs to add an activity to the model. Then, he decides which label fits best to that activity. The satisfaction level of the modeler apparently affects the label quality of the respective model. For example, one person might prefer "execute payment" and another person "pay order". Novices might have different satisfaction levels when compared to experts, regarding guidelines, conventions or social aspects (for a group).

Consequently, the understanding of a process model does not rely only on comprehension, but also in intentional factors. We can associate this with the gulf of evaluation, presented in Norman [22], which refers to the differences between what a model tries to express and what the model reader interprets. An interpretation can be valid in terms of language, but differing to the intention of the modeler. Thus, it is possible to suggest a problem of social quality when an agreement about the representation does not exist. The same problem occurs in queries for model reuse. The persons that search will use their particular vocabulary, which could be different of the vocabulary used to create the model, leading to false positives or the impossibility to retrieve relevant models to the user.

Let us also consider the problem of ambiguous label grammar [18]: the label "measure processing" could refer to the processing of a measure or to the measurement of a processing. Another example are the words "items" and "products". They are not synonyms according to WordNet[1], however they can be used as such in some contexts. The latter problem also relates to social questions, thus it is very important that all the stakeholders understand the words significance in the right - for the group - meaning.

2.3 Potential usage of ontologies

For the sake of simplicity, we assume that process models are created from a requirements specification presented as text. Following the analogy of the scissor, on one side we have the modeler and on the other we have the world to be modeled plus the requirements specification. In order to improve one side of the scissor, we believe that ontologies can serve as agreement descriptors for the process modelers and readers. In this case, a tool that allows the creation of models with ontology support is needed. The tool can allow the selection [9] or association of terms from ontologies to the inserted labels in the model.

However, the use of ontologies in this context has to be done carefully, otherwise it can generate extraneous cognitive load [31], which could deteriorate its benefits. Thesaurus or dictionaries could be used instead. However, ontologies are more expressive, and therefore potentially avoid extraneous cognitive load. Also, ontologies may allow inferences over the activity labels. Thus, besides verification of the model structure [34], we can verify the labels, for example by means of activity patterns [33].

One of the most well known definitions of ontologies states that an ontology is a formal explicit specification of a shared conceptualization [30]. Here, conceptualization relates to a representation of a portion of reality by means of concepts. It points out that the concepts and constraints are explicitly represented. Formal in this context means that the ontology definition is machine readable. Finally, being shared expresses the idea of consensus, where the representation is accepted by a group of people.

The conceptualization in the ontologies definition has a different sense if compared with the meaning triangle [23]. The triangle conceptualization relates to the concepts in the mind of the person. In formal ontologies [30], it relates to the representation by means of concepts, using a specification language. The use of the term conceptualization in the meaning triangle is nearer to the philosophical meaning of ontologies that Guarino calls of conceptualization [11]. It is

[1]http://wordnet.princeton.edu

different to the engineering artifact, by means of a specification language (e.g. OWL), called also ontology by Guarino. In this point of view, two ontologies (engineering artifacts) can have two different vocabularies and represent a unique conceptualization. For example, two companies within the same domain may use two different representations for the same conceptualization. Since each person has their own way to express and understand the world, we can have many representations for one unique portion of world. Following this idea, one can ask [7]: who decides what makes a "good" model and "correct" representation? As George E. P. Box notes all models are wrong, but some are useful [2]. This means that even if we have a model that follows quality guidelines, some aspects can still represent bad quality from some point of view grounded in the individuality of peoples' way of thinking. Agreements can change the good quality for one person to good quality for a group of involved persons. Based on this individuality and the shared characteristics presented by ontologies, ontologies can be used to represent a vocabulary of a company, for example. The textual symbols used by the process modelers would be selected from the ontologies that represent the application of a company. The group involved in process modeling or interpretation will have access to this shared vocabulary representation. The main idea is to change the moment of agreement, instead of achieve agreements after the modeling of each process, the agreement would occur in the ontologies creation and the modelers would use and improve it.

In this paper, we do not investigate how ontologies would be applied to improve the process modeling; instead, it intends to show the problems that support the need for ontologies. We have implemented a prototypical tool that allows for the creation of process models with ontology support.

2.4 Hypotheses

A possible operationalization of the representation issues has been described as the vocabulary problem. Furnas [8] may have been the first to demonstrate empirically that people often disagree in the words used to name things. We can relate this findings to the meaning triangle [23]. The symbol in the triangle is just one way to access the concept in our mind. If by some reason someone teaches you to name actual dogs as "house", that would be how you will call dogs. Also, the more distinct the words used by people, the more distinct might be the satisfaction points where people stop their search process, regarding the bounded rationality.

Our initial conjecture is that people will represent activities in many different ways. For example, given the same textual description about an activity of a process model for N persons, a significant amount of different labels will be generated. This conjecture is based on the bounded rationality and the vocabulary problem. Having different labels is not always bad, however it may affect directly model comprehension and model reuse. For comprehension, examples that could generate ambiguities were given previously in this paper. For reuse, besides comprehension, search techniques may have problems to find relevant models and may retrieve false positives. We do not intend to show which specific factors influence these differences, only to verify its relevance in the context of process models.

As presented previously, the components of a modeling scenario would be the modeler, the world to be modeled and the requirements. In this sense, our second conjecture is that changing the requirements specifications will affect the produced labels. The operationalization of this change refers to the explicitness of words that should be used in the resultant activity label. More specifically, a description that lacks in specific words for labeling is termed as "without explicit elements" and a description which presents specific words is termed as "with explicit elements". For example: "...ship the material..." is very explicit, in the other hand "...the material will be taken..." it is not. Regarding the vocabulary problem, in the second example, one could refer to the shipment as "transfer", "send", or "deliver". The vocabulary problem in the example interacts with bounded rationality, the modeler could choose "transfer" instead of "send" just because he prefers this word.

For our third conjecture, we need a structured label. Researches have presented the verb-object style as the most suitable for labeling activities of process models [18]. Examples of this style are: "receive invoice" or "submit paper". Complementarily, we use a third element, the subject of the task, composing a triplet: e.g. Subject (Author); Verb (Submit); Object (Paper). This style can also be found in practice [15]. In this context, our third conjecture relies on the idea that if one element is known (let's call it anchor), the following elements will be less distinct. We base this argumentation on the idea that the search space, namely the alternatives that can be considered by a person as presented in the bounded rationality, will be smaller. When an anchor is present, let's say subject (Author), the search space for the verb would be smaller and consequently people would use less distinct words, influencing the amount of different labels. Similar ideas for naming process models have been formulated in [16].

In a BPMN model, the subject would be a Pool or Lane, which are used to represent the participants of a process. The verb and object would form the label of an activity within the related Pool or Lane. We used this triplet representation as basis for our experiment. Our hypotheses are:

- **H1:** Textual representations with explicit element(s) will significantly decrease the number of distinct words for the element(s) in question;
- **H2:** Textual representations with explicit elements will significantly decrease the number of distinct words for the other elements;
- **H3:** Textual representations with explicit element(s) will significantly decrease the number of words with distinct meanings for the element(s) in question;
- **H4:** Textual representations with explicit elements will significantly decrease the number of words with distinct meanings for the other elements;

The first two hypotheses relate to syntax, as in [8]. The words "dog" and "cat" will be considered as totally distinct. The last two hypotheses relate to semantics, the words "dog" and "cat" will have some degree of similarity.

3. RELATED WORK

Furnas [8] tested the vocabulary problem in the context of command access, where a designer defines a name to a command and the user has to access it. However they argue that the problem might occur in any naming usage. They found that in average, 80 to 90% of people use different words to name things. Zhao [38] studied the vocabulary problem in information retrieval. They found that terms of queries do not appear on relevant documents to the user on an average of 40 to 50%. Which means that many documents that are relevant to the user will not be retrieved. In addition, many false positives will occur due to the vocabulary problem, which they call vocabulary mismatch. Similar observations have been made for ontology matching [6] and process model matching [35].

Concerning process models, other researches discussed aspects as qualities involved in the modeling. Krogstie [14] revised the Semiotic Quality Framework (SEQUAL), a framework about process models quality, and define quality in terms of physical quality, empirical quality, syntactic quality, semantic quality, perceived semantic quality, pragmatic quality, social quality and organizational quality. Our work relates to semantic, perceived semantic, pragmatic and social qualities. Mendling [19] suggests seven guidelines to model process models in order to improve quality. Mendling [18] discusses different labeling styles and their use in process modeling. Based on an experiment with process modelers, they suggest that the style verb-object is the most suitable for describing activity labels in process models.

Recker [27] analyses factors for model comprehension, namely model presentation and user characteristics. They conclude that different grammars have little influence on understanding, but user characteristics such as experience with a modeling grammar are important. Mendling and others [20] find that conceptual knowledge and process modeling expertise are important personal factors. As model factors, they found that - for syntactical comprehension - labels represented by words instead of letters (A, B or C) make the model understanding more difficult.

Pinggera [25] presents a study on informal specifications of process models. They argue that structuring the specification may improve the quality of the resultant process models. They tested three options of structuring: a depth-first description, a breadth-first description and a random description. Their findings suggest that depth-first and breadth-first descriptions indeed result in significantly better process models. No significant difference was observed between breadth-first and depth-first.

Important contributions have been done on quality of process models. However, the vocabulary problem in the context of process models is hardly understood. We believe that relating activity labels and words from ontologies can achieve good results in improving the semantics. Therefore, we test the vocabulary problem in the context of process models in order to check the need for using ontologies. Also, we test the effects of different textual representations (as requirements specifications) on labels created by modelers.

4. RESEARCH METHOD

The conducted experiment was twofold and intended to:

- Analyze the probability of two persons using the same word to represent a concept (vocabulary problem [8]);
- Analyze whether different levels of explicitness in the description of process model activities have influence in the words used to create the labels.

In order to test our hypotheses, we need to compare the labels created by different participants. To this end, we conducted an online experiment. Each person answered questions about demographics and represented activity labels based on different textual descriptions.

4.1 Population

The experiment intends to analyze novices of process modeling and was performed in a within-subjects form. The participants were students, which are described as adequate proxies for novice modelers [3]. We analyzed the answers of 22 participants, 12 master students, 8 bachelor students and 2 Ph.D. students from two computer science classes at Federal University of Rio Grande do Sul in Brazil.

4.2 Research Design

We considered three options for gathering activity labels in order to test our hypotheses:

1. Provide a requirements specification (textual) to participants. Ask them to model the process;
2. Remove labels of one or more activities from a process model. Ask the participants to complete the model, similar to a Cloze test [32];
3. Provide the participants with descriptions of specific activities and ask them to write the labels for each description.

Option 2 was not possible to our case because we wanted to test different levels of explicitness. In Option 1, we would have to select which activity from each participant should be compared with other participants activities. However, Option 3 allows for more control and would cause less bias. Also, for Option 1 we cannot ensure that all participants create the same activities, what could cause problems regarding pairing. Therefore, Option 3 was chosen. The task for the participants was thus to read a text which describes specific activities of a process model and to create labels. Each participant's answer was composed by a subject, a verb and an object. These elements are the basis for our dependent variables.

The experiment intends to test if explicit elements (our independent variable) and the interactions between elements decrease the vocabulary problem. Whether the subject is given or not, the verbs will have the same number of distinct words or it will decrease for the explicit subject. To that end, the experiment has five different levels of element explicitness:

- **E1:** none of the words necessary to represent the labels are explicit.
- **E2:** the subject was presented in the text, however the participant of the experiment was not told to use it. The rest was not made explicit.
- **E3:** the subject and the verb were presented in the text, however the participant of the experiment was not told to use them. The object was not made explicit.
- **E4:** subject was presented and the user was obliged to use it.
- **E5:** subject and verb were presented and the user was obliged to use it.

For clarification purposes regarding explicitness, we present examples of questions (Q1, Q2 and Q5) used in the experiment:

- **Q1:** This task occurs in the context of someone going to board on a flight. More specifically, when passing through security; Who will fly have to give access to things that can be carried on the plane for the security personnel, so that they can verify it. **P.S.** Describe the task performed by who will fly, in the context presented above.

- **Q2:** This task occurs in the context of someone ordering a book on an online store. More specifically, when the books seller informs things like address or preferential date to the entity that will take the book(s) to the buyer's address. **P.S.** Describe the task performed by the entity in the context presented above.

- **Q5:**This task occurs in the context of a user searching for information in the web. More specifically, when the website process what the user inserted to search. **P.S.** Describe the task performed by the website, in the context presented above.

Question 1 and 2 were part of level E1, thus none of the words presented in the label were given in the question. It's original labels were, for Q1: Passenger; Manage; Luggage. For Q2: Shipping Company; Receive; Shipping Manifest. Question 5 was part of level E3, thus subject and verb were explicitly presented in the question. It's original label was: Website; Process; Query.

4.3 Materials and tasks

The experiment was accessible via an online website. Each question was created based on an activity of a process model from an online collection[2]. The activities were chosen as follows: an algorithm selected models with 20 activities and copied each to a new folder, which received a random number from 0 to 10000 in the beginning of its name, to be randomly sorted. In addition, a file containing a random ranking from 0 to 20 was created to each model. The chosen activities were picked following the ranking and each model was verified as follows:

1) Are labels concrete (e.g. not A, B, C) and in English?
2) The model has Pools/Lanes?
3) Is the domain of common understanding and the model of good quality?
4) Is the task X with label style verb-object and it includes 2 or 3 words? If not, go to next.
   a) Models were read in depth-first fashion. When a gateway was found, the reading was done always from top-down, for the outgoings edges.
   b) If 3 first tasks are not good go to next model. A model can be of common understanding, but have activities that are not, skip them;

For step 3, we did not use models with specific terms, but followed 7PMG [19]. Ten activities plus subject (pool or lane) were selected. After that, a question for each activity was created, following the five levels of explicitness described above.

[2]http://bpmai.org/download/index.html

Each level of explicitness had two variants (questions) in order to decrease possible bias due to possible facets, language and domain issues, making a total of 10 questions. One activity may allow more variants than another. Some people may have more knowledge than others about a domain. Also, some words may have more synonyms or hyponyms. Thus, two questions for each level may decrease those differences. A larger number of questions per case would be better, however carryover effects may appear while the number of questions increases.

We performed a pre-test with 5 persons to verify if the questions were representing what they were meant to. Changes were performed based on that pre-test. In the experiment, before presenting the questions, an example of business process model was shown. After that, a description of the verb-object style was presented. We pointed out that subjects and objects are normally represented by nouns and have one or two words. Verbs are represented by one word. A table with five examples of proper and not proper labels, according with verb-object style, were presented. Also, one example of question with answers was shown. The questions were presented on the same page, thus the participant had access to the examples during the whole time. After answering a question, the next one was shown and the previous was hidden, so answers could not be changed. In the first 6 questions (E1, E2 and E3) the participant answered the subject, verb and object for the specific activity. For questions 7 and 8 (E4), the participant answered the verb and object, the subject was already filled. For questions 9 and 10 (E5) the participant had to answer only the object.

4.4 Data Cleansing

The questionnaire was answered by 65 persons. People who have used: "abc", "I don't know" or "123" in any of the questions were removed (7 persons). People that gave answers in Portuguese were removed (3 persons). We used the WordNet::Similarity to find problems in specific terms. Terms in plural form were changed to singular, so WordNet can find them. Prepositions were removed. In cases like: "noun A" of "noun B" or "noun A" of the "noun B", we changed it to "noun B" "noun A". This is the expected format and it was showed in the examples. At last, in order to increase internal validity [26], we used WordNet::Similarity to compare the labels of each person against the original labels, which originated the questions. Participants that achieved a score smaller than 0.5, comparing with the original labels, in any of the questions were removed (33 persons). After these steps, 22 participants remained and were used in the tests.

5. RESULTS

In approach A we compared the number of distinct words used (vocabulary problem) and in approach B we compared the meaning of words through similarity comparisons. For both approaches, we compared each element (subject, verb and object) of the label of a participant against the elements used by all other participants. As a result a matrix was generated and the inferior diagonal was used in the tests. All statistical tests were made with 95% of confidence.

First, we present snippets from our result set for clarification. We describe why we have two questions for each explicitness level. The data from Table 1 present how many people gave equal answers (pairwise) for subject and verb elements. We do not present data from E4 and E5 because subjects and verbs (E5) were given in the answers.

Table 1: Equal answers for subject and verb elements

		Subject		Verb	
	Question	**Count**	**%**	**Count**	**%**
	1	121	52	25	11
E1	2	14	6	36	16
	3	89	39	45	19
E2	4	231	100	42	18
	5	210	91	51	22
E3	6	155	67	80	35

For subjects, except for level E3, there was a big discrepancy between each question of the same level, the use of two questions decreases the issue. In the original labels, each explicitness level had a subject with one word and another with two words (by chance). The questions with lower scores are the ones with two words. This data can cause doubts regarding validity, thus Table 2 present snippets of answers from question 1 and 2.

The snippet of results presented in Table 2 provide an idea about why questions 1 and 2 have a big difference regarding equal answers. Question 2 presented a wider scope of alternatives for subject element. In the examples, both questions have one alternative that are arguably wrong: Security and Seller, respectively. In fact, those alternatives are not exactly wrong, they describe valid alternatives, however they are not the most adequate ones. The questions presented the direction that should be used to write the labels: passenger and shipping company. Thus, based in the bounded rationality, we could draft some possible causes

1) Lack of cognitive resources to understand the question;
2) Lack of time availability, so the participant did not read properly the description;
3) Lack of will, so the participant did not read carefully;
4) The participant did not consider answering the question as a valuable outcome, so he did not apply much effort.

Table 2: Examples of answers for E1

	Subject	Verb	Object
	Passenger	Show	Hand luggage
	Passenger	Authorize	Verification
	Security	Check	Baggage
	Customer	Show	Personal items
Q1	Passenger	Scan	Belongings
	Delivery Company	Receive	Client Data
	Carrier	Analyze	Buyer info
	Seller	Report	Data
	Delivery service	Deliver	Book
Q2	Mail Centre	Deliver	Book

Summarizing, many reasons could have led the participant to non proper answers; however, they are still valid in the context of the question. The rest of the subjects presented in Table 2 suffers from the vocabulary problem. For passenger, we have an example of "customer", which is not wrong from the company point of view. For question 2, the alternatives are not that similar as Passenger and Customer, however they are suitable for the case.

5.1 Results for approach A

Here we analyzed the probability of two persons to use the same terms to name each of the elements of a label. The results are binary, element X is either equal or not to element Y, compared by simply matching the strings. The probabilities can be seen in Table 3.

Table 3: Probabilities of using equal terms (pairwise)

		E1	E2	E3	E4	E5
	Subject	29%	69%	79%	100%	100%
	Verb	13%	19%	28%	14%	100%
Exact	**Object**	8%	6%	5%	2%	12%
	Subject	30%	83%	88%	100%	100%
	Verb	14%	19%	29%	14%	100%
Contains whole	**Object**	13%	18%	15%	6%	31%

We can see in Table 3 that the results are in line of Furnas [8] who found a probability of 10% to 20%. The elements with 100% should not be considered, because they were given and could not be changed by the participant. In the "Exact" set, the subjects from E2 and E3 had the higher decrease in the vocabulary problem. It might be due to the level of explicitness of terms for those elements in the questions.

The "Contains whole" set of Table 3 shows bigger probability increase for Subjects and Objects than for Verbs. The reason is because subjects and objects may have more than one word, so "confirmation message" and either "confirmation" or "message" would match. However, the individual words are less specific and might cause ambiguity in some cases. As the comparisons in this point result in binary values and we have paired within-subjects, we performed McNemar's test to verify statistical significance on the different explicitness levels. We tested whether a person uses the same word as another or not. Table 4 presents crosstabs and McNemar's test for subject element.

In Table 4 we can see that the number of subjects increased from distinct (Diff.) to equal (Eq.) when comparing E1 with both E2 and E3. The equal subjects increased in smaller range from E2 to E3. Subjects of E4 and E5 were given, so both are equal in all cases and thus were not used in the tests. The table at the lower right corner present results of McNemar's test, which shows that the differences presented in the crosstabs are significant from level E1 to both E2 and E3, which had the subjects explicitly presented in their questions. Comparing levels E2 and E3, which had the same level of explicitness for subject, we have significant difference as well. However, we believe that it does not affect H1. The result supports H1 for subjects, Figure 1 clarifies this support.

Table 4: Crosstabs and McNemar's test for Subject element

	E2	
E1	**Diff.**	**Eq.**
Diff.	61	266
Eq.	81	54

	E3	
E1	**Diff.**	**Eq.**
Diff.	79	248
Eq.	18	117

	E3	
E2	**Diff.**	**Eq.**
Diff.	9	133
Eq.	88	232

	P-value
E1-E2	0.000[a]
E1-E3	0.000[a]
E2-E3	0.003[a]

a. Binomial distr.

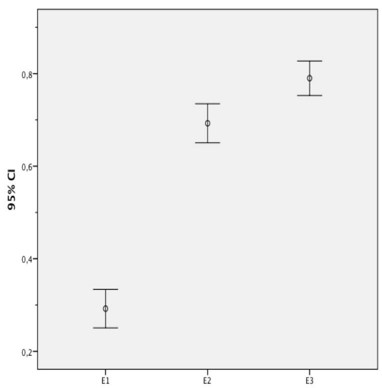

Figure 1: Error bar for subject element.

Table 5 presents crosstabs and McNemar's test for verb elements. Level E3 is the only level with explicit verbs in its questions, in all comparisons the results show significant difference from E3 against the other levels. These results support H1 for verbs. However, the other verbs do not show significant difference, except for E1 vs. E2, thus we reject H2 for verbs.

Table 6 presents the frequencies for objects. They were expected to vary as much as other elements. However, they present a descendent distribution instead of ascendent, except for E5 which had the greater mean, which was expected. However, as the other elements do not follow the expected behaviour, we reject H2 for objects. In fact, the interaction between label elements (H2) seems to follow the Zipf distribution, at least for the most used term of each element in the label. It seems that the explicitness helps up until a certain point and then the bounded rationality takes control. Table 7 shows the distribution on our data.

In all our tests, the Zipf distribution was more similar to level E2, which makes sense and reinforces the effects of explicitness. We are unable to compare levels E4 and E5 because they have given elements and have the highest possible scores for the subject. The terms for each element, however, do not seem to follow Zipf's distribution according to our data.

5.2 Results for approach B

The hypotheses H3 and H4 request semantic measures, thus we used the software package WordNet::Similarity [24].

Table 5: Crosstabs and McNemar's test for Verb element

	E2	
E1	**Diff.**	**Eq.**
Diff.	326	75
Eq.	49	12

	E3	
E1	**Diff.**	**Eq.**
Diff.	284	117
Eq.	47	14

	E4	
E1	**Diff.**	**Eq.**
Diff.	343	58
Eq.	54	7

	E3	
E2	**Diff.**	**Eq.**
Diff.	276	99
Eq.	55	32

	E4	
E2	**Diff.**	**Eq.**
Diff.	325	50
Eq.	72	15

	E4	
E3	**Diff.**	**Eq.**
Diff.	281	50
Eq.	116	15

	McNemar (p-value)
E1 vs. **E2**	0.024[a]
E1 vs. **E3**	0.000[a]
E1 vs. **E4**	0.777[a]
E2 vs. **E3**	0.000[a]
E2 vs. **E4**	0.057[a]
E3 vs. **E4**	0.000[a]

a. Binomial distribution used.

Table 6: Frequencies for object element

	E1	E2	E3	E4	E5
Mean	0.080	0.065	0.041	0.017	0.123
SD	0.272	0.247	0.199	0.131	0.329

Table 7: Zipf distribution for elements of an activity label

	E1	Zipf	E2	Zipf	E3	Zipf
Subject	20.0	20.0	35.0	35.0	39.0	39.0
Verb	15.0	10.0	16.0	17.5	23.0	19.5
Object	10.0	6.6	10.0	11.6	8.0	13.0

Which are Perl modules that allow measuring the similarity and relatedness of two terms. We used the WUP measure [37], however, other measures can be used instead. It is a similarity measure that supports us in checking whether persons use equal words or not. WUP yields a result in the range from 0 to 1, which is based on path length up to the least common subsumer (LCS) and the root node. Note the difference to the vocabulary problem. In the vocabulary problem, two synonyms are a problem, since they are represented by different words. Here, the WordNet::Similarity will give a high score for two synonyms. Thus, we are measuring not only syntactically equal words, but also similar meanings. An important point is that, as we are using only labels that are at least 50% similar with the original labels, the range for differences decreases significantly.

The expected number of words was one or two for subjects and objects and one for verbs. This was made explicit in the examples and followed by most part of the participants. Based on that premises, the similarity comparison between each element had four possibilities:

- **One word vs. one word:** WA1 vs. WB1;
- **Two words vs. two words:** ((WA1 vs. WB1) + (WA2 vs. WB2))/2;
- **One word vs. two words:** WA1 vs. WB1 and WA1 vs. WB2, get the higher and divide by two;
- **Two words vs. one word:** WA1 vs. WB1 and WA2 vs. WB1, get the higher and divide by two;

We compared subjects and objects as nouns and verb as verb. Because a word with verb as part of speech and another with noun can have a high similarity, so we cannot compare all parts of speech to search for the highest similarity. An example is "show" and "check" both can be used as noun or verb and with the measure WUP the higher similarity is between noun and verb.

As the experiment was performed in a with-in fashion and we have the same persons answering each question, a pairwise t-test was used to compare the means. In this approach we also consider a mean of the label elements, resulting in a value from 0 to 1. This works because, differently from approach A, the values are continuous. The results for elements mean and for subjects are presented in Table 8.

Table 8: Paired Sample T Test for label mean and for subject element

		E1-E2	E1-E3	E2-E3
	Mean	0.000	0.000	0.351
P-Value	**Subject**	0.000	0.000	0.001

The results suggest that presenting explicitly the subject as in E2 and E3 and the verb in E3 results in higher similarity of labels as comparing to E1, which presented all elements in an abstract way. This supports partially H3. Also, data suggest that there is no significant difference between E2 and E3, thus presenting the subject and verb (E3) explicitly do not affect the label against presenting only the subject (E2), rejecting partially H3.

The results for subjects suggest that there is a significant difference in E1 vs. E2 and E3. Thus, explicit subject (E2 and E3) cause significant difference against not explicit (E1). Also, E2 vs. E3 - both have explicit subjects in their questions - have significant difference in the tests but have similar means 0.897 and 0.938, respectively, against 0.699 of E1. Therefore, this supports H3 for subjects. Figure 2 clarifies this support.

The elements Verb and Object did not follow the expected distribution. Table 9 presents the descriptive statistics.

For Verbs, E3, which has explicit verbs in their questions, had the lowest result. All verbs have practically equals means, potentially due to our cut in data cleansing. Thus, we reject H3 for verbs. In addition, E4 is very similar to E1, thus we reject H4 for verbs. As the results were not in the

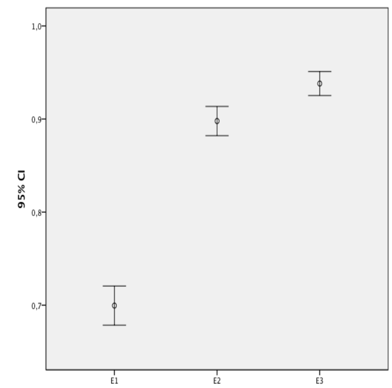

Figure 2: Error bar for subject element.

Table 9: Descriptive statistics for Verb and Object elements

		E1	E2	E3	E4	E5
	Mean	0.664	0.663	0.630	0.643	1.0
Verb	**SD**	0.218	0.223	0.276	0.238	0.0
	Mean	0.675	0.618	0.633	0.526	0.722
Object	**SD**	0.198	0.207	0.191	0.242	0.175

expected distribution, we did not present a pairwise t-test for this element. For Object elements, Table 9 shows unexpected results as well, in particular E1 has no explicitness and is equal or higher than E2, E3 and E4. Thus, we did not present a pairwise t-test for this element. We reject H4 for object element.

5.3 Discussion

It is important to note that the scope of process modeling was narrowed down considerably for this experiment. First, the use of verb-object style plus subject element had to be followed by participants. Then, we removed any answer that did not achieve 50% of similarity with the seed labels. All this is necessary for the experiment, but we believe that in this sense the problem is much wider in real world scenarios. The difference of probabilities presented in Table 3 is very big for the presented case.

Our study suggests that the vocabulary problem indeed occurs in process models with novice modelers. The explicit presentation of label elements in specifications seems to decrease the vocabulary problem for subject and verb elements, specially for subject elements. However, the subject is used to represent participants in a process model, by means of a Lane or Pool. The elements actually presented in activity labels are verb and object, which were less affected by the different levels of explicitness provided in our experiment. In fact, the triplet elements (subject, verb, object) seem to follow Zipf's distribution.

Regarding element interaction, our hypotheses were rejected for all cases. This means that having an anchor does not affect the other elements. If the subject is given, it does not affect the verbs chosen by persons. If the subject and verb are given, it does not affect the objects. The vocabulary problem still occurs, a possible view is that the bounded rationality is stronger than what we called element interaction.

Therefore, the vocabulary problem has to be considered regarding process models quality. Based on our discussion in Section 2 altogether with the results of the experiment, we believe that ontologies should be used to provide a shared vocabulary to avoid this problem. Using ontologies it is possible to use different symbols (words) to refer to a concept and still keep track of the original symbol used to represent that concept. To study this matter further, we plan to perform experiments with ontologies in the future.

Our study suggests that explicit information in the requirements results in a higher similarity of terms. It does not solve the vocabulary problem, but the terms appear to be more similar, which might improve models quality. In this way, at least the subjects - Roles and Lanes - should be presented explicitly in process model specifications. A consistent vocabulary, using the same term every time a concept appears in the whole specification might help as well.

The vocabulary problem may also affect the retrieval of process models for reuse. Besides false positives and missing relevant documents, it may cause error propagation when the reused models are not appropriate for the cases. This could lead to many problems, such as the anchoring and adjustment effect [1], which needs to be studied in the future.

5.4 Threats to validity

Regarding the experiment, we understand that a drawback is the low number of participants and a wider experiment should be done. We see two ways of doing it: (i) compare words of person X against person Y or (ii) compare words of person X against all other participants. In (i) selecting pairs may cause a bias and the sample needs to be quite big to alleviate that. In (ii) a smaller number of participants allows having bigger sample. We performed option (ii), where 22 participants with two questions per explicit level made a total of 462 pairs. In addition, our results are quite similar to results presented in literature for other domains. Due to the ratio of (ii), the number of pairs can grow fast and cause problems for statistical tests. P-values calculated from very large samples may result zero and support for results may not have practical significance [17].

Our tests related only to novice modelers. We assumed the verb-object style for internal validity. However, in real world scenarios, people might not use this style as a standard, which may increase the vocabulary problem for process models. In WordNet, each word can have many senses. To not interfere in the measurements, we got the higher value between two words, since we could not choose the senses manually. Thus, some words could have been related to a different sense from what it should. We set the part of speech (e.g. noun or verb) and the context should be similar due to questions, thus we believe that it would not affect the results strongly.

6. CONCLUSIONS

The documentation of business process by means of models is an important task for companies in practice. In this paper, we have presented conceptual discussions about the use of ontologies to support this modeling. In addition, we have presented experimental results concerning the vocabulary problem in the context of business process models. Both discussions are in line with the growing research interest in process models quality.

Our findings suggest, based on empirical data, that the vocabulary problem indeed exists for process models created by novice modelers. Solutions based purely on vocabulary do not seem to be enough to solve the problem. The interaction between elements does not solve the problem either. In fact, elements seem follow Zipf's distribution in this regard. However, we see the need to present explicitly at least the subjects of activities in process specifications, which might improve the quality of process models created by novices. Yet, requirements specifications are not always available and might have many different representations.

In order to solve the problem, ontologies might be considered to provide terminological support for process modeling. Concerning the bounded rationality, if well applied, ontologies can help to improve the nature of the environment, which might consequently improve the modeler's performance. We have discussed ontologies and aspects that could be related with the vocabulary problem in this paper as well. In future work we plan to investigate other aspects such as personal factors and the use of ontologies in this context by means of empirical research. Our work is part of a recent effort to improve process models quality, in different levels, which would consequently improve companies information systems.

7. ACKNOWLEDGMENTS

This work has been partially supported by Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) and Coordenação de Aperfeiçoamento de Pessoal (CAPES), Brazil.

8. REFERENCES

[1] G. Allen and J. Parsons. Is query reuse potentially harmful? anchoring and adjustment in adapting existing database queries. *Info. Sys. Research*, 21(1):56–77, Mar. 2010.

[2] G. Box and N. Draper. *Empirical model-building and response surfaces*. Wiley series in probability and mathematical statistics: Applied probability and statistics. Wiley, 1987.

[3] A. Burton-Jones and P. Meso. The effects of decomposition quality and multiple forms of information on novices' understanding of a domain from a conceptual model. *J. AIS*, 9(12), 2008.

[4] I. Davies, P. Green, M. Rosemann, M. Indulska, and S. Gallo. How do practitioners use conceptual modeling in practice? *Data Knowl. Eng.*, 58:358–380, September 2006.

[5] A. Dreiling, M. Rosemann, W. M. P. van der Aalst, and W. Sadiq. From conceptual process models to running systems: A holistic approach for the configuration of enterprise system processes. *Decis. Support Syst.*, 45(2):189–207, may 2008.

[6] J. Euzenat and P. Shvaiko. *Ontology Matching, Second Edition*. Springer, 2013.

[7] P. Fettke, A.-L. Vella, and P. Loos. From measuring the quality of labels in process models to a discourse on process model quality: A case study. HICSS '12, pages 197–206, Washington, DC, USA, 2012. IEEE Computer Society.

[8] G. W. Furnas, T. K. Landauer, L. M. Gomez, and S. T. Dumais. The vocabulary problem in

human-system communication. *Commun. ACM*, 30(11):964–971, Nov. 1987.

[9] J. B. Gassen, A. Machado, L. H. Thom, and J. P. M. de Oliveira. Ontology support for home care process design. In L. A. Maciaszek, A. Cuzzocrea, and J. Cordeiro, editors, *ICEIS (2)*, pages 84–89. SciTePress, 2012.

[10] G. Gigerenzer and R. Selten. *Bounded Rationality: The Adaptive Toolbox*. Dahlem Konferenzen. MIT Press, 2002.

[11] N. Guarino. *Formal Ontology in Information Systems: Proceedings of the 1st International Conference June 6-8, 1998, Trento, Italy*. IOS Press, Amsterdam, The Netherlands, The Netherlands, 1st edition, 1998.

[12] M. Indulska, P. Green, J. C. Recker, and M. Rosemann. Business process modeling : perceived benefits. 2009.

[13] B. D. Jones. Bounded rationality. *Annual Review of Political Science*, 2(1):297–321, 1999.

[14] J. Krogstie, G. Sindre, and H. Jørgensen. Process models representing knowledge for action: a revised quality framework. *Eur. J. Inf. Syst.*, 15(1):91–102, Feb. 2006.

[15] H. Leopold, R.-H. Eid-Sabbagh, J. Mendling, L. G. Azevedo, and F. A. Baião. Detection of naming convention violations in process models for different languages. *Decision Support Systems*, 56:310–325, 2013.

[16] H. Leopold, J. Mendling, H. A. Reijers, and M. L. Rosa. Simplifying process model abstraction: Techniques for generating model names. *Inf. Syst.*, 39:134–151, 2014.

[17] M. Lin, H. C. Lucas, and G. Shmueli. Research commentary—too big to fail: Large samples and the p-value problem. *Information Systems Research*, 24(4):906–917, 2013.

[18] J. Mendling, H. Reijers, and J. Recker. Activity labeling in process modeling: Empirical insights and recommendations. *Information Systems*, 35(4), 2010.

[19] J. Mendling, H. A. Reijers, and W. M. P. van der Aalst. Seven process modeling guidelines (7pmg). *Inf. Softw. Technol.*, 52(2):127–136, Feb. 2010.

[20] J. Mendling, M. Strembeck, and J. Recker. Factors of process model comprehension-findings from a series of experiments. *Decis. Support Syst.*, 53(1):195–206, Apr. 2012.

[21] D. Moody. The physics of notations: Toward a scientific basis for constructing visual notations in software engineering. *IEEE Trans. Softw. Eng.*, 35(6):756–779, Nov. 2009.

[22] D. A. Norman and S. W. Draper. *User Centered System Design; New Perspectives on Human-Computer Interaction*. L. Erlbaum Associates Inc., Hillsdale, NJ, USA, 1986.

[23] C. Ogden and I. RICHARDS. *Meaning of Meaning: A Study of the Influence of Language Upon Thought and of the Science of Symbolism*. New York, 1923.

[24] T. Pedersen, S. Patwardhan, and J. Michelizzi. Wordnet::similarity: Measuring the relatedness of concepts. In *Demonstration Papers at HLT-NAACL 2004*, HLT-NAACL–Demonstrations '04, pages 38–41, Stroudsburg, PA, USA, 2004. Association for Computational Linguistics.

[25] J. Pinggera, S. Zugal, B. Weber, D. Fahland, M. Weidlich, J. Mendling, and H. A. Reijers. How the structuring of domain knowledge helps casual process modelers. In *Proceedings of the 29th International Conference on Conceptual Modeling*, ER'10, pages 445–451, Berlin, Heidelberg, 2010. Springer-Verlag.

[26] J. Recker. *Scientific Research in Information Systems: A Beginner's Guide*. Progress in IS. Springer, 2012.

[27] J. C. Recker and A. Dreiling. The effects of content presentation format and user characteristics on novice developers' understanding of process models. *Communications of the Association for Information Systems*, 28(6):65–84, February 2011.

[28] J. C. Recker, M. Indulska, M. Rosemann, and P. Green. The ontological deficiencies of process modeling in practice. *European Journal of Information Systems*, 19(5):501–525, September 2010.

[29] H. Simon. *Models of man: social and rational; mathematical essays on rational human behavior in society setting*. Wiley, 1957.

[30] R. Studer, V. R. Benjamins, and D. Fensel. Knowledge engineering: Principles and methods, 1998.

[31] J. Sweller. Cognitive load theory, learning difficulty, and instructional design. *Learning and Instruction*, 4(4):295 – 312, 1994.

[32] W. Taylor. Cloze Procedure: A New Tool for Measuring Readability. *Journalism Quarterly*, 30:415–433, 1953.

[33] L. Thom, M. Reichert, and C. Iochpe. Activity patterns in process-aware information systems: Basic concepts and empirical evidence. *International Journal of Business Process Integration and Management (IJBPIM)*, 4(2):93–110, 2009.

[34] I. Weber, J. Hoffmann, and J. Mendling. Beyond soundness: on the verification of semantic business process models. *Distributed and Parallel Databases*, 27(3):271–343, 2010.

[35] M. Weidlich, R. M. Dijkman, and J. Mendling. The icop framework: Identification of correspondences between process models. In B. Pernici, editor, *Advanced Information Systems Engineering, 22nd International Conference, CAiSE 2010, Hammamet, Tunisia, June 7-9, 2010. Proceedings*, volume 6051 of *Lecture Notes in Computer Science*, pages 483–498. Springer, 2010.

[36] D. A. Wiegmann, D. F. Dansereau, E. C. McCagg, K. L. Rewey, and U. Pitre. Effects of knowledge map characteristics on information processing. *Contemporary Educational Psychology*, 17(2):136 – 155, 1992.

[37] Z. Wu and M. Palmer. Verbs semantics and lexical selection. In *Proceedings of the 32Nd Annual Meeting on Association for Computational Linguistics*, ACL '94, pages 133–138, Stroudsburg, PA, USA, 1994. Association for Computational Linguistics.

[38] L. Zhao and J. Callan. Term necessity prediction. In *Proceedings of the 19th ACM International Conference on Information and Knowledge Management*, CIKM '10, pages 259–268, New York, NY, USA, 2010. ACM.