



## Overspecified references: An experiment on lexical acquisition in a virtual environment



Alexandra Luccioni<sup>a,\*</sup>, Luciana Benotti<sup>a</sup>, Frédéric Landragin<sup>b</sup>

<sup>a</sup>Logic, Interaction and Intelligent Systems Group, FAMAf, Universidad Nacional de Córdoba/CONICET, Medina Allende s/n, Ciudad Universitaria, Córdoba CP 5000, Argentina

<sup>b</sup>LaTITCe Laboratory, ENS, UMR 8094, 1, Rue Maurice Arnoux, F-92120 Montrouge, France

### ARTICLE INFO

#### Article history:

Available online 12 March 2015

#### Keywords:

Referring expression generation  
Overspecification  
Lexical acquisition  
Virtual worlds  
Computer Assisted Language Learning

### ABSTRACT

This study explores the role of quantity of information on vocabulary acquisition in a virtual world. Previous studies have shown that, although it makes interpretation more lengthy, speakers include more information in their referring expressions than is strictly necessary to identify an object – they *overspecify*. We aim to study the impact of this kind of overspecification on the acquisition of new lexemes in a foreign language. Our hypothesis is that using overspecified expressions during the practice of recently acquired vocabulary will help learners to better remember the lexemes and to exploit them more efficiently later on. In this paper, we describe an experimental study designed to evaluate this hypothesis, comparing two groups of learners who received overspecified and minimally specified referring expressions while practising newly acquired lexemes in the context of a language learning game in Russian. The game is situated in a virtual environment and the interaction is similar to that of a video game. Our results, based on experimental data from participants' performance as well as a post-experiment questionnaire, support the claim that overspecification improves lexical acquisition rates compared to minimal specification. Some pedagogical suggestions are provided for the design of referring expression generation algorithms in Technology-Enhanced Language Learning (TELL) and Computer Assisted Language Learning (CALL) Systems.

© 2015 Elsevier Ltd. All rights reserved.

### 1. Introduction

In our increasingly connected and internationalized world, learning a foreign language at an adult age is something that more and more people are facing. However, due to the number of learners worldwide, it is difficult to provide personalized, one-on-one teaching to all of them. Given the potential of technology to enhance and improve education in general, it is to be expected for learners to turn to CALL (Computer Assisted Language Learning) to learn languages (Dickinson & Herring, 2008; Johnson, 2007; Zhao, 2003; Sørensen & Meyer, 2007). This brings new questions as to the characteristics that CALL software should have in order to provide the most efficient learning experience.

Language acquisition is a complex process, with many variables that can affect a learner's progress on his or her unique learning path. In terms of vocabulary acquisition, the quantity of new

lexemes to be provided at once during the language acquisition process is the subject of much debate (Hulstijn & Laufer, 2001; Sagarra & Alba, 2006). We propose that instead of teaching isolated lexemes or minimally specified expressions, a CALL system should provide learners with overspecified referring expressions (REs). This is based on previous research in the field of Referring Expression Generation (REG), which has shown that, although it makes interpretation more costly, redundant information is frequently used in the REs produced by speakers (Engelhardt, Bailey, & Ferreira, 2006; Pechmann, 1989; Wu & Keysar, 2007). For example, Viethen and Dale (2011) found that, when identifying a big red ball sitting next to a big red cube in Fig. 1, most people use the referring expression “*the big red ball*”. They do so even though there are no small red balls in the figure, and “*the red ball*” provides enough information to identify the target object. We say that the RE “*the big red ball*” is *overspecified* because it includes more information than is necessary to univocally distinguish the referent.

In the literature, there are two explanations that have been proposed as contradictory for the overspecification phenomenon. Explanation (1) holds that overspecification impairs communication and is a result of a cognitive limitation of the speakers. Explanation (2) argues that overspecification is a useful part of

\* Corresponding author at: Laboratoire de Gestion, Diffusion et Acquisition de Connaissances, Université du Québec à Montréal, PK-4285, 201 Avenue du Président-Kennedy, H2X 3Y7 Montréal, Canada. Tel.: +1 514 436 4865.

E-mail addresses: [alexandra.vorobyova@gmail.com](mailto:alexandra.vorobyova@gmail.com) (A. Luccioni), [benotti@famaf.unc.edu.ar](mailto:benotti@famaf.unc.edu.ar) (L. Benotti), [frederic.landragin@ens.fr](mailto:frederic.landragin@ens.fr) (F. Landragin).

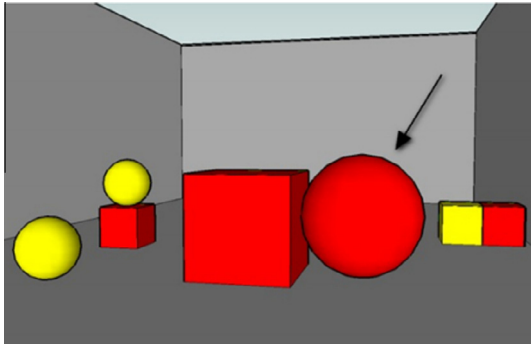


Fig. 1. An example of a visual scene with multiple objects. Source: Viethen and Dale (2011).

communication because it encourages lexical alignment between the speaker and the listener. Below we discuss the cognitive mechanisms for both explanations in turn.

Explanation (1) holds that the cognitive effort of producing non-redundant (i.e., minimal) referring expressions is too high (Pechmann, 1989). Pechmann's *Principle of least effort* suggests that language production is an incremental process and that it is not cognitively efficient for speakers to evaluate the entirety of their surroundings before producing a minimal specification. Waiting for such a complete representation of the context can be too costly for the speaker and can result in postponing the completion of language production processes. On the contrary, including salient and redundant properties in a referential expression as they come to mind is less cognitively demanding than explicitly verifying that they are redundant in order to omit them. Experimental results supporting this theory see overspecification as a result of a cognitive limitation of speakers, which impairs listeners' comprehension (Engelhardt, Demiral, & Ferreira, 2011). Engelhardt et al. found that listeners take longer to process overspecified referring expressions in comparison to minimal ones and concluded that listeners are confused and their comprehension impaired by overspecification. To the best of our knowledge, no empirical data exists showing that it takes longer to produce minimal referring expressions compared to overspecified ones.

Explanation (2) claims that overspecification is a useful part of communication because it allows speaker and listener to align by increasing their *common ground*, that is, the information they share regarding their environment (Nadig & Sedivy, 2002). As a result, future communication becomes more effective (Wu & Keysar, 2007). This hypothesis is also supported by empirical data: Clark and Wilkes-Gibbs (1986) have found that both speaker and listener use redundant properties in early references to an object, and reuse these properties in later references to the same object or to other objects with the same properties. This phenomenon can be perceived as a mechanism aiming to provide the listener with more information than is strictly necessary in order to compensate for potential perceptual difficulties or interference, as well as to coordinate on which lexemes are suitable for describing controversial physical properties. For example, a speaker may refer to a particular jacket as red using the overspecified RE "the red jacket on the sofa" when this is the only jacket on the sofa. Even if his listener thinks that the jacket is pink, she will nonetheless be able to resolve the RE because the property red is redundant. Later on, the listener may begin to refer to it as red herself in order facilitate communication.

Contrary to what has been argued in previous work, we believe that explanations (1) and (2) are not contradictory. They differ in how they study communication. Explanation (1) analyzes only single turns between speaker and listener while explanation

(2) consider multiple turns. In the short term, any delay to the listeners interpretation such as the one found by Engelhardt et al. (2011) can be seen as detrimental for communication. However, in the long run, overspecification leads to more shared information, that is, to more lexical alignment between speaker and listener. This in turn, leads to more effective communication. We believe that communication is a multi-turn process and therefore we adhere to explanation (2).<sup>1</sup>

We see second language lexical acquisition as an extreme case of alignment, since learners are obliged to use the same vocabulary and grammatical constructions as their teachers due to the limited choice of linguistic resources that they can use (Atkinson, Churchill, Nishino, & Okada, 2007). A study such as ours, focusing on lexical acquisition, permits us to control the speakers knowledge of a language and therefore to target the 'square one' from where alignment starts.

In this paper, we aim to support explanation (2) by empirically evaluating the effect of overspecification on lexical acquisition in second language (L2) learning. To this end, we created a system that produces minimal and overspecified REs describing objects in a 3-dimensional virtual world. We divided the participants into two groups: one group received minimal REs and the second group received overspecified REs during the practice phase following an initial exposure to new lexemes. We aim to see if, by giving overspecified REs during the practice period of lexical acquisition, the effectiveness of vocabulary learning increases. If this is the case, it has theoretical and practical implications. On the theoretical side, it provides evidence to support explanation (2), which goes to show that overspecification is useful for communication. On the practical side, it has a direct application in the design of referring expression generation algorithms for Computer-Assisted Language Learning (CALL): the descriptions used during lexical acquisition should be overspecified in order to increase lexical acquisition rates.

The rest of the paper proceeds as follows. In Section 2, we describe the various practical aspects of our experiment: the participants, the virtual world used, and the procedure. We present both the objective and subjective results of the experiment in Section 3, specifying the metrics we looked at and their pertinence to evaluating our hypothesis. The discussion of the results follows in Section 4, where we compare the results of the two groups of participants studied. Finally, Section 5 concludes and discusses potential applications to expert systems useful for CALL.

## 2. Materials and methods

We designed an experiment that uses a virtual environment setting, screenshots of which are shown in Figs. 2 and 3, and the map of the environment can be found in Appendix A. We programmed an instruction-giving system to guide participants in the world and permit them to acquire new vocabulary words. By comparing participants who practised new words with minimal and overspecified referring expressions, we aim to observe the utility of overspecification for L2 lexical acquisition.

### 2.1. The participants

We recruited 50 participants (15 women and 35 men) by sending out an invitation to test a language learning game. All of the participants were university students from the Universidad Nacional de Cordoba in Argentina. The average age was 28. All of the participants were native Spanish speakers who had no prior

<sup>1</sup> Communication is a multi-turn process both in speech-based interactions and in text-based interactions such, as the chat used in our experiment.

Download English Version:

<https://daneshyari.com/en/article/350307>

Download Persian Version:

<https://daneshyari.com/article/350307>

[Daneshyari.com](https://daneshyari.com)