In: Proceedings of PRE-CogSci 2009, 29 July 2009, Amsterdam, The Netherlands

Salience and Pointing in Multimodal Reference

Paul Piwek (P.Piwek@open.ac.uk)
Centre for Research in Computing, The Open University, Walton Hall
Milton Keynes, MK7 6AA United Kingdom

Abstract

Pointing combined with verbal referring is one of the most paradigmatic human multimodal behaviours. The aim of this paper is foundational: to uncover the central notions that are required for a computational model of human-generated multimodal referring acts. The paper draws on existing work on the generation of referring expressions and shows that in order to extend that work with pointing, the notion of salience needs to play a pivotal role. The paper investigates the role of salience in the generation of referring expressions and introduces a distinction between two opposing approaches: salience-first and salience-last accounts. The paper then argues that these differ not only in computational efficiency, as has been pointed out previously, but also lead to incompatible empirical predictions. The second half of the paper shows how a salience-first account nicely meshes with a range of existing empirical findings on multimodal reference. A novel account of the circumstances under which speakers choose to point is proposed that directly links salience with pointing. Finally, a multi-dimensional model of salience is proposed to flesh this model out.

Keywords: Generation of Referring Expressions; Multimodal Reference; Salience; Pointing Gestures; Deixis.

Introduction

Researchers on human pointing gestures have observed that pointing is essentially a means to "reorient the attention of another person so that an object becomes the shared focus of attention" (G. Butterworth, 2003). Somewhat surprisingly, this insight does not seem to have a counterpart in computational models of multimodal referring expression generation. In these accounts, *focus of attention*, *accessibility* and *salience*, three notions whose interrelationships we examine in more detail in the next section, are absent. Pointing is treated as either a fallback strategy for when verbal means fall short, or as expressing a property (i.e., as denoting a set of objects) in the same way that words, such as 'red' or 'bird', express properties.

For example, Lester, Voerman, Towns, and Callaway (1999) describe a system that only produces a pointing act, when a pronoun does not suffice to identify the target. Similarly, Claassen (1992) introduces an algorithm which only uses pointing if no purely verbal means of identification is possible, and Sluis and Krahmer (2001) describe an algorithm that only generates a pointing act if a purely verbal referring act becomes too complex. More recently, Krahmer and Sluis (2003) treat pointing acts not very different from words: as expressing a property. A pointing act identifies a subset of objects in the domain. Their algorithm assigns costs to the properties that are included in a referring expression. A graph-based algorithm is employed to find the cheapest combination of properties for referring to an object.

This is not to say that none of the models of referring expression generation and interpretation use notions such as attention, accessibility, or salience – a notion that will occupy a central place in the model that is offered in this paper. For example, visual salience plays a pivotal role in the interpretation and generation algorithms of Kelleher, Costello, and Genabith (2005). Similarly, Choumane and Siroux (2008) model visual salience for interpretation. Neither of these accounts do, however, directly relate salience to pointing gestures: Kelleher et al. (2005) only deals with verbal referring acts, whereas Choumane and Siroux (2008) view pointing acts rather narrowly as designating an object, rather than playing the dynamic role of changing the focus of attention.

The aim of this paper is to unpick the relation between salience and pointing and lay the foundations for a computational account based on this relation. The next section makes the assumptions behind the current approach explicit, and spells out the relation between the notions of salience, accessibility and focus of attention. Next, the role of salience in the generation of referring expressions is examined. We distinguish between two opposing approaches for dealing with salience: salience-first and salience-last accounts, and argue that these differ not only in computational efficiency, as has been pointed out previously, but also lead to diverging empirical predictions. The second half of the paper shows how a salience-first account nicely meshes with a range of existing empirical findings on multimodal reference. A novel account is put forward of the circumstances under which speakers choose to point. This account directly links salience with pointing. Finally, it is fleshed out by introducing a multi-dimensional model of salience for multimodal reference.

Assumptions and Terminology

The situations that we aim to model have three main ingredients: a speaker, an addressee and a visually shared domain of discourse. The speaker's goal (or intention) is to identify an object, the target, for the addressee in the domain of discourse. To achieve this goal, the speaker can use both language and pointing gestures. The scope of the model is restricted to cases in which the speaker is referring to objects in the visually shared domain and, if the speaker points, the target is among the objects that the speaker points at. This excludes cases such as those discussed by Clark, Schreuder, and Buttrick (1983) and Goodwin (2003). For example, Clark et al. (1983) discuss a speaker who says 'I worked for those people' whilst pointing at a newspaper. In this instance, the speaker referred to the publishers of the newspaper. Cases like this one, where the speaker refers to an object that is not in the visually shared domain and points at an object which is different from the target, are beyond the scope of the current study.