Modelling Synchrony and Diachrony of DOM in Stochastic OT

Romance languages exhibit differential object marking (DOM), whereby animate and specific objects are marked by what is known in the Romance literature as the prepositional accusative *a* (*pe* in Romanian) while others stay unmarked. For example in (1), from Corsican, we see the presence of *a* on the specific noun phrase and the absence of the marker on the non-specific one.

(1)   Prete Sartoli cuniscia l'   omi in ginerale e   **a**   Ziu   Don Salto in
       priest sartoli knew   the man in general and OBJ uncle Don Salto in
       particulare.
       particular
       'Priest Sartoli knew men in general and uncle Don Salto in particular.'

Previous analyses resort to a functional explanation in terms of Markedness Reversal (cf. Comrie 1989, Aissen 2003): what is unmarked for prototypical subjects (e.g. animate, definite, and specific) is marked for objects, and such marked, "subject-like" objects must be morphosyntactically marked (for disambiguation). Such an analysis makes the prediction that only those objects that are more like prototypical subjects will receive object marking. While this may be true for other DOM languages like Turkish, it is not sufficient to explain the full range of facts in Romance languages: in Romance we not only find object marking for those objects that resemble prototypical subjects, but also variably for those that are low in animacy and specificity, as illustrated in (2) from De Jong (1996).

(2)   a.   El entusiasmo vence   (a)   la difficultad.
            the enthusiasm conquer.3SG (OBJ) the difficulty
            'Enthusiasm conquers difficulties.'
       b.   A   la difficultad vence   el   entusiasmo.
            OBJ the difficulty   conquer.3SG the enthusiasm
            'Enthusiasm conquers difficulties.'

To explain these and similar facts De Hoop and Narasimhan (to appear) and De Swart (2003, to appear), argue to view DOM (in Romance) as a dynamic process in which not only the properties of the object but rather the properties of the whole transitive clause are taken into consideration, in line with the classical notion of transitivity put forward by Hopper and Thompson (1980). In this view DOM is motivated by the principle of 'Distinguishability'— the two participants of a transitive clause must be distinguishable. In case the two participants are not distinguishable because they share animacy and specificity features, the object

has to be marked.

Further support for this view of DOM comes from historical data in Brazilian Portuguese (BP). According to Tarallo (1996), from 1825 to 1981 the use of overt 3rd person DO pronouns in BP decreased from 81% to 18% (60% in a mid-point 1880), giving rise to a DOM system in contemporary BP in which overt DO pronouns are used only for animate and specific referents (Schwenter and Silva 2003). Conversely, in the same time span the use of overt subject pronouns *increased* from 16% (in 1825) to 79% (in 1981; 33% in 1880). These figures show an intricate interplay between subjects and objects: the concomitant, otherwise unnatural, shift from a null to overt subject language must have developed as a means of disambiguating grammatical functions, as the language shifted from an overt to a null object language. Moreover, the development of DOM in BP also resulted in the concomitant loss of VS word order (Berlinck 1989). This suggests that in synchronic grammar, too, variable marking of objects may be conditioned by factors beyond referential properties of objects, like word order (as also suggested by (2b)).

In evaluating the two approaches sketched above we claim that the Markedness Reversal approach gives a natural account of 'static' or grammaticalized DOM systems like Turkish, which always case-marks the marked objects (specifics) and never the unmarked, regardless of the subject type. It cannot, however, account for the more dynamic aspects of DOM, i.e. variable marking of inanimate objects as in (2). The Distinguishability approach can better explain such dynamic aspects of DOM but provides a less natural account for 'static' aspects. This holds especially true in well-known cases of grammaticalization of a-marking. For Upper Engadian, for instance, Bossong (1991) reported that the 1st person pronoun *me* fused with the *a* to form *ame*. Interestingly this new pronoun still receives a-marking resulting in *a(d) ame*. It is unnatural to analyze these cases as resulting from a need for distinguishability. Rather a constraint seems at work which states that all pronouns must be a-marked (cf. the Markedness Reversal approach). In the present work we propose to combine forces of the two models sketched above in order to account for both the static and dynamic aspects of DOM.

Taking a historical perspective, we show that stochastic bidirectional OT (cf. Jäger 2003) enables us to model the grammaticalization path where a-marking originated as a means of distinguishing subject and object due to the loss of case inflection in Late Latin and then developed into a marker that identifies animate, specific objects. This means that over time the DOM system is changing from a dynamic system (based on distinguishability) to a static one (captured by Markedness Reversal). In our model this change is captured by letting the constraints from the Markedness Reversal model, which refer to 'identifying' properties of objects, gain weight over time with respect to the Distinguishability constraint.