



HAL
open science

Coreference in the light of pronouns with indefinite reference

Frédéric Landragin

► **To cite this version:**

Frédéric Landragin. Coreference in the light of pronouns with indefinite reference. Conference on R-impersonals, Feb 2015, Paris, France. halshs-01153989

HAL Id: halshs-01153989

<https://shs.hal.science/halshs-01153989v1>

Submitted on 20 May 2015

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Coreference in the light of pronouns with indefinite reference

Frédéric Landragin
Lattice-CNRS/ENS/Université de Paris 3

Introduction

When studying references to human beings in narrative texts, we observe examples with exact and unambiguous referring expressions, like (some) proper nouns and definite descriptions, i.e., expressions that refer to a specific and precise referent. When personal pronouns are immediately interpreted as anaphora, it is then easy to identify coreferring relations. By taking into account transitivity, it is then possible to build on coreference chains. Because of its links with theme, topic continuity and referents' salience, a coreference chain is a kind of discourse structure that is taken into account by an increasing number of linguistic studies and Natural Language Processing (NLP) applications.

But coreferring relations are often considered as strict: each referring expression of a text fully belongs to one coreference chain, i.e., following an all-or-nothing relationship.

Some other referring expressions, for instance those that refer to groups of persons, organizations, or generic referents, raise some problems when attributing the expression to a coreference chain. In NLP applications, those fuzzy cases are often ignored or simplified. In order to go deeper into the modelling and formalization of reference and coreference, we propose in this talk a model that implies three degrees instead of all-or-nothing coreferring relations. We show how pronouns with indefinite reference, and in particular the French R-impersonal pronoun *on*, put a spotlight on such a model and allow reconsidering the semantic properties of such a complex marker.

Three degrees for coreferring relations and links with *on*

First degree: “strict coreference”. In this case referents are identical, whether they are individual, groups or generic. Examples of such a coreferring₁ relation are the following, i.e., when *on* refers strictly to a precise referent, for instance the speaker (1), when the reference of *on* is explicitly and unambiguously mentioned (2), and when several occurrences of *on* are clearly coreferring, for instance because a referent change would break up the narration (3).

- (1) **On** se dit ça, **on** se prépare (Jean-Luc Lagarce, *Juste la fin du monde*, page 21).
- (2) **On** dormait un peu, **leur père et moi**, sur la couverture (J.-L. Lagarce, page 28).
- (3) **On** gratte, **on** gratte et puis très vite **on** respire mal, **on** sue, il se met à faire terriblement chaud (Jean Echenoz, *L'occupation des sols*).

Second degree: “inclusive coreference”. In this case one of the referents is a group, more precisely a fuzzy group, that is, a group for which we cannot say how many individuals are included – as opposed to a strict group, for instance the one with the mother and the father in (2). Together with the fuzzy group, the coreference chain includes another referent(s), which can be an individual like in (4) where *on* in *on klaxonnait* refers to the father who is driving, and the other occurrences of *on* refer to the whole family. This example emphasizes how the reference of *on* can be vague in context, as does example (5), where the three referring expressions are linked to different referents – the mother, the father, and both of them – but where all of them can be grouped into one coreferring₂ relation, which reflects the discursive intention (and which is a practical alternative to a model including three coreference chains).

- (4) **On** disait qu'**on** « partait en vacances », **on** klaxonnait, et le soir, en rentrant, **on** disait que tout compte fait, **on** était mieux à la maison (J.-L. Lagarce, page 28).
- (5) **On** travaillait, **leur père** travaillait, **je** travaillais (J.-L. Lagarce, page 25).

Third degree: “fuzzy coreference”. In this case both referents are fuzzy groups. The intersection itself between these groups can be fuzzy, as in (6), where a lot of interpretations of *on* can be proposed (group of characters, generic character, witness-narrator), and in (7) where the reader is unable to list the referents of *on* and *ils*.

- (6) Ah ! nom de Dieu ! oui, **on** s’en flanqua une bosse ! Quand **on** y est, **on** y est, n’est-ce pas, et si l’**on** ne se paie qu’un gueuleton par-ci par-là, **on** serait joliment godiche de ne pas s’en fourrer jusqu’aux oreilles. Vrai, **on** voyait les bedons se gonfler à mesure [...] (Emile Zola, *L’assommoir*, cf. Maingueneau 2000).
- (7) Et **on** renonce à moi, **ils** renoncèrent à moi (J.-L. Lagarce, page 30).

Back to the values of *on*

To us, *on* is the referring form that fits at best the notion of fuzzy group, and therefore the three degrees of coreferring relations we propose. With this study, we want to put forward a way to consider *on* in narrative texts. Instead of searching for several potential referents and for complex links between them – complex person, nondescript member of a group, witness-narrator, partial schizophrenia, a-definites, selfhood, person enallage, etc. – we prefer consider *on* as an intrinsic fuzzy marker. Of course, there are a lot of cases where strict reference can be identified, but there is the need for fuzzy aspects in order to model these simple cases and complex examples together. Talking about fuzzy reference is a way to take into account the multiple values of *on* in a unified framework, and to allow NLP exploitations.

It is also a way to emphasize the fact that identifying the exact referent of *on* is not mandatory. Sometimes, the reader is not able to identify the referent of a referring expression, and can nevertheless go on reading and understanding the text. The resolution of the reference can be left unfinished, and “good-enough” approaches to language processing as well as the third degree of coreference is a way to model that phenomenon. Doing that, the importance of the referring act is diminished compared to the importance of the predicative act. Examples (1), (3), (4), (6) and (7) illustrate well this hypothesis. Events are essential, not referents. Discursive aspects are therefore more salient than the characters.

References

- ANSCOMBRE C. (2005) « Le *On* locuteur : une entité aux multiples visages », In : Bres J. *et al.* (Eds.) *Dialogisme, polyphonie : approches linguistiques*, De Boeck Duculot, pp. 75-94.
- BEGUELIN M.-J. (2012) « La concurrence entre *nous* et *on* en français », *Colloque « Noi – Nous – Nosotros. Sur les traces d’un pronom »*, Université de Zürich.
- BLANCHE-BENVENISTE C. (2003) « Le double jeu du pronom *on* », in : Berré M. *et al.* (Eds.), *La syntaxe raisonnée*, De Boeck Supérieur « Champs linguistiques », pp. 41-56.
- CABREDO HOFHERR P. (2008) « Les pronoms impersonnels humains – syntaxe et interprétation », *Modèles linguistiques* XXIX-1, vol 57, pp. 35-56.
- CREISSELS D. (2011) « Impersonal Pronouns and Coreference: two Case Studies », *Workshop on Impersonal Pronouns*, téléchargeable sur <http://www.umr7023.cnrs.fr>.
- DETRE C. (1998) « Entre ipséité et altérité : statut énonciatif de « on » dans *Sylvie* », *L’information grammaticale* 76, pp. 29-33.
- DEULOFEU J. (1989) « Les couplages de constructions verbales en français parlé : effet de cohésion discursive ou syntaxe de l’énoncé », *Recherches sur le français parlé* 9, pp. 111-141.
- FLØTTUM K., JONASSON K. & NOREN C. (2007) *On : pronom à facettes*, De Boeck Duculot, Bruxelles.
- KOENIG J.-P. & MAUNER G. (1999) « A-definites and the semantics of implicit arguments », *Journal of Semantics* 16(3), pp. 207-236.
- LANDRAGIN F., TANGUY N. (2014) « Référence et coréférence du pronom indéfini *on* », *Langages* 195, p.99-115.
- LEEMAN D. (1991) « *On* thème », *Linguisticae Investigationes* 15, pp. 101-113.
- MAINGUENEAU D. (2000) « Instances frontières et angélisme narratif », *Langue française* 128, pp. 74-95.
- RABATEL A. (2001) « La valeur de *on* pronom indéfini / pronom personnel dans les perceptions représentées », *L’information grammaticale* 88, pp. 28-32.
- SCHAPIRA C. (2006) « *On* pronom indéfini », in : Corblin F. *et al.* (Eds.), *Indéfini et prédication*, Presses de l’Université Paris-Sorbonne, Paris, pp. 507-518.
- SIEWIERSKA A. (2011) « Overlap and complementarity in reference impersonals », In: Malchukov A., Siewierska A. (Eds.), *Impersonal Constructions: A cross-linguistic perspective*, pp. 57-90.
- VIOLLET C. (1988) « Mais qui est *On* ? », *Linx* 18, pp. 67-75.