Coreference in the light of Pronouns with Indefinite Reference

Frédéric Landragin
CNRS, Lattice Laboratory

Conference on R-impersonal pronouns
February 2015

Content

- **Context and objectives**
  - Coreference chains in French
  - Referents and links between referents
  - R-impersonal pronouns, “on”

- **Data and analyses**

- **Model**
  - Three degrees for coreferring relations
  - Importance of “on” in the three degrees

- **Back to the interpretation of “on”**
  - Vagueness and discursive aspects
  - Salience
Context and objectives

• Several studies of coreference chains in (short) written texts in French, mostly narrative texts

• Approach: corpus linguistics using (manual) annotation tools

• Objectives
  ◦ Exploring referents types and links between referents:
    specific versus generic,
    singular versus plural,
    strict group versus fuzzy group…
  ◦ Exploring the nature of coreference chains: most important elements and less important ones
  ◦ Towards a linguistic model for coreference
  ◦ Natural Language Processing (NLP) applications
  ◦ Corpus linguistics methodology: improving annotation tools
  ◦ First results: thematic issue of Langages journal, number 195 (sept 2014)

Preliminary corpus

• Jean Echenoz, L’occupation des sols (1988)
  ◦ Short story, 1787 words
  ◦ Contemporary literature, with several "on" (25 occurrences)
  ◦ Specificity of the text: a character is presented together with its representation (painting) on the wall of a building (advertising)
  ◦ Cf. articles in RSL journal, number 3, http://rsl.revues.org/

• Jean-Luc Lagarce, Juste la fin du monde (1990)
  ◦ Contemporary theatre, with numerous “on”
  ◦ Specificity: complex relations between parents and children, full of ambiguities – numerous ambiguous referring expressions

• A set of remarkable examples
  ◦ Found in French grammar handbooks
Coreference chains in French

- Set of all the referring expressions that refer to the same referent
- Here: exact and unambiguous referring expressions

Elements of coreference chains that are not referring expressions
In parallel with coreference chains:
annotation of referring expressions

- Features for the annotation of an occurrence of “on”

Annotation and visualization of coreference chains using ANALEC

- ANALEC = “Analyse de l’écrit”
  - Tool designed and implemented by Bernard Victorri (Lattice Laboratory)
  - Free and open source (JAVA), cf. http://www.lattice.cnrs.fr/
  - Adapted to the study of reference and coreference in written texts
  - Still in development…
Reference of “on”

- Sometimes very clear
  - 1st person plural ("nous", when identifying the referents is easy)
  - Episodic use (someone, useless to search for a precise referent)
  - Generic use (everyone, useless to search for a precise referent)
  - Some stylistic uses, when “on” clearly stands for “je”, “tu”, “vous”…
    (a doctor at the hospital: “Alors, on a bien dormi cette nuit ? ”)

- Sometimes very vague
  - Some stylistic uses: when the logical contextual interpretation does not correspond to any classical value of “on”, in scientific articles…
  - 1st person plural: like for “nous”, it is sometimes impossible to list the referents (some of them can be listed, but the borders of the group are vague – for instance, the reader may be implicitly included)
  - Ambiguity between generic use and 1st person plural
  - In all these cases, specifying coreference chains is quite difficult

“on” linked to “je”

- Typical examples in Juste la fin du monde
  - (1) “Je vous ennuie, j’ennuie tout le monde avec ça, les enfants, on croit être intéressante” (page 14)
  - (2) “On se dit ça, on se prépare” (page 21)
  - (3) “On pleure. On est bien. Je suis bien” (page 44)

- Comments
  - (1) – Perhaps an ambiguity between “je” and a generic referent
  - (2) et (3) – Maybe the 1st person, maybe a generic referent
  - (3) – Example that comes after a long alternation between “on” and “je”; in which the reference of “on” is evolving from a generic person to “je”

- Coreference chains
  - (1) – One chain with “je”, “j’” and “on”?
  - (3) – Potential problem: with one chain, we loose the nuances between “on” and “je”; with two chains, we loose the strong link between “on” and “je”…
“on” linked to “nous”

- Examples in *Juste la fin du monde*
  - (4) « On dormait un peu, leur père et moi, sur la couverture » (page 28)
  - (5) « On travaillait, leur père travaillait, je travaillais » (page 25)
  - (6) « Nous sommes toutes les trois, comme absentes, on les regarde » (page 69)
  - (7) « Cela nous rend service et on n’est pas toujours obligées de demander aux autres » (page 24)
  - (8) « Les patrons nous connaissaient et on y mangeait toujours les mêmes choses » (page 28)
  - (9) « Le dimanche nous allions nous promener. Pas un dimanche où on ne sortait pas » (page 26)
  - (10) « Les autres jours nous allons chacun de notre côté, on ne se touche pas » (page 56)

- Comment: no ambiguity here, “nous” and “on” are always coreferent

“on” linked to… nothing clear

- A typical example of *Juste la fin du monde*
  - (11) « Oui ? On est là ! » (page 57, uttered by Suzanne just after the mother called “Louis !” – Louis is not present, but Suzanne and Antoine are – so “on” may include Suzanne and Antoine but not Louis)

- Comments
  - Theoretically, “on” must include Louis, and may include Suzanne (speaker-inclusive in this kind of dialogue)
  - The reference is blurred and the ambiguity contribute to the misunderstanding that characterizes the text
  - “Je” was not possible, “il” was not possible, “nous” would have sound strange… “on” sounds strange but perhaps less strange than “nous”
  - “On” = referring expression of last resort?
  - “On” = the only referring expression that is able to (discreetly) convey an intentional ambiguity?

- Coreference chain: this “on” may be left alone…
“on” linked to… nothing clear

- A typical example of *L’occupation des sols*
  - « Comme on ne possédait plus de représentation de Sylvie Fabre, il s’épuisait à vouloir la décrire toujours plus exactement »

- Comments
  - “on” may refer to the group including the son and his father
  - “on” may not refer… as if the sentence were « Comme il n’y avait plus de représentation de Sylvie Fabre »
  - Why “on” and not “ils”? (cf. Charolles & Storme, 2015): « *On* est beaucoup plus subtil que *ils* car, même dans les emplois où il se contente de poser l’existence d’un événement, indépendamment des participants au procès, il manifeste toujours une sorte d’implication du locuteur ou du rédacteur dans celui-ci »

- Coreference chain: this “on” may also be left alone…

“on” linked to genericity

- Examples in *Juste la fin du monde*
  - (12) « *On* ne peut pas plaisanter » (page 16)
  - (13) « *On* ne peut pas m’accuser » (page 68)
  - (14) « *On* me comprend » (page 17)
  - (15) « Je n’ai rien dit, *on* dit que je n’ai rien dit » (page 16)
  - (16) « Et *on* renonce à moi, *ils* renoncent à moi » (page 30)
  - (17) « Et alors il faut te chercher, *on* doit te chercher » (page 59)
  - (18) « Comment est-ce qu’*on* dit ? “d’une pierre deux coups” » (page 64)
  - (19) « *On* ne devrait jamais se lâcher, serrer les coudes, comment est-ce qu’*on* dit ? » (page 68)

- Comments
  - (16) (and (17)) – The repetition seems to constraint the coreference
  - (18) and (19) (and (15)) – Has “on” inside a fixed form a referent?
  - (19) – Among other examples, here is a case where two occurrences of “on” in the same sentence are not coreferent
“on” linked to cohesion, coherence and discursive aspects

- Several “on” in the same sentence in Juste la fin du monde
  - (20) « 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 » (page 28)

- Comments
  - Most of the occurrences of “on” refer to the whole family: the parents and the three children
  - “On” in “on klaxonnait” raises a problem: the father is driving, so only the father honks (but it was probably the intention of the whole family)
  - “On disait qu’on partait en vacances, le père klaxonnait […] on…” sounds less coherent
  - Then the choice of “on” has also discursive reasons

- Coreference chains: only one in this example, grouping the five occurrences of “on”

“on” linked to story progression

- In L’occupation des sols, the interpretation of “on” evolves as the story progresses
  - At the beginning, “on” refers to the group “father + son”
  - After some occurrences with a generic or an episodic interpretation, “on” is getting more and more vague
  - In « regarde un peu le soleil qu’on a », “on” includes the father, the son, and potentially anyone who stands there (and looks at the sky) (cf. Charolles & Storme, 2015)
  - At the end, “on” is a sort of indistinct person, an agglomerate of the father and the son, and maybe other persons that are not described: “on” = “anyone who, like Fabre and his son, would live such a situation”
  - This is a point that allows us to make a link between L’occupation des sols and Juste la fin du monde
  - “On” = 1st person plural + episodic referent?
    1st person plural + generic referent?
Towards a model for coreference

- **Three degrees instead of all-or-nothing coreferring relations**
  - Important elements (strong links, “maillons forts” in French) and less important elements (weak links, “maillons faibles”)
  - 1st degree = strict coreference
  - 2nd degree = inclusive coreference (“on klaxonnait” – “on partait”)
  - 3rd degree = fuzzy coreference (“le soleil qu’on a” – “on gratte”)

- **Consequences**
  - A coreference chain is not just a set of referring expressions
  - For NLP applications, there does not exist any framework (annotation structures, coding recommendations, evaluation metrics) that can take into account such structures
  - “On” seems to be the referring form that fits at best the notion of fuzziness, and therefore the three degrees of coreferring relations
  - Instead of notions like ‘nondescript member of a group’, ‘a-definites’, ‘partial schizophrenia’, we prefer consider “on” as an intrinsic fuzzy marker

Last example

- **From Emile Zola, L’assommoir & discussed in (Maingueneau 2000)**
  - “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 […]”

- **Comments**
  - A lot of analyses of the occurrences of “on” in this example: several values, and then several coreference chains
  - To us: 3rd degree coreferring relations (one coreference chain)

- **Is the identification of the referents of “on” 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 3rd degree of coreference is a way to model this phenomenon
Conclusion and future works

- **Discursive roles of “on”**
  - The reference of “on” can be vague
  - The coreference with “on” can be vague
  - The value of “on” can evolve as the text progresses
  - Several “on” in the same sentence can be linked by coreferring relations for cohesion and coherence reasons
  - Then: “on” sometimes seems to be more linked to discursive aspects than to referential aspects

- **Salience**
  - In *L’occupation des sols*, « on gratte, on gratte »: what is salient is not the referent (“father + son”) but the action
  - In *Juste la fin du monde*, “le soleil qu’on a”, “on klaxonnait”: idem
  - When we read a sentence beginning with “on”, we do not have the exact referent in mind: the reference may be calculated at the end of the sentence, which means that the situation (and not the referent) is salient (cf. Charolles & Storme, 2015)