

Grammaire de *on* et genres textuels

Denise Malrieu

UMR MODYCO – Paris X

dmalrieu@u-paris10.fr

1 - Les objectifs

2 - La démarche

3 - L'approche quantitative (fréquences brutes)

4 - Les valeurs indexicales et non indexicales de *on*

5 - La répartition des valeurs de *on* selon les genres et les DR

1. Les objectifs

L'objectif poursuivi à travers les analyses contrastives de l'usage du *on* dans différents genres textuels est de dessiner une démarche opératoire pour définir les contraintes interprétatives portées par le genre. Dans une optique largement discutée par Coseriu avec sa notion d'entours, par F. Rastier dans son herméneutique matérielle ou par S. Bouquet dans sa volonté de joindre linguistique de la langue et linguistique de la parole, la démarche voudrait préciser les hypothèses de Bakhtine sur le normativité des genres.[...]

1.1. L'historique : La démarche de caractérisation des genres par des variables morpho-syntaxiques

Je ne reviendrai pas ici sur les apports des travaux pionniers de Bronckart et J.M.Adam dans le domaine francophone. Dans le domaine anglophone, les travaux de Biber ont été à l'origine de la *linguistique de corpus dans son acception forte* et des grammaires "corpus-driven"

Biber observe que les distributions d'un lexème ou d'un morphème sont bien différentes selon les registres et que certains mots sont spécifiques à certains registres; que les ambiguïtés grammaticales de la forme "-ed" en anglais par ex, ou la catégorie grammaticale (verbe/nom/adjectif) pour un même mot se répartissent de façon contrastée selon que l'on est dans des textes expositifs ou fictionnels. Il en vient donc à la définition de grammaires et dictionnaires intégrant les variations de la norme linguistique (*Longman Grammar of Spoken and Written English*, Biber, Johansson, Leech, Conrad, & Finegan, élaborée à partir d'un corpus de 40 millions de mots dans 4 registres : conversation, fiction, articles de presse, textes académiques).

Les analyses de Rastier et Malrieu à partir des variables issues du corpus de Synapse

Une première démarche déductive sur un corpus de 2541 ouvrages étiquetés par Synapse avait permis de montrer (Malrieu et Rastier, 2001) en mettant en œuvre l'analyse univariée, l'existence de variations morphosyntaxiques systématiques selon les domaines, champs génériques et genres textuels; l'analyse en composantes principales montrait que le facteur principal dans l'analyse factorielle est en rapport avec les variables liées à l'oral représenté. La démarche inductive par la classification hiérarchique ascendante sur les 250 variables morphosyntaxiques obtenues par CORDIAL redonne de façon nette les découpages en domaines et champs génériques (juridique vs théâtre vs poésie vs essais); cependant, la classification des genres narratifs s'avérait moins évidente.

Classification hiérarchique ascendante
de 2600 ouvrages de différents genres

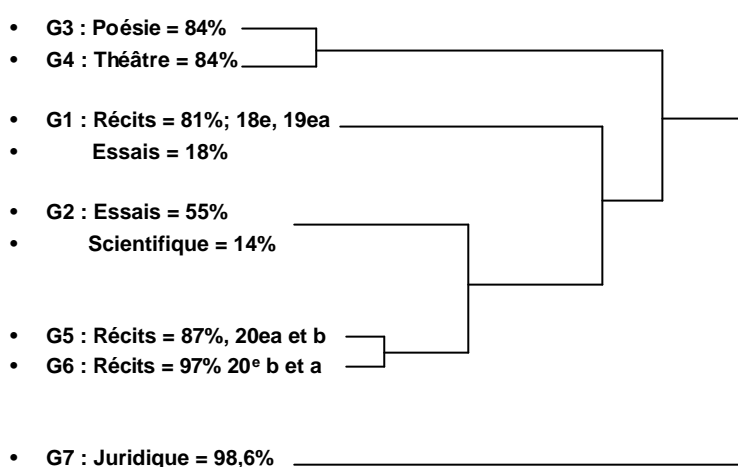


Tableau 1 : classification hiérarchique ascendante en domaines et champs génériques

1.2 – Les limites de l'approche globale des textes pour leur caractérisation

M'intéressant particulièrement aux variations des récits fictionnels et à la caractérisation des narrateurs, mes essais pour décrire de façon plus fine des familles narratologiques au sein du roman "sérieux" se sont avérés peu concluants en utilisant les mêmes variables ou des sous-ensembles de celles-ci (le premier facteur n'expliquant pas plus de 30% de la variance). La démarche inductive de classification des romans rencontrait plusieurs obstacles :

- non disponibilité de corpus raisonnés de romans et suffisamment étoffés.
- nécessité de repenser les variables morphosyntaxiques à prendre en compte dans la classification (insuffisance des variables disponibles dans Cordial : nous en avons redéfini environ 300).
- les calculs sur le texte dans sa globalité effectuent un lissage qui dilue les différences : la plupart des genres sont par essence des textes composites : par ex le fort taux de 1PS ne peut différencier le roman homodiégétique et le roman hétérodiégétique fortement dialogué. Le constat est donc que les contraintes interprétatives portées par le genre impliquent la prise en compte des séquences inférieures au texte, séquences informées par le genre mais qu'il faut caractériser en tant que telles.

1.3 – Reformulation des objectifs :

La démarche de caractérisation morphosyntaxique statistique des genres peut permettre le profilage et la classification des textes à l'intérieur de domaine de pratique précis, mais ne peut guère servir pour la désambiguïsation syntaxique ou sémantique.

Je reprendrai ici les propositions de S. Bouquet (*Langages* n° 153) concernant l'affectation du sens à un énoncé : la linguistique de la langue dans ses grammaires ne décrit que des parties détachées de sens, la linguistique de la parole va définir une pragmatique du sens qui a trait au genre et à la façon dont des traits suprasegmentaux du texte global et de ses sous-ensembles, va contraindre l'analyse grammaticale du segmental.

La démarche

2.1. Le corpus et les genres

L'idée est de définir un corpus comme un ensemble de données expérimentales dont on va faire jouer certaines variables pour l'analyse comparative.

La définition des genres textuels : elle tient compte à la fois des domaines d'activité (littérature vs journalisme), des champs génériques (genres narratifs vs théâtre vs poésie), des genres (dans les genres narratifs, on distinguera le conte, le roman, la biographie, etc.).

J'ai constitué un corpus (évolutif) de textes d'un même auteur (la variable auteur est ainsi neutralisée) mais de genres divers. Ce corpus est composé de romans, d'entretiens, de recueils d'articles de presse.

Dans chacun de ces genres, ces catégories doivent être affinées, principalement en précisant les variations de dispositif énonciatif à l'intérieur du genre :

- en ce qui concerne *le roman* : type de narrateur : intra- vs extradiégétique, hétéro vs homodiégétique, nature des diégèse enchâssées si elles existent, adresse narrateur /narrataire présente vs absente, récit de la diégèse principale au présent vs au passé, les séquences de DR.
- En ce qui concerne *les entretiens* (échange dual en face à face): on peut distinguer les entretiens informels (discussion à bâton rompu) vs interview dissymétrique vs propos recueillis (rapport sélectif et retravaillé des paroles d'autrui)
- En ce qui concerne les *recueils d'articles de presse*, chaque article est affecté à une classe; ce sont tous des articles de presse grand public, mais ils diffèrent par le domaine, le mode d'interlocution, le type d'élaboration : j'ai distingué les catégories suivantes : récit, récit autobiographique, entretiens, propos recueillis, commentaire d'œuvre, article de critique, commentaire d'actualité, réflexions générales, préface, lettre, article nécrologique, recette culinaire.

Dans cet environnement, on pourra donc comparer le récit extradiégétique des articles de presse et d'une nouvelle ou d'un roman, le récit autobiographique et le roman homodiégétique, la conversation représentée dans le roman et la discussion de l'entretien informel, etc.

Genre	Nombre	Nb de mots
romans	6	399366
Entretiens (bâton rompu)	1	68822
Entretiens retravaillés	1	35345
Recueils d'articles de presse	2	140119

Tableau 2 : Composition du corpus

2.2. Le balisage du corpus :

Je ne m'étendrai pas sur cet aspect : le balisage adopté concerne d'une part l'enrichissement du header proposé par la TEI pour ce qui concerne le <textclass> : domaine, champ générique, genre et sous-genres définis plus haut; le balisage de la structure logique du texte est celui de la TEI : niveaux de division du texte; dans les séquences textuelles, m'intéressant aux discours rapportés, j'ai introduit des balises de types de discours rapportés, leurs segments introducteurs et incisives, les balises d'îlots textuels de citation et d'activités métalinguistiques et autonymiques. Cela permet de situer chaque occurrence de *on* à l'intérieur des univers de discours représentés (DR ou discours du narrateur).

Je donne ici la liste des discours rapportés balisés :

- le discours direct (DD) sous ses différentes formes : isolé, en dialogue, en DD rapporté marqué ou pas, en DD introduit par *que*, en DD libre
- le discours indirect, (DI), soit simple soit narrativisé (DIN)
- le monologue intérieur soit simple au DD (MI), soit au DI (MII), soit narrativisé (MIN)
- le discours rapporté narrativisé (DRN souvent désigné DIL) : "Il n'était pas bien, disait-il, et préférait rentrer"

3 – L'analyse fréquentielle de *on* selon les genres et les DR

Un stylesheet XSLT permet d'extraire les DR à l'intérieur de chaque genre : le passage de ces extraits sous CORDIAL donne les fréquences. La valeur comparative choisie pour le taux de fréquence est non pas le % de *on* / nb de mots du genre ou du DR mais le % de verbes dont *on* est le sujet (calculé dans CORDIAL).

On présentera ici les résultats commentés de l'analyse comparative des taux de fréquence de *on* intra- et inter-genres.

4 - Les valeurs indexicales et non indexicales de *on*

Les définitions de *on* varient selon les linguistes : certains le voient comme un nom car nient la possibilité d'un usage anaphorique, d'autres (Riegel et al.) rangent *on* avec les pronoms personnels déictiques dont le référent est identifié à partir de la situation de discours où ils sont employés.

La fréquentation des textes montre que *on* peut remplacer tous les autres pronoms personnels sujets, qu'il est aussi bien déictique, qu'anaphorique ou indéfini.

Je m'appuierai pour l'analyse sur la grammaire des pronoms personnels proposée par S. Bouquet : on commentera l'arborescence suivante (représentation simplifiée)

La démarche adoptée consiste à projeter l'arborescence présentée ci-dessous sur le corpus, en affectant à chaque occurrence de *on* une position dans l'arborescence.

On décrira la répartition des *on* sur les différents nœuds de l'arborescence dont le tableau 3 ci-dessous synthétise les nœuds supérieurs :

- indexicalité extra-textuelle, déictique
- indexicalité intra-textuelle : anaphores
- non indexicalité - Générique

Les nœuds représentés dans le corpus sont les suivants :

1.1. Indexicalité extratextuelle pure :

1.1.1. +IDX/+EXT/+ITL/+LOC/-AMP/+SING/ (interlocutoire du locuteur non amplifié) **ON = Je**

1.1.2. +IDX/+EXT/+ITL/+LOC/+AMP/+ALLO+SING/+EXCL/+MASC vs -MASC (amplification du locuteur à l'allocutaire singulier exclusivement); **ON = nous = moi + toi**

1.1.10. +IDX/+EXT/+ITL/-LOC/-AMP/+SING/+MASC vs -MASC (indexicalité extratextuelle interlocutoire de l'allocutaire non amplifié); **ON = Tu, vous** (politesse)

1.2. Indexicalité extratextuelle + anaphorique :

1.2.1. +IDX/+EXT/+ITL/+LOC/+AMP/+ALLO/+SING/-EXCL/+TIERSING/+MASC vs -MASC (amplification du locuteur à l'allocutaire singulier non exclusivement, incluant un tiers singulier anaphorique) **ON = nous = moi + toi + lui ou elle**

1.2.2. +IDX/+EXT/+ITL/+LOC/+AMP/+ALLO/+SING/-EXCL/-TIERSING/+MASC vs -MASC (amplification du locuteur à l'allocutaire singulier non exclusivement, incluant un tiers pluriel) **ON = nous = moi + toi + eux**

1.2.3. +IDX/+EXT/+ITL/+LOC/+AMP/+ALLO/-SING/-EXCL/+TIERSING/+MASC vs -MASC (amplification du locuteur à l'allocutaire pluriel non exclusivement, incluant un tiers singulier) **ON = nous = moi + vous (vous = toi + n toi) + lui**

1.2.4. +IDX/+EXT/+ITL/+LOC/+AMP/+ALLO/-SING/-EXCL/-TIERSING/+MASC vs -MASC (amplification du locuteur à l'allocutaire pluriel exclusivement, incluant un tiers pluriel) **ON = nous = moi + vous (vous = toi + n toi) + eux**

1.2.5. +IDX/+EXT/+ITL/+LOC/+AMP/-ALLO/+SING/+MASC vs -MASC (amplification du locuteur au non-allocutaire singulier) **ON = nous = moi + lui**

1.2.6. +IDX/+EXT/+ITL/+LOC/+AMP/-ALLO/-SING/+MASC vs -MASC (amplification du locuteur au non-allocutaire pluriel) **ON = nous = moi + eux**

"Nous, ce qu'on veut, c'est partir."

2 - Indexicalité intra-textuelle

2.1. +IDX/-EXT/+SING/+MASC/ (indexicalité intratextuelle singulière masculine) **ON = il,**

2.2. +IDX/-EXT/+SING/-MASC/ (indexicalité intratextuelle singulière féminine) **ON = elle**

2.3. +IDX/-EXT/-SING/+MASC/ (indexicalité intratextuelle plurielle)

2.3.1. on = lui + lui :

2.3.2. on = lui + elle :

2.3.3. on = lui + eux :

2.3.4. on = lui + elles :

2.3.5 : (indexicalité intratextuelle plurielle masculine) : **on = ils,**

2.3.6 : (indexicalité intratextuelle plurielle féminine) : **on = elles,**

3.2 : on = actant collectif représentant l'autorité, le pouvoir : "on les déporta aux bouches de Cattaro"

3.3 +IDX/-EXT/-SING/+MASC/-DEF (indexicalité intratextuelle singulière masculine non définie, actant prototypique à l'intérieur d'un scénario)

3.8 +IDX/-EXT/-SING/+MASC/-DEF (indexicalité intratextuelle singulière : construction de la scène "on pouvait voir..., on entendait")

3 – Non indexicalité : valeur GÉNÉRIQUE :

3.1 : modalisation épistémique : "on dirait que, on eût dit" : catégorie problématique

3.4 : énoncé déontique : "Apprends que quand le brutal gronde, on ne montre jamais d'or"

3.5 : énoncé gnominique, vérité universelle : "la vérité comme on la dit à un vieil ami"

3.6 : activités métalinguistiques, autonymie : "comme on l'appelle"

3.7 : la doxa, l'opinion publique : "c'est ce qu'on dit"

	ACN	Bar- rage	Dix Heu- res	Mod- erat- o	Cha- rtreu- se	Vie- mat	Vie- mat	Out- side	Out- side	Mon- de ext	Par- leu- ses	ACN	Bar- rage	Bar- rage
	Non DR	Non DR	Non DR	Non DR	Non DR	récit auto bio	récit auto bio	récit auto bio	récit auto bio	DD	DD	DD	DD	récit 1S
Index extra-text	15,1	2	0	0	9,3	16,8	39,3	9,7	58,6	37,9	23,5	63,7	71,6	74,7
Index intra-text	50,0	38,8	82	88,2	63,8	19,9	17,	53,5	19,5	13,8	17,2	10,5	6,3	12,
Non Index Gen	34,9	58,2	18	11,8	26,9	62,8	43,7	36,8	21,8	48,3	59,3	25,7	22,1	13,3
Nb total de on	126	98	50	34	483	191	135	155	87	58	565	171	222	75

Tableau 3 : répartition des genres selon les nœuds supérieurs de l'arborescence (% en colonnes)