

J.-P. BENZÉCRI

## **Sur l'étude des textes russes d'après les occurrences des formes de mots**

*Les cahiers de l'analyse des données*, tome 19, n° 1 (1994),  
p. 7-34

[http://www.numdam.org/item?id=CAD\\_1994\\_\\_19\\_1\\_7\\_0](http://www.numdam.org/item?id=CAD_1994__19_1_7_0)

© Les cahiers de l'analyse des données, Dunod, 1994, tous droits réservés.

L'accès aux archives de la revue « Les cahiers de l'analyse des données » implique l'accord avec les conditions générales d'utilisation (<http://www.numdam.org/conditions>). Toute utilisation commerciale ou impression systématique est constitutive d'une infraction pénale. Toute copie ou impression de ce fichier doit contenir la présente mention de copyright.

NUMDAM

Article numérisé dans le cadre du programme  
Numérisation de documents anciens mathématiques

<http://www.numdam.org/>

## SUR L'ÉTUDE DES TEXTES RUSSES D'APRÈS LES OCCURRENCES DES FORMES DE MOTS

### [TEXTES RUSSES]

J.-P. BENZÉCRI

#### 1 Traitement du texte et dénombrement des formes en langue russe

##### 1.1 Langue russe et langues indo-européennes

Des corpus littéraires en grec, latin et espagnol, ainsi que des articles et documents en français, ont été analysés en traitant les textes bruts par des algorithmes de dénombrement de formes, dont la conception et la mise en œuvre ne requièrent du statisticien que la saisie des textes, sans aucune élaboration lexicale ou grammaticale; (cf., e.g., [TEXTES GRECS, 1 & 2], Voll. XVI, n°1, et XIX, n°2; [TEXTES LATINS], Vol.XVI, n°4; [SIÈCLE d'OR], Vol.XVII, n°4; [CAD XII-XVII, 1-3], Vol.XVIII, n°1); et nous estimons, d'après un premier essai (cf. [MOTS ARABES]), que la même méthode s'appliquera utilement à la langue arabe.

La langue russe, se présente, *a priori*, avec des caractères généraux qui appartiennent à d'autres langues indo-européennes; la plus semblable, parmi celles que nous avons déjà étudiées, pouvant être le latin. D'une part, la déclinaison multiplie les formes des noms, adjectifs et pronoms; d'autre part, le nom est généralement isolé: l'usage du démonstratif ne s'étant pas étendu jusqu'à faire de celui-ci un article. Mais, de plus, le verbe *être* est particulièrement défectif; à telle enseigne qu'au présent, la phrase simple formée d'un sujet et d'un attribut, se réduit au couple sujet-prédicat, sans couple verbale interposée, comme dans la phrase nominale arabe.

Ces particularités pouvaient faire craindre que l'analyse statistique ne fût gênée, à la fois par l'émiettement des formes des outils, et la faible fréquence globale de ceux-ci relativement aux mots pleins. L'expérience dont rend compte le présent article atteste, croyons-nous, que la voie est ouverte à une stylométrie des textes russes fondée sur les dénombrements mécaniques de formes. Cette perspective nous intéresse d'autant plus que nous ne connaissons que deux applications de la statistique multidimensionnelle à la

langue russe: d'une part, de J.-J. CASTOT, une typologie de noms et de verbes russes; ceux-ci caractérisés par le profil des compléments qu'ils admettent et ceux-là par leur profil d'emploi aux divers cas, in *Pratique de l'analyse des données en linguistique et lexicologie*, Prat Ling, LC1 n°9; Dunod; Paris; (1981); d'autre part, une remarquable analyse, non publiée, de Mr. VOLOCHINE, consacrée au vocabulaire et au style de M. CHOLOKHOV.

### **1.2 Particularités de l'écriture de la langue russe**

Un texte latin peut s'écrire sans recourir à d'autres caractères que les capitales et minuscules d'un clavier usuel; même si l'on introduit des ligatures {æ œ}, des trémas, des accents, ceux-ci disparaissent dans l'élaboration du texte; et la liste des formes ne comporte aucun autre signe que les minuscules d'une police usuelle.

Le français, l'espagnol et, plus encore, le grec offrent quelques difficultés; mais on peut, en introduisant un petit nombre de signes diacritiques, utiliser une même police de caractères pour la saisie du texte et le listage des formes; parce que l'ordre des octets qui correspondent aux touches frappées est, à très peu près, l'ordre alphabétique même propre à la langue.

En russe, comme en arabe, mais à un bien moindre degré, la frappe du texte sur un clavier usuel, suivant un code naturel, aboutit à une suite d'octets ne respectant pas l'ordre alphabétique propre à la langue; en sorte qu'il faut, pour le tri des formes, renuméroter les octets du texte saisi; et, pour l'affichage, présenter ces octets avec une police dont la frappe directe au clavier ne pourrait être complètement guidée par des associations mnémoniques naturelles entre lettres russes et touches du clavier français.

Nous avons donc créé deux polices, Russ et Ruz: la première, commode pour la frappe; la deuxième respectant l'ordre alphabétique. Ces polices sont conçues pour les ordinateurs Macintosh, mais pourraient être adaptées à d'autres systèmes de gestion des octets.

### **1.3 La frappe du texte russe**

Nous donnons, sur deux lignes parallèles, l'ensemble des minuscules de la police Russ, et la suite des lettres qui correspondent à leurs octets dans une police latine usuelle. Il n'y a pas lieu de présenter les capitales qui s'obtiennent comme celles des signes latins correspondant, avec des frappes qui, pour les signes spéciaux, dépendent de la gestion du clavier par le système: par exemple, le 'chtch', чч s'obtient au clavier français en frappant la touche ç; et, avec le système utilisé pour composer cet article, la capitale 'Chtch', Чч, vient en frappant la même touche, ç, en même temps que la commande de caractère spécial, comme pour obtenir Ç dans un texte français.

абсдефгиіжклмноцрстуvwшхчз яэъьйѳю щ а ж љ ѣ ѵ ѿ ѳ ѵ  
 abcdefghijklmnopqrstuvwxyz àèêîïòù ç ã õ æ œ π ô ß ø ÿ

L'ensemble des caractères est présenté comme un alphabet latin; suivi d'un bloc de voyelles accentuées; du signe isolé ç; et d'un bloc complémentaire. La correspondance entre lettres russes et touches du clavier est généralement régie par le son: a pour а, b pour б, v pour в...; il suffira donc de rendre raison des nécessaires exceptions à cette règle.

Le s russe, с, qui s'écrit comme notre c, est obtenu en frappant c, et non s: en effet un caractère dessiné s, figure dans l'écriture slave ancienne, et nous en avons réservé la place.

En frappant les touches {h i}, on obtient les caractères {и і}: ce dernier est inusité en russe moderne, mais il se trouve dans d'autres langues utilisant l'écriture cyrillique, ainsi que dans les textes russes imprimés avant 1917; on a réservé la place du i, et assimilé le и au h: car il est issu du èta grec, qui, en capitale, coïncide avec notre H.

On ne peut trouver, en russe, d'analogue phonétique à notre lettre q (qui fait double emploi avec k): en frappant q, on obtient ц: ce qu'on retiendra parce que, d'une part, Q a un crochet, comme ц; et, d'autre part, le ц russe résulte par transformation du k d'autres langues slaves: e.g., en russe цвет, fleurs (collectif); en polonais kwiat, fleur.

Le y russe est à la place du u français, lequel se prononce 'ou' dans la plupart des langues de l'Europe.

Les lettres {ш х ч}, {ch kh tch}, du cyrillique sont logées, d'après une analogie de forme, à la place des lettres latines {w x y}.

Au delà de l'alphabet usuel, se trouve un bloc de voyelles accentuées; qui correspondent à des voyelles ou semi-voyelles du russe. On a, de préférence, utilisé l'accent grave, qu'on obtient généralement en une seule frappe sur le clavier français. L'analogie nous semble claire: à pour я, ja; ù pour ю, ju... Avec i, on a utilisé les trois accents. Le ѳ, frappé ò, qui marque, en bref, l'absence de mouillure, offre la particularité d'avoir presque totalement disparu de l'orthographe contemporaine; tout en subsistant dans certains cas. Quant au ѳ, frappé ê, il a totalement disparu.

Du ц, chtch, russe, la lettre française ç rappelle le son sifflant et le crochet.

Pour être complet, il faut signaler le ë, dont la fréquence d'emploi varie selon le style orthographique; et les accents, dont l'usage diacritique est exceptionnel dans un texte destiné à des russes (e.g. временнóе, temporel ≠

временное, temporaire); mais qu'on introduit pour indiquer la prononciation aux étrangers.

Enfin, est réservée la place d'autres caractères cyrilliques {А Ъ Ы Ъ Ц ѡ ѣ ѐ ѱ...}; qui ne sont plus utilisés en russe; et dont seuls les derniers, correspondant au thêta et à l'upsilon grecs, mais prononcés respectivement f et i, se trouvent encore imprimés au début du siècle dans des mots venus de Byzance: e.g. СКѦЫ, les Scythes, écrit aujourd'hui Скифы (cf. *infra*, §2.6.1).

#### 1.4 Du texte frappé à la liste des formes

On a, pour la langue russe, un programme 'triruss' qui transforme le fichier de texte saisi, en une suite de formes, chacune étiquetée par verset et chapitre, selon le numérotage introduit dans le texte.

Aux particularités près qu'implique l'écriture du russe, ce programme ne diffère pas essentiellement des programmes 'trigrec', 'trilat', 'trigalac', 'trihisp'... utilisés pour le grec, le latin, le français, l'espagnol..., dans de précédentes études.

```
@ABCDEFGHJKLMNOPQRSTUVWXYZ [ \ ] ^ _ abcde
абвгдежзийклмнопрстуфхц чщщъььѣюя
```

La police 'ruz' de sortie, restreinte aux minuscules et logée dans la suite d'octets qui, dans le code ASCII représentent les caractères de @ à e, respecte l'ordre de l'alphabet russe; mais n'offre qu'une analogie imparfaite entre lettre russe et lettre latine de même rang.

Les difficultés de transcription viennent de ce que l'on peut disposer de textes qui, notamment pour des raisons philologiques, ont été saisis suivant toute sorte de norme orthographique. On a donc pris le parti d'appliquer au texte saisi la réforme orthographique la plus radicale: on supprime tout accent ou tréma: ë est écrit e; de même ѣ est transcrit e; le signe dur, ъ, est éliminé; (même si ѣ et ъ ont leur place dans 'ruz', en vue d'applications éventuelles à d'autres langues). Malgré cela, on ne peut réduire mécaniquement à une norme unique les textes contemporains et ceux, imprimés au début du siècle, où l'on distingue, selon le genre, deux formes du nominatif pluriel des adjectifs: par exemple pour bon: добрый et bleu, синій:

nominatif pluriel ancien : masculin : добрые , синіе ;  
fémnin et neutre : добрыя , синія ;

nominatif pluriel moderne: pour les trois genres: добрые , синие ;

mais ceci pourrait être repris par traitement de texte avant de passer au programme 'triruss' et à la suite des traitements.

## 2 Le corpus des textes russes

L'écriture cyrillique des langues slaves est établie avant la fin du premier millénaire de l'ère chrétienne et une tradition continue lie les plus anciens manuscrits aux livres imprimés en Russie jusqu'au règne de Pierre le Grand. Celui-ci eut recours à des maîtres Hollandais pour doter la Russie d'une typographie de style latin; et quand, le 17 (28, selon le calendrier grégorien) Mars 1708, sortit des presses de Moscou, le premier livre de la *Géométrie*, ce fut comme la naissance d'une écriture nouvelle.

Il fallut un siècle pour que les auteurs russes parvinssent, dans tous les genres littéraires, à s'exprimer, avec aisance, en un style qui, autant que la typographie, fût en harmonie avec le concert des muses de l'Europe; où la russe fut désormais reçue parmi les premières.

Le corpus que nous avons saisi ne conte pas, dans toutes ses étapes, cette glorieuse histoire; mais il en traverse plus de deux siècles; depuis une Ode écrite en 1747, jusqu'à une biographie parue en 1981. Nous énumérons les auteurs et les textes; en précisant les sigles des 62 fragments - tous de longueur équivalente - qui servent de base à nos dénombrements de formes.

### 2.1 М.В. ЛОМОНОСОВ (1711-1765) : Ode pour l'avènement au trône de sa majesté Elizaveta Petrovna (1747): {Lm1 Lm2}

Историк, ритор, механик, минеролог, художник и стихотворец, он все испытал и все проник...

“Histoire, rhétorique, mécanique, minéralogie, arts figuratifs, poésie, il a tout tenté, tout pénétré...” C'est ainsi qu'en substance, POUCHKINE honore l'avidité de savoir de celui dont la grande Université de Moscou porte aujourd'hui le nom: M. V. LOMONOSOV. Poète, s'il n'a pas laissé de chef d'œuvre accompli, il a donné à la langue une ferme assurance: léguant à la postérité les règles du vers; et cet heureux alliage de slavon et de russe qui est comme le bronze dont furent coulés les chefs d'œuvre.

L'Ode que nous avons saisie comprend deux parties: éloge de la paix; et éloge de la minéralogie, si utile pour découvrir les trésors que recèle la Russie. Les savants d'aujourd'hui n'ont pas la même élégance quand ils suggèrent aux chefs d'État un favorable partage de leur budget!

### 2.2 Г.Р. ДЕРЖАВИН (1743-1816) : quatre poésies: Dieu; la Cascade; prise d'Izmail; Felitsa (1782); et un discours en prose: {ΔBog Δvd1... Δvd4 Δiz1... Δiz3 Δfa Δfb Δπα Δπб}

2.2.1 Du *Discours raisonné sur la poésie lyrique*, notre édition donne des fragments séparés, où le poète explique des qualités que doit avoir une Ode:

*Inspiration; force et fracas de l'entrée; hauteur de ton;*

*Désordre lyrique; diversité; nouveauté; vraisemblance...*

Le poète, hors de l'enthousiasme, ne touchera pas sa lyre; le torrent des images doit être plus fort que la logique; la clarté, l'ordre, la vraisemblance restant saufs.

Dans ses odes, DERJAVINE s'entend à concilier tant de qualités inconciliables! mais, oserons-nous dire que, toujours brillant, sublime et profond, sans pécher contre l'ordre, il méprisa parfois les transitions.

2.2.2 Au cours d'une lente carrière dans les fonctions militaires et civiles, DERJAVINE s'est appliqué à devenir poète. Agé de 35 ans, il se fait distinguer en dédiant, à la grande Catherine, la "Felitsa".

Brodée sur un conte que l'Impératrice avait, elle-même, écrit pour son petit fils, le futur Alexandre I-er, l'Ode commence par louer la souveraine de n'avoir point les défauts de ses favoris; et termine en prisant en elle une largeur d'esprit, que le poète avait quelque motif de solliciter: et il entreprend ... "de citer parmi ce qu'on dit":

И о себе не запрещаешь  
И быль и небыль говорить;  
Что будто самым крокодилам,  
Твоих всех милостей зоилам  
Всегда склоняешься простить.

*Que, de Toi, Tu n'interdis pas  
De conter le vrai, ni le faux;  
Et qu'à ces crocodiles mêmes,  
De tous tes bienfaits les censeurs, (zoïles)  
Toujours Tu tends à pardonner.*

Le tout fut récompensé d'une tabatière, ornée de brillants, et emplie de pièces d'or. Mais la franchise du poète n'eut pas toujours même récompense.

En suivant DERJAVINE, la langue russe eût pu rendre vie à la beauté antique, dont les autres langues de l'Europe moderne étaient incapables; elle s'arrêta, pour suivre le temps, avec KARAMZINE.

2.3 Н.М. КАРАМЗИН (1766-1825) : lettres de Paris (Mars-Avril 1790); épisodes d'un récit (1792): {Кпа Кпâ Кпб Кпс КНу КNz}

2.3.1 Les lettres se trouvent-elles en français? citons, de ce gentilhomme russe, pénétré de l'esprit français mieux qu'un français ne pourrait l'être aujourd'hui, l'enthousiasme et les embarras.

"Я в Париже!" .....

"Что было мне известно по описаниям, вижу теперь собственными глазами – веселюсь и радуюсь живою картиною величайшего, славнейшего города в свете, чудного, единственного по разнообразию своих явлений.

“*Je suis à Paris .....*”

“*Ce que je connaissais par des descriptions, je le vois maintenant de mes propres yeux. Je fais mon bonheur et ma joie du spectacle de la ville, la plus superbe, la plus illustre au monde; merveilleuse, unique par la variété de ce qu'elle montre.*”

“Система Декартовых вихрей могла родиться только в голове француза, парижского жителя. Здесь все спешит куда-то; все, кажется, перегоняют друг друга, ловят, хватают мысли, угадывают чего вы хотите, чтоб как можно скорее вас отправить.

“*Le système des tourbillons de Descartes n'a pu naître que dans l'esprit d'un français, d'un parisien. Ici, chacun court quelque part; on se bouscule, on saisit les pensées au vol; on devine ce que vous voulez, afin de se défaire de vous au plus vite...*”

Pour ne rien dire d'autres tourbillons: ceux des ruisseaux bourbeux qui coulent au milieu des rues; même de la fameuse rue Saint Honoré,

“*всех длиннее, всех шумнее и всех грязнее.*”

“*de toutes, la plus longue, la plus bruyante et la plus sale.*”

2.3.2 Dans un salon de Paris, KARAMZINE a entendu lire une dissertation sur l'amour; dont il rend compte... avec humour.

Mais il donne, dans “*Nathalie fille de Boïar*”, un modèle de ce que les contemporains eux-mêmes ont appelé le *sentimentalisme*.

Fils d'un seigneur mort en exil, fuyant le Tsar auprès de qui il avait été injustement dénoncé, *Alexis* rencontre *Nathalie*, fille du Boïar *Mathieu*. Les jeunes gens s'épousent secrètement et vivent cachés; jusqu'au jour où une invasion Lithuanienne offre à *Alexis*, secondé de *Nathalie*, tenant le rôle d'un jeune page, frère du héros, l'occasion de rentrer en grâce par un fait d'armes qui renverse, en faveur des Russes, la fortune du combat...

Nous avons saisi le combat, du départ à la victoire; et le retour triomphal à la cour du Tsar.

**2.4 А.С. ПУШКИН (1799-1837): Ode: Évocation dans le parc de Tsarskoe Selo (1814); récit: La demoiselle paysanne (1834): {πvс1 πvс2 πB1... πB9 πBa πBb}**

Tandis que les salons de Petersbourg se partageaient entre Derjavinistes et Karamzinistes, POUCHKINE, habitant du XIX-ème siècle, mais épris de la libertine liberté du XVIII-ème, reçut, sans embarras, le double héritage.



2.4.1 Élève à 15 ans, dans le noble Lycée qui confine au parc du Versailles russe, POUCHKINE évoque, devant DERJAVINE attendri d'admiration, l'image de la grande Catherine; puis la guerre, qui gronde encore: bataille de la Moskova; incendie de Moscou; retraite de Napoléon dans les neiges... et jusqu'à Paris:

В Париже росс! – где факел мщенья?  
 Поникни, Галлия, главой.  
 Но что я вижу? Росс с улыбкой примиренья  
 Грядет с оливою златой.

*Le Russe est dans Paris! Torche de la vengeance!  
 Gaule, courbe la tête.  
 Mais quoi? Sourire et Paix:  
 Le Russe au rameau d'or, s'avance.*

2.4.2 Au sentimentalisme, POUCHKINE fit, en russe, le même parti que MUSSET lui fit, en français.

Héritiers de deux voisins que l'on peut croire irréconciliables, *Liza* et *Alexis* se rencontrent secrètement: celui-ci, tient son propre rôle de gentilhomme; celle-là, a pu satisfaire à la curiosité de connaître son voisin en se parant, sous le nom d'*Akulina*, des atours d'une paysanne. À la fortune d'une rencontre, les pères se réconcilient. Ils entreprennent de marier leur progéniture: *Alexis* regimbe, s'apprête à déclarer son projet de fuire avec *Akulina*: quand il découvre que celle-ci est la *Liza* qu'on lui destine.

On ne résume pas ces épisodes, ces dialogues où tintent, entre les chants des merles, un écho des illusions contradictoires qui ont bercé plusieurs générations d'européens. Nous dirons seulement que, tandis que la *Nathalie* de KARAMZINE a la blancheur des lys, *Liza* est une brunette à qui sied la vive couleur du Sarafan rustique; et que POUCHKINE s'arrête net sur ces mots:

Читатели избáвят меня́ от изли́шней обяза́нности  
 опи́сывать разв́язку.

*Les lecteurs me dispenseront de l'inutile obligation de  
 décrire le dénouement.*

mais KARAMZINE, ayant écrit:

Читатель вообразит себе все последующее.

*Le lecteur imaginera tout ce qui s'ensuivit.*

charge encore d'une page de détails l'imagination du lecteur.

**2.5 В. И. ЛЕНИН (1870-1924) : Thèses de philosophie et de politique:**  
{LМа... LMe LGa... LGe}

**2.5.1 LÉNINE: Matérialisme et Empiriocriticisme;** 1-ère édition: 1908; 2-ème édition: Septembre 1920; nous utilisons une édition du Госполитиздат, 1946: {LМа... LMe}

Du chapitre IV: “Les philosophes idéalistes, alliés et héritiers de l’Empiriocriticisme”, on a saisi la majeure partie du §1: “Critique du kantisme du point de vue de la gauche et de la droite”.

Ces titres indiquent bien, que, sans considérer en son fond la critique des impressions sensorielles et de leur interprétation en terme d’objets, dont des observations contradictoires peuvent faire ultérieurement douter...; LÉNINE décrit un champ de bataille politique.

**2.5.2 LÉNINE: L’État et la Révolution ;** 1-ère édition: Août 1917; 2-ème édition: Décembre 1918; nous utilisons une brochure du Госполитиздат, 1947: {LGa... LGe}

NB le 2-ème fragment, LGb, consiste essentiellement en longs passages, traduits de K. MARX.

On a saisi les §§1 & 2 du chapitre V: “Bases économiques du dépérissement de l’État”. Une citation suggérera au lecteur le sens idéal qu’a, pour un marxiste, ce terme de dépérissement:

“И только тогда демократия начнет отмирать в силу того простого обстоятельства, что, избавленные от капиталистического рабства, от бесчисленных ужасов, дикостей, нелепостей, гнусностей капиталистической эксплуатации, люди постепенно привыкнут к соблюдению элементарных, веками известных, тысячелетиями повторявшихся во всех прописях правил общежития, к соблюдению их без насилия, без принуждения, без особого аппарата для принуждения, который называется государством.

*“Et seulement alors la démocratie [l’État démocratique de la société communiste] commencera à dépérir, de par ce simple fait que, délivrés de l’esclavage capitaliste, des terreurs sans nombre, de la barbarie, de l’ineptie, de l’ignominie, de l’exploitation capitaliste, les gens s’habitueront graduellement à observer les principes de la vie sociale connus depuis des siècles, répétés depuis des millénaires dans toutes les prescriptions, à les observer sans contrainte, sans obligation, sans subordination, sans cet appareil spécial de contrainte, qu’on appelle l’État.*

Le lecteur fera, par lui-même, l’application de ce projet aux États contemporains, russe et non russes.

**2.6.1 Александр Б ЛОК (1880-1921): poèmes: {Бскf B12a B12b B12c}**

**2.6.1 BЛОК: les Scythes (20 Février 1918): {Бскf}**

Au temps d'Hérodote et de Plutarque, les Scythes, Скифы ou plutôt Скѳы, Σκύθαι (cf. *supra* §1.4), tribus nomades Iraniennes, chevauchent dans la steppe, au nord de la Mer Noire; et Athènes recrute parmi eux ses policiers. Héritiers des Scythes menaçants, mais en faction à l'Est de l'Europe et épris de ses trésors, tels sont les Russes:

Мы лю́бим всё – и жар хо́лодных числ,  
И дар божѣственных видѣний,  
Нам вня́тно всё – и о́стрый га́льский смысл,  
И сýмрачный герма́нский гѣний...

*Tout nous est cher - la brûlante froideur des nombres,  
Et le don des visions célestes,  
Nous avons tout saisi - le sel de l'esprit français,  
Et le sombre génie allemand...*

Ainsi parle BЛОК; avec en dernière strophe cette invitation, fraternelle et menaçante:

В послѣдний раз – опóмнись, стáрый мир!

*Écoute, écoute encor, monde de la vieillesse:  
Une dernière fois tu peux te souvenir...*

**2.6.2 BЛОК: les douze (3 Mars 1918): {B12a B12b B12c}**

Par une nuit de l'hiver de 1918, une patrouille de douze soldats de la garde rouge marche dans les rues enneigées de Pétrograd. Au moindre signe insolite, répond une raffale de projectiles.

L'un des gardes, *Petia*, aperçoit *Katia* qu'il aime; et avec elle un homme, *Vanka*; il vise:

Трах-тарарáх! Ты б́удешь знать,  
Как с дѣвочкой чужо́й гуля́ть!...

*trac-tararac! ça t'appendra  
à sortir avec la fille d'un autre!...*

*Vanka* tombe; et *Katia* avec lui. Le garde s'en émeut quelque temps... Sans doute *Katia* représente-t-elle pour le poète, tous ceux que l'histoire foule sous ses pas...

La patrouille, enveloppée de flocons, avance toujours, tirant au hasard.

Le poème se termine sur une vision, qui justifie le choix du nombre apostolique de 12.

...Так идёт державным шагом -  
 Позади - голодный пёс,  
 Впереди - с кровавым флагом,  
 И за выюгой невидим,  
 И от пули неведим,  
 Нежной поступью надвьюжной  
 Снежной россыпью жемчужной,  
 В белом венчике из роз -  
 Впереди - Исус Христос.

..Ils marchent: marche impérieuse -  
 Derrière eux - un chien affamé;  
 Devant - avec un drapeau de sang,  
 Invisible au travers des flocons,  
 Inaccessible aux balles,  
 Tendresse et douceur dans son pas,  
 La neige le pare comme de perles;  
 Couronné de blanches roses -  
 Devant eux - c'est Jésus Christ.

**2.7 Г.Н. ПОВАРОВ : Avant-propos à la traduction russe de l'ouvrage de Edmund C. BERKELEY: Symbolic logic and intelligent machines (Moscou:1961): {loga logb logβ logc}**

“Будет ли “полная мозговая машина” (и снабженный ею автомат) обладать разумом, сознанием, жизнью или, несмотря на все свое функциональное сходство с живым мозгом, она останется его мертвой, хотя и страшно сложной и подвижной, моделью? Или, быть может, “полная мозговая машина” будет обладать лишь разумом и сознанием, но не жизнью?

Pour POVAROV, ces “*machines intelligentes ne sont que des outils mentaux*”.

Imaginant une “machine cérébrale complète” (*couplée à un automate*) il se demande si celle-ci, “*pour semblable qu'elle soit, quant à la fonction, à un cerveau vivant, serait douée de raison, de conscience, de vie...*”.

Sans prononcer les mots de matérialisme et d'idéalisme, l'auteur donne à ce qu'il appelle les “doctrines philosophiques bourgeoises” une réplique que nous qualifierions de spiritualiste.

**2.8 А.Х. КАСЫМЖАНОВ : Le problème de l'identité de la dialectique, de la logique et de la théorie de la connaissance; (d'après les Cahiers philosophiques de V. I. LÉNINE) ; (Édition de l'Académie des Sciences du Kazakhstan; Alma-Ata; 1962): {KJa... KJd}**

De la Partie I: “L'analyse par Lénine des prémisses théoriques de la dialectique de Marx”, on a saisi le début du chapitre I: “Dialectique, théorie de la connaissance et logique dans la philosophie de KANT”.

Recueil de longs passages, copiés chez HEGEL, FEUERBACH, et d'autres auteurs allemands ou russes, et brièvement annotés par LÉNINE, les fameux "Cahiers" ont servi en URSS d'anthologie d'auteurs non réédités. Sous l'autorité de MARX, ENGELS et LÉNINE, KACEMJANOV prône l'étude de HEGEL; et donc de KANT, dont est parti HEGEL.

"Вопреки имевшему место в нашей литературе нигилистическому отношению к Гегелю следует отметить, что Маркс и Энгельс, а вслед за ними и Ленин, высоко ценили гегелевскую философию, ее диалектику.

*"En dépit de l'attitude négative (littéralement: nihiliste) qui est celle de notre littérature vis-à-vis de HEGEL, il faut noter que MARX et ENGELS, et LÉNINE à leur suite, ont hautement prisé la philosophie de HEGEL, sa dialectique."*

On a voulu comparer à celui de LÉNINE le discours de KACEMJANOV sur KANT.

**2.9 Валерий СЕРГЕЕВ: Biographie du peintre André Roublov (Moscou: Collection des Biographies d'hommes illustres; 1981): En contemplant la Sainte Trinité {†Ra... †Rf} ; avec un avant-propos de l'Académicien Л.С. ЛИХАЧЕВ: {†RL}**

L' Icône de la Trinité, mille fois reproduite, est aujourd'hui sous les yeux de tous. Valère Serguéiev décrit, ou plutôt imagine, ce que fut le travail du moine André ROUBLOV, contemplant la lumière du Ciel, au delà de celle de la terre:

"Прозрачны и легки весенные дни. По утрам в иной из дней яркое солнце, свежий еще холодок, на ветру шумит молодая листва. После общей трапезы – сразу за работу.

*"Transparents et légers les jours de printemps! Certains matins, le soleil est éclatant, le froid reste vif, les jeunes feuilles frissonnent dans le vent. Sitôt finie la messe conventuelle, André est à l'ouvrage."*

Et, avec un commentaire théologique aussi suggestif qu'il est précis, l'auteur décrit l'épisode biblique, sa signification et la représentation, lumineuse et dépouillée, qu'en a donnée Roublov:

"И вот однажды, когда сидел он на пороге своего дома в Мамбрийской дубраве, в полуденный зной явился ему сам Бог. Невидимое, непостижимое, не имеющее образа божество, для общения с человеком принявшее вид трех путников.

*“... un jour - Abraham est assis à l'entrée de sa demeure, dans la chesnaie de Mambré - sous le soleil brûlant de midi, Dieu lui-même lui apparaît. La divinité invisible, inaccessible, celle qui n'a point de figure,... a pris, pour communiquer avec l'homme, la forme de trois voyageurs.*

“в образах трех ангелов была явлена миру тайна троического божественного единства - “Троица единосущная и нераздельная”.

*“sous la forme des trois anges a été révélé au monde le mystère de l'unité trinitaire de la divinité: de la “Trinité consubstantielle et indivisible”.*

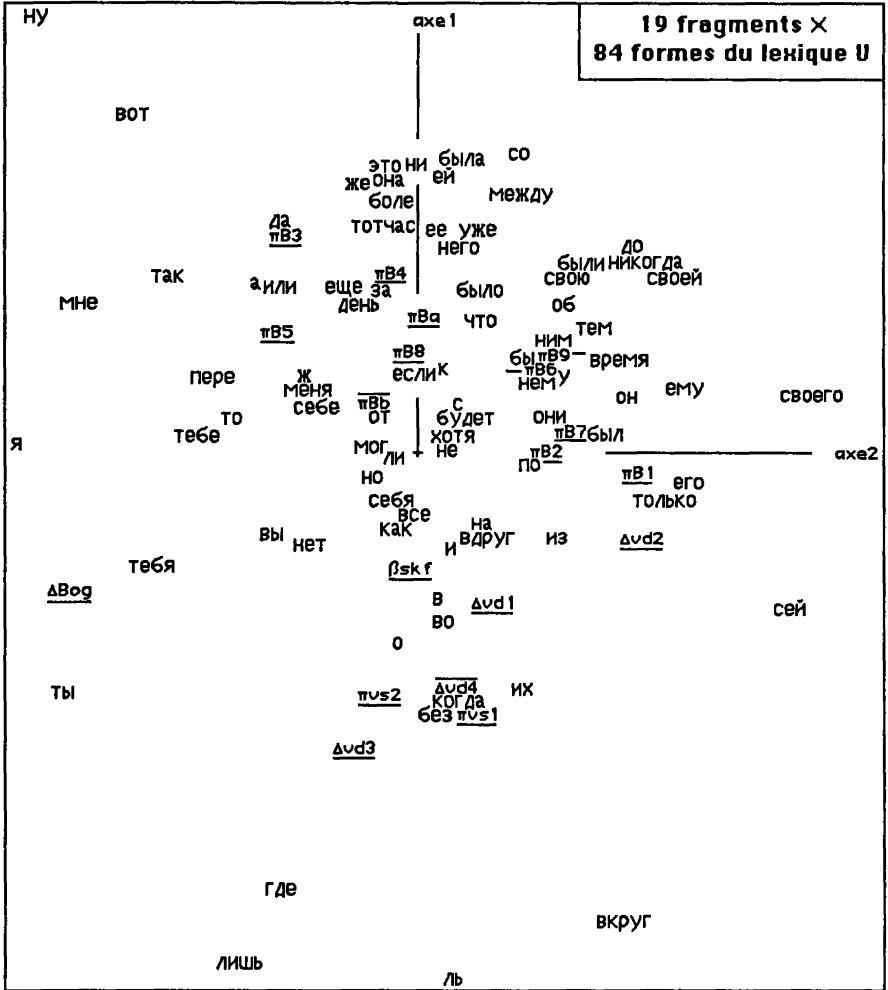
“Был день, когда Рублев окончил “Троицу”. Несказанным, весенним и тихим сиянием засветились ее краски. Наверное, старый Андрей чувствовал - в этом творении вершина и итог его жизни.

*“Vint le jour où Roublov acheva la “Trinité”. Les couleurs resplendissaient d'un éclat indicible - à la fois printanier et paisible. Le vieil André sentit que cette œuvre était le zénith et le bilan de sa vie.*

### 3 Élaboration et analyse des textes

#### 3.0 Les programmes utilisés

Aux particularités près du numérotage des caractères russes (cf. §§1.3 et 1.4) le programme ‘triruss’ ne diffère pas des programmes utilisés pour d'autres langues afin de créer des listes triées alphabétiquement où chaque forme occupe une ligne de rang impair, suivie, sur une ligne paire, de son étiquette par chapitre et verset. L'élaboration se poursuit sans distinction de langue, le tri alphabétique se faisant suivant l'ordre des octets, sans référence à la valeur linguistique de ceux-ci. Ici, on a utilisé quatre programmes ‘qamus’, ‘trimu§’, ‘trimu’ et ‘tridic’: ‘qamus’ crée le dictionnaire du texte, liste triée où chacune des formes distinctes attestées dans le texte se rencontre une fois, occupant une ligne, avec, sur la ligne précédant chaque forme, le nombre des occurrences de celle-ci dans le texte; ‘trimu§’ trie le dictionnaire dans l'ordre des fréquences croissantes, en une liste où, chaque forme est sur une ligne, avec sa fréquence sur la ligne précédente. D'après ce dictionnaire complet trié, on choisit le lexique des formes retenues afin de créer un tableau croisant formes et fragments; ‘trimu’ donne, dans l'ordre alphabétique, la liste des formes retenues. À partir des fichiers triés du texte et du lexique, et d'un fichier qui spécifie le découpage du texte en fragments affectés chacun d'un sigle, ‘tridic’ crée le tableau de correspondance entre formes et fragments.



### 3.1 Analyse préliminaire

84 mots de U X 19 fragments

trace : 9.243e-1

rang : 1 2 3 4 5 6 7 8 9 10

lambda : 1687 1260 924 788 649 526 463 426 398 359 e-4

taux : 1825 1364 1000 852 702 569 501 461 430 388 e-4

cumul : 1825 3189 4189 5041 5743 6312 6813 7274 7704 8093 e-4

L'analyse porte sur un tableau 84 x 19. Avec des données restreintes, on peut, sur un graphique complet, placer, à la fois, toutes les formes et tous les fragments.

L'ensemble J des textes comprend les 8 fragments de 4 poésies:  $\beta skf$ , les Scythes de BЛОК (cf. supra §2.6.1); { $\pi vs1 \pi vs2$ }, ВОСПОМИНАНИЯ В Царском Селе, par POUCHKINE, Évocation dans le parc du palais (cf. supra §2.4.1);  $\Delta Bog$ , БОГ, Ode théologique de DERJAVINE; { $\Delta vd1... \Delta vd4$ }, l'Ode Водопад, où le même poète, dans le fracas d'une *Cascade*, évoque le destin de chefs de guerre qui illustrèrent le règne de la grande Catherine; (de cette Ode, pour POUCHKINE le chef d'œuvre de DERJAVINE, la composition est manifestement reprise dans l'Évocation) ; et les 11 fragments de prose { $\pi B1... \pi B9 \pi Ba \pi Bb$ } en lesquels a été subdivisé le récit de POUCHKINE: *La Demoiselle paysanne* (cf. §2.4.2).

Le lexique U compte 84 formes: mots outil: conjonctions, adverbes de temps, pronoms (y compris ceux des personnes 1 et 2); ainsi que les noms {ВРЕМЯ: temps; ДЕНЬ: jour} dont l'emploi est aussi général que celui d'adverbes de temps {УЖЕ: déjà; НИКОГДА: jamais}; et la forme verbale (МОГ: pourrait).

Sur l'axe 1, les fragments de poésie ont tous une coordonnée F1 nettement négative. La prose est du côté ( $F1 > 0$ ); les fragments { $\pi B1 \pi B2 \pi B7 \pi Bb$ }, qui ne contiennent pas de dialogue entre les deux jeunes gens, ou entre *Liza* et sa servante et confidente *Nastia*, sont proches de l'origine;  $\pi B1$ , présentation en forme des deux pères de famille qui s'affrontent, étant seul du côté ( $F1 < 0$ ).

Du côté ( $F1 < 0$ ) on trouve non seulement des formes fréquentes en poésie mais rares en prose (ВКРУГ: alentour; et non ВОКРУГ: autour); mais aussi {ГДЕ: où; КОГДА: quand} qui ouvrent des perspectives dans l'espace et le temps; et (ЛИШЬ: seulement) par lequel le poète montre dans cette perspective l'exception sur laquelle se porte son regard.

Avec la prose ( $F1 > 0$ ) sont les formes qui rendent la vivacité du mouvement ou du dialogue: {ТОТЧАС: tout-de-suite; ВОТ: voici; ЭТО: ce, ça}

Sur l'axe 2 négatif, se détache l'Ode  $\Delta Bog$ : DERJAVINE se place devant la majesté divine, pour l'invoquer dans une méditation théologique, où les pronoms (je:Я), (tu:ТЫ ТЕБЯ ТЕБЕ) représentent la créature et le Créateur. C'est pourquoi les fragments { $\pi B3 \pi B5$ }, dialogues de *Liza* avec *Nastia* ou avec *Alexis*, tendent aussi vers ( $F2 < 0$ ), attirés par ces mêmes pronoms.

Est confirmé le fait, déjà remarqué ailleurs, qu'en comptant dans le lexique les pronoms des personnes 1 et 2, on établit, entre fragments, des proximités qui ne procèdent, directement, ni du style ni du genre.

De la САН, nous retiendrons seulement l'association entre  $\Delta Bog$  et les mots: {себя то тебе мог я без лишь тебя ты}.



### 3.2 L'analyse principale et ses variantes: stabilité du plan (1 × 2)

146 mots de S (sans 'Je' ni 'Tu') × 62 fragments

trace :	1.563e+0									
rang :	1	2	3	4	5	6	7	8	9	10
lambda :	1277	1110	716	701	653	554	520	483	458	435 e-4
taux :	817	710	458	449	418	354	333	309	293	279 e-4
cumul :	817	1527	1985	2433	2851	3206	3539	3848	4141	4419 e-4

151 mots de Z (sans 'Je' ni 'Tu') × 62 fragments

trace :	1.636e+0									
rang :	1	2	3	4	5	6	7	8	9	10
lambda :	1272	1138	798	721	676	551	534	493	476	461 e-4
taux :	777	695	488	441	413	337	326	302	291	282 e-4
cumul :	777	1473	1961	2401	2814	3151	3477	3779	4070	4352 e-4

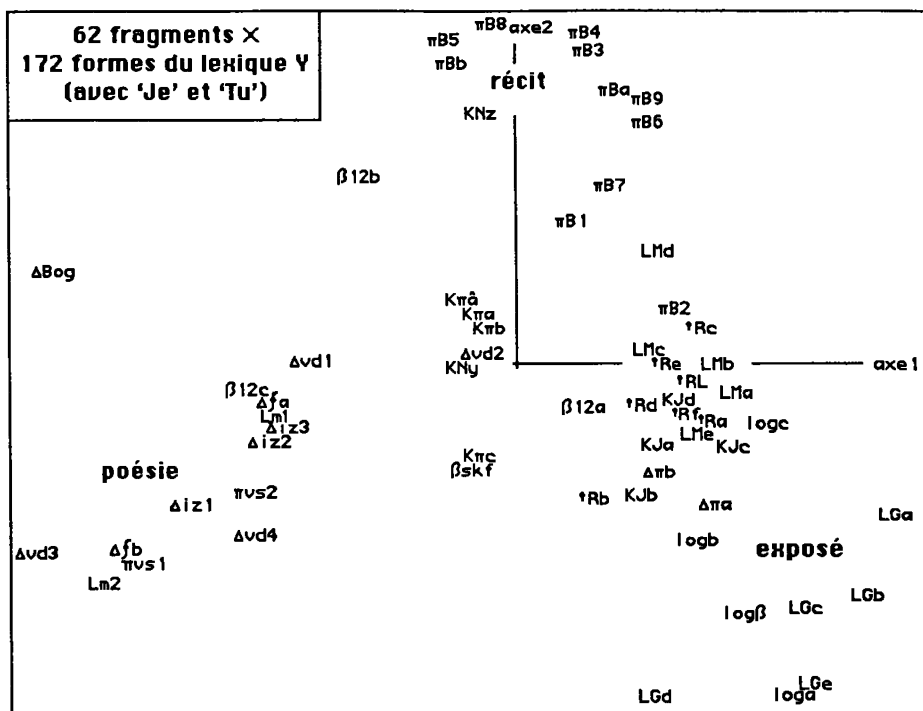
172 mots de Y (avec 'Je' et 'Tu') × 62 fragments

trace :	1.827e+0									
rang :	1	2	3	4	5	6	7	8	9	10
lambda :	1606	1207	883	779	721	680	583	559	529	490 e-4
taux :	879	661	484	426	395	372	319	306	289	268 e-4
cumul :	879	1540	2023	2449	2844	3216	3535	3841	4131	4399 e-4

Par analyse principale, nous entendons analyse croisant l'ensemble des 62 fragments de textes, présentés au §2, avec l'un des trois lexiques de mots outil: {S Z Y}, comptant respectivement 146, 151 et 172 formes: celles des pronoms *je* et *tu* figurant dans Y, mais non dans S et Z; tandis que les lexiques S et Z, créés à différentes étapes de la saisie du corpus, diffèrent seulement en ce que Z compte 15 formes absentes de S; et S, 10 formes absentes de Z. La différence, pourtant faible, entre S et Z, montrera son effet au §3.3, sur la classification des fragments. (On trouvera, au §3.4, sur le tableau d'une classification, la liste des formes du lexique S; et, en NOTE finale, le bilan des différences entre les lexiques S et Z.)

Un tel lexique ne diffère du lexique U, considéré au §3.1, qu'en ce qu'avec un corpus plus étendu, on a pu conserver un plus grand nombre de formes (le seuil de fréquence choisi étant 12). Par exemple, on a signalé dans U la forme (МОГ: pourrait); s'y ajoutent, dans les lexiques {S Z Y} les deux formes {МОЖЕТ: peut; МОЖНО: possible}, construites sur la même racine et servant à exprimer la même notion dans des contextes différents. Dans les trois analyses, la décroissance des valeurs propres est lente, avec moins de la moitié de l'inertie dans l'espace engendré par les 10 premiers axes. Mais la deuxième valeur propre est nettement séparée de la troisième, ce qui autorise à considérer pour lui-même le plan (1,2).

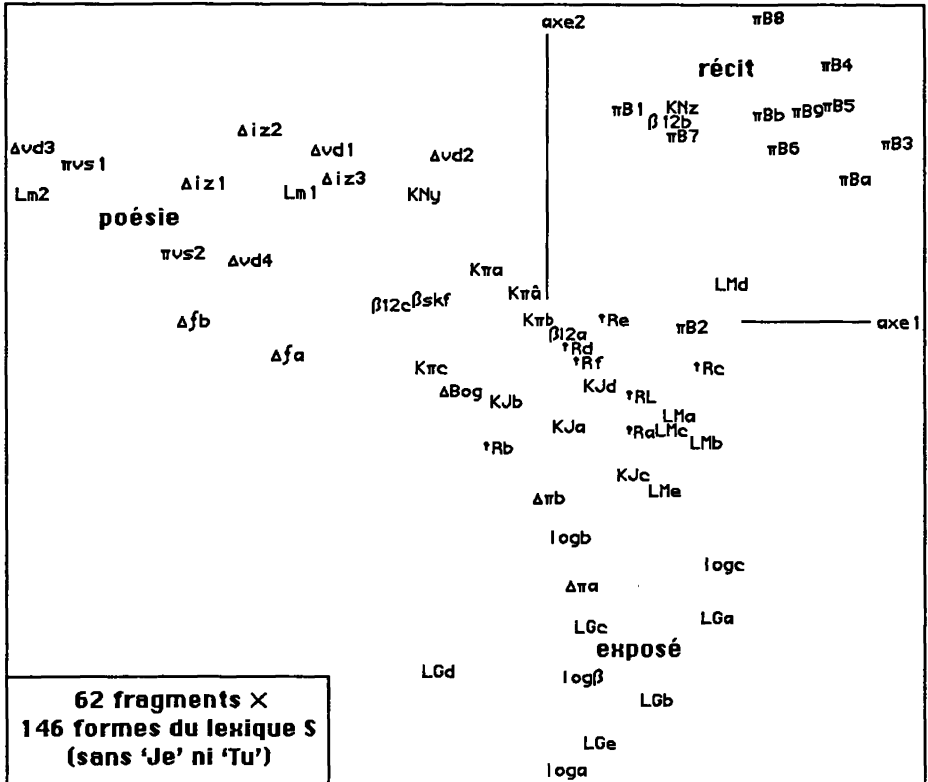
L'interprétation de ce plan est claire; au delà nous ne reconnaissons pas de dimension qu'on puisse, d'après notre corpus, interpréter de façon sûre en terme de genre et de style; même si, d'autre part, (cf. §§3.3 et 3.4), les axes de rang supérieur à 2 contribuent à rendre compte, de façon cohérente, de la répartition des fragments en œuvres.



Dans les trois analyses, le nuage des 62 fragments, présente, dans le plan (1,2), une figure à trois branches, étiquetées {poésie récit exposé}; la poésie, étant essentiellement représentée par des odes, dans notre corpus. Seule varie l'orientation des branches relativement aux axes. Avec S et Z, on a, sur l'axe 1 poésie opposée à récit et exposé; celui-ci étant proche du centre.

L'introduction de {Je Tu} dans Y accroît l'écart entre l'exposé et les autres genres littéraires, particulièrement la poésie lyrique; sur l'axe 1, poésie s'oppose à exposé, le récit étant proche du centre. Les formes de {Je Tu} sont, sur l'axe 1, du même côté que la poésie lyrique; elles s'étalent suivant l'axe 2: {Я:je МОЙ:mon ТЕБЕ,ТЕБЯ:toi} sont du côté du récit; les formes {ТВОЕЙ ТВОИХ ТВОЙ} de l'adjectif possessif *ton* s'y opposent; (Ты:tu) est à l'origine sur l'axe 2.

Dans l'analyse fondée sur le lexique Y, les formes de {Je Tu} apportent à l'axe1 environ un tiers de son inertie, créant une nette dénivellation entre les deux premières valeurs propres. Cependant, la figure à trois branches n'est guères modifiée, même dans le détail des positions relatives des divers fragments: ce qui prouve qu'il s'agit d'une structure qui n'est pas créée par l'absence de *Je* et *Tu* dans nos exposés; non plus que par le tutoiement

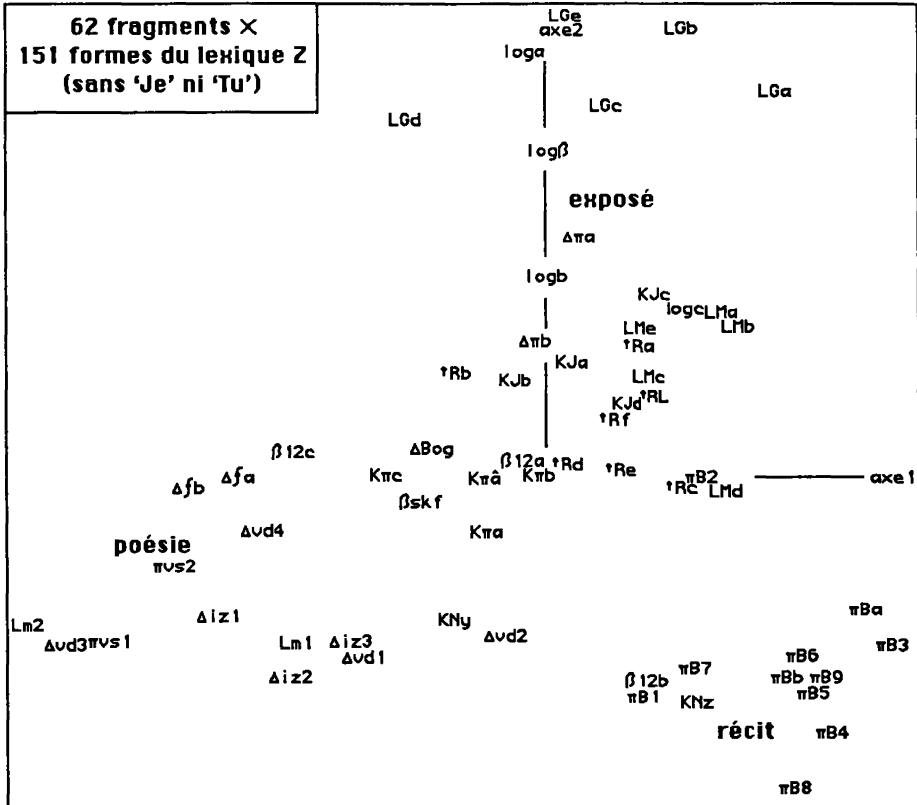


emphatique propre à la poésie lyrique (cf. *supra*, §2.2.2, DERJAVINE tutoyant la grande Catherine).

Reste à considérer dans ses détails la distribution des fragments; ce que nous ferons sur le plan (1,2) issu de l'analyse fondée sur le lexique S.

Dans la branche de la poésie, on trouve les fragments de cinq des six odes (une de LOMONOSOV, une de POUCHKINE, quatre de DERJAVINE); toutefois, (comme au §3.1), le fragment Δvd2 de la Cascade tend vers le récit: Δvd2, sans fracas héroïque, fait passer en un songe, devant ROUMIANTSEV, l'image de ses victoires, de sa disgrâce et de la mort de POTEMKINE, en faveur de qui il avait été évincé. Et ΔBog, 'Dieu', ode théologique, tend vers l'exposé.

Aucun des fragments de BLOK, n'a de position écartée dans la branche lyrique: toutefois les *Scythes* (βskf: cf. §2.6.1) et la fin mystique des *Douze* (β12c: cf. 2.6.2) sont dans cette branche: mais β12a, écho syncopé de discours politiques des soviets de Péetrograd, confine à l'exposé; et β12b, passage de *Katia* et propos mélancoliques de la patrouille dans la neige, se place dans le



récit; branche constituée de 10 sur 11 des épisodes de la *Demoiselle paysanne* (cf. §2.4.2); dont seul se sépare πB2, présentation d'*Alexis* assaisonnée de propos de sociologie, qui va vers l'exposé.

Dans la branche de l'exposé, on trouve non seulement des fragments de LÉNINE ou de contemporains traitant de la philosophie (cf. §§2.7 et 2.8); mais aussi, de deux siècles antérieur, {Δπα Δπα}, le discours de DERJAVINE sur la poésie lyrique (cf. §2.2.1). Un fragment de *Matérialisme et Empiricriticisme* (cf. §2.5.1), LMd, tend vers le récit: LÉNINE y montre MACH et AVENARIUS relevant chez KANT une "connivence avec le réalisme"; mais il ne développe pas d'arguments philosophiques.

Proche de l'origine est la biographie de ROUBLOV (cf. §2.9); ainsi que la correspondance de KARAMZINE (cf. §2.3.1) dont l'enthousiasme confine au lyrisme; mais s'interrompt parfois pour des commentaires sociologiques. Du même auteur, les fragments de *Nathalie fille de Boïar* (cf. §2.3.2) sont l'un en poésie (KNy: bataille), l'autre dans le récit (KNz: retour à la cour du Tsar).

Telle est l'image du plan (1,2): mais n'ayant, dans ce plan, que moins d'un sixième de l'inertie du nuage, on doit chercher dans une classification prenant en compte l'ensemble des facteurs une représentation fidèle des proximités immédiates entre fragments.

### 3.3 Variantes de la classification des fragments

Dans les classifications des fragments fondées sur les lexiques S et Y, on reconnaît, au sommet de l'arbre, le partage entre prose et poésie; avec Z, la partition en 3 classes, définie par les 2 nœuds les plus hauts, compte 1 classe de poésie, 2 classes de prose. Les exceptions à ces règles générales, peu nombreuses, (et dont plusieurs étaient attendues, d'après le plan (1,2),) sont signalées par des étoiles dans les tableaux de contenu des classes.

Nous présentons d'abord les CAH fondées sur Z et Y, puis celle sur S, étiquetée d'après les classes de formes

c	Partition en 8 classes d'après Z : Sigles des textes de la classe c										
112	*ΔBog	Δπa	κπc	*Δfa	logc	Δπb					(poésie) prose
109	κπb	†Rb	KNy	κπa	κπâ*β12a*βskf						prose
111	KJa	†RL	†Rf	†Re	†Rd	†Rc	†Ra	πB7	πB1		
114	KJc	LMe	KJd	KJb	πB2	LGa	LMa	LMc	LMd	LMB	
102	LGb	loga	logb	logβ							prose
115	LGc	LGe	LGd*β12c								
107	Δvd4	Δiz1	Δvd2	Δiz2	lml	Δiz3	Δvd1				poésie
103	πvs1	Δvd3	Lm2	πvs2	Δfb						
116	KNz	πBb	πBa	πB9	πB6	πB8*β12b	πB5	πB4	πB3		prose

112							121		122
109		113					119	120	
111									
114									
102							117		
115									
107							118		
103									
116									

**Avec le lexique Z:** Se sépare en haut de la hiérarchie la classe j116 qui contient avec 9 des 11 épisodes de la *Demoiselle paysanne* et le fragment β12b (*Katia*) des *Douze*. Se sépare ensuite la classe j118 qui contient l'ensemble de la poésie lyrique (de LOMONOSOV, DERJAVINE et POUCHKINE) à l'exception des deux fragments {ΔBog Δfa} (l'Ode théologique et le début de la Félicita: cf. §2.2.2). Reste j121, classe de prose; où sont trois fragments dus à BLOK {βskf β12a β12c}; et {ΔBog Δfa} qui sont dans la subdivision j112, avec le discours de DERJAVINE sur la poésie lyrique {Δπa Δπb}.

**Avec le lexique Y:** Comme sur le plan (1,2), KNY, le combat dans *Nathalie*, rejoint la poésie; tandis KNz, *retour à la cour*, est, comme  $\beta 12b$  (*Katia*), avec la *Demoiselle paysanne*. De la *Cascade*, le deuxième fragment,  $\Delta vd2$ , déjà remarqué dans le plan (1,2) va, avec ainsi que  $\beta 12a$ , dans la prose. Enfin,  $\beta skf$ , les *Scythes*, se lieut aux lettres de KARAMZINE (ce que nous ne tenterons pas d'expliquer). De l'étiquetage de la CAH, on donne seulement ce qui a rapport avec les trois classes de formes {i318 i327 i316} qui contiennent des pronoms des 1-ère et 2-ème personnes du singulier.

c   Partition en 15 classes d'après Y : Sigles des textes de la classe c	
1	$\Delta Bog$ i318+++++
31	$\beta 12c$ (la vision finale) i318+ i327+
90	$\Delta fb \Delta vd3$ i327++++ i316+++
109	$\Delta fa \pi vs2 *KNy Lm2$ i316+++++
105	$\pi vs1 \Delta iz1 \Delta iz3 \Delta iz2 \Delta vd4 Lm1 \Delta vd1$ i327++++
84	$\pi Ba \pi B9 \pi B6$
108	$*\beta 12b \pi B5 \pi B3 KNz \pi Bb \pi B4 \pi B8$ 318+
39	$K\pi c$
107	$K\pi b K\pi a K\pi a*\beta skf$ 318+
101	$Lm c Lm d Lm b Lm e Lm a$
93	$LGd LGe LGc$
99	$\pi B7 \pi B1*\Delta vd2$
106	$\dagger Rb*\beta 12a \dagger Ra \dagger Rf KJa \dagger RL \dagger Re \dagger Rd \dagger Rc \pi B2 LGa KJb KJd KJc$
96	$logb log\beta loga$
103	$logc \Delta\pi b LGb \Delta\pi a$
1	118
31	110 115 116
90	
109	
105	
84	113 122
108	
39	111 119 121
107	
101	
93	117 120
99	112
106	
96	114
103	
318	ЧТОБ ТОБОЙ МЕНЯ МНЕ СОБОЮ МОЙ Я j1+++++ j108+
327	се вокруг свой коль ль иль тебя ты j1+ j90+++ j105++
316	который тебе твоей твоих твой там где j90+++ j109++++

c	Partition en 12 classes d'après S: Sigles des textes de la classe c												
110	ΔBog	Δfa	β12c	Δfb	*†Rb				i269+	i270+	i153++++		
99	Lm2	Δvd3	πvs2	πvs1	Δiz1					i270+++	i153++++		
107	*Κπâ	βskf	Lm1	*Κπα	Δvd1	*Κπb			i269+	i239++	i274++		
97	Δvd4	Δiz3	Δiz2	Δvd2					i254++	i270++++			
103	LGe*β12a	LGc	LGd						i260++	i246+++++			
81	logb	logβ	loga									i273++++	
83	LMe	Lmb	Lmd	Lmc					i272+++++	i264+			
108	Κπc	Δπb							i172+++	i45++++			
111	logc	Δπα	LGb	LGa						i45++	i257+++	i273++	
105	†Rc	KJa	†Ra	†Rd	†Re	†Rf	LMa	†RL	πB2	KJd	KJb	KJc	i268++++
112	KNy	KNz	πB7	πB1	πB9	πB6			i254++	i264++	i271+		
109	πB8	πB4	πB3	πBb	πBa*β12b	πB5					i238++	i271++++	
110				116		118							
99													
107				113									
97													
103				115				120			122		
81													
83								119					
108				114									
111													
105						117			121				
112													
109													

Avec le lexique S: dans la classe j118 de la poésie, les seuls fragments de prose sont d'une part †Rb, ROUBLOV peignant dans sa cellule illuminée par l'éclat soudain du printemps russe (cf. §2.9); et d'autre part {Κπα Κπâ Κπb} (cf. §2.31), enthousiasme de KARAMZINE arrivant à Paris; le fragment {Κπc}, le plus chargé de commentaires sociologiques, restant avec la prose (ainsi que {KNy KNz}), les deux chapitres de *Nathalie*, mêlés, dans j112, à des épisodes de la *Demoiselle paysanne*).

Des *Douze* de BLOK, (cf. 2.6.2), seule la vision finale, β12c, est rangée avec la poésie (ainsi que βskf: les *Scythes*); tandis que β12a et β12b sont les seuls fragments de poésie dans la classe j122: β12a, écho de slogans politiques, va dans j103, avec des fragments de *l'État et la Révolution*; β12b, mort de *Katia* dans la neige, va dans j103, avec {πB3...}, conversations de la *Demoiselle Paysanne*.

Dans le tableau ci-dessus, les 12 classes de la partition retenue sont étiquetées suivant les classes de formes du lexique S; ainsi, on a noté, pour la classe j103: i260++ et i246+++++; parce que la classe j103 de fragments se distingue, au sein du corpus, par une fréquence élevée des formes de la classe i260; et une fréquence très élevée de celles de i246.

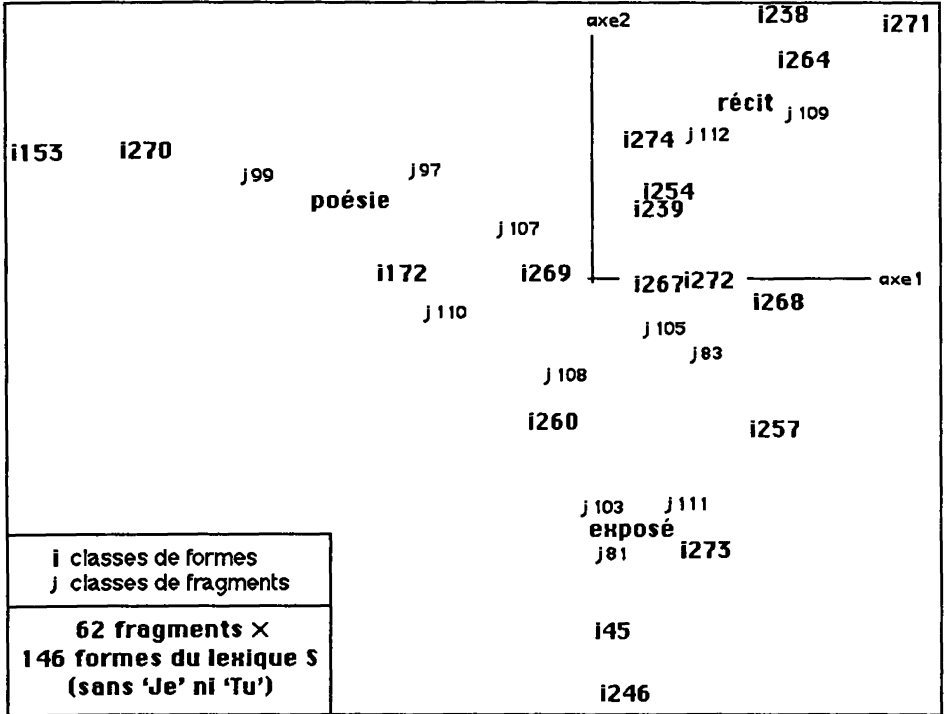
c	Partition du lexique S en 18 classes: formes de la classe c
267	a что совсем этой так не им как
268	никогда того об раз к от своих
269	теперь время со же той свое
272	всегда всего собой времени ней них это были была
239	нею очень тут своими свои все
269	о над их с на и во но здесь из в
272	эту до то себе вот за
274	нем сей ли
254	чтоб ж тотчас него какое одно вдруг
264	которого таким себя мог собою
264	своего у один сам всем по было уже ему тех был своей его
264	он свою столъ перед
238	ни уж ним хорошо другой ей
271	опять да она наконец ее
260	чтобы когда если будет ничего только без нет
246	оно при такой для
172	которых через всех
45	или
257	давно других которые даже самом конечно хотя бы
273	эти еще сейчас
270	может быть есть чем они между тем одна можно более
270	кто лишь который б свой коль пред под се вокруг ль
153	там где

267	_____	≈cdg	_____	279	_____	283	_____	286	_____	289	_____	290	_____
268	105++++	_____	_____	_____	_____	_____	_____	_____	_____	_____	_____	_____	_____
239	107++++	_____	_____	275	_____	_____	_____	_____	_____	_____	_____	_____	_____
269	_____	≈cdg	_____	_____	_____	_____	_____	_____	_____	_____	_____	_____	_____
272	83++++++	_____	_____	_____	_____	_____	_____	_____	_____	_____	_____	_____	_____
274	107++	_____	_____	_____	281	_____	_____	_____	288	_____	_____	_____	_____
254	97++	_____	112++	_____	276	_____	_____	_____	_____	_____	_____	_____	_____
264	83+	_____	112+++	_____	_____	_____	_____	_____	_____	_____	_____	_____	_____
238	109+++++	_____	_____	_____	278	_____	_____	_____	_____	_____	_____	_____	_____
271	109+++++	_____	112++	_____	_____	_____	_____	_____	_____	_____	_____	_____	_____
260	103++++	_____	_____	_____	280	_____	_____	_____	287	_____	_____	_____	_____
246	103+++++	_____	_____	_____	_____	_____	_____	_____	_____	_____	_____	_____	_____
172	_____	108+++++	_____	277	_____	_____	_____	285	_____	_____	_____	_____	_____
45	111+++	_____	81+	_____	108++++	_____	_____	_____	_____	_____	_____	_____	_____
257	111+++	_____	_____	_____	_____	_____	_____	_____	_____	_____	_____	_____	_____
273	111+++	_____	81++++	_____	_____	_____	_____	_____	_____	_____	_____	_____	_____
270	_____	99+++	_____	97+++	_____	_____	_____	_____	_____	284	_____	_____	_____
153	110+++	_____	99++++	_____	_____	_____	_____	_____	_____	_____	_____	_____	_____

### 3.4 Classification des formes du lexique S

De même, les classes de formes sont étiquetées par les classes de fragments.





On trouvera, dans le plan (1,2), une représentation simultanée des deux nuages de centres qui ne montre que partiellement les proximités existant dans l'espace multidimensionnel et traduites par l'étiquetage.

Dans la partition de S en 18 classes, partition définie par les 17 nœuds les plus hauts, deux classes, marquées '≈cdg' se distinguent par leur proximité au centre de gravité du nuage, visible sur le plan (1,2) et s'étendant à toutes les dimensions:

la classe i267, avec 18% de la masse totale (i.e. de l'ensemble des occurrences des formes retenues dans S), n'apporte à l'inertie globale que 0,4%;

la classe i269, avec 35% de la masse ne contribue que pour 0,7% à l'inertie; des onze formes que compte i269, sept sont des prépositions très fréquemment employées;

de ces deux classes, seule i269 est mentionnée dans l'étiquetage des fragments: et elle y contribue très peu.

Sans prétendre expliquer exhaustivement les affinités entre formes et fragments, nous ferons quelques remarques

Dans tous les fragments de LM, *Matérialisme et empiriocriticisme*, de LÉNINE, la forme себе (*soi* au cas prépositionnel) est particulièrement fréquente: il s'agit généralement de вещь в себе ou d'emplois analogues, traduits de l'Allemand *Ding an sich*. D'où l'affinité mutuelle entre i272 et j83.

Les formes {там где}, {là-bas où}, servent au poète lyrique pour évoquer les lointains, l'espace: ainsi la classe i153 est-elle associée aux fragments de j110 et j99.

Est également associée à la poésie lyrique la classe i270 où sont, notamment, {перед вокруг} {devant, avant; autour}, formes slavonnes, dont les équivalents usuels sont {перед вокруг}; au §3.1, on a proposé, comme une analogie avec le français: (вкруг: alentour; вокруг: autour). La forme {перед}, qu'on a, elle aussi, retenue pour le lexique S, est rangée dans la classe i264, associée à j112: fragments de la *Demoiselle paysanne* de POUCHKINE, et de la *Nathalie* de KARAMZINE.

La conjonction de coordination {или}, {ou}, constitue, à elle-seule, la classe i45, associée aux fragments de j111, j108 (voire j81), où la dialectique joue sous diverses formes.

NOTE: L'ordre de décroissance des fréquences fluctue avec la composition du corpus: d'où, entre les lexiques S et Z, créés à des étapes différentes de la saisie des textes, des différences fortuites. Nous donnons ci-après le détail de ces différences (dont le bilan a été annoncé au §3.2):

*formes de Z qui ne sont pas dans S*

{будто вместе вперед всей должно именно которая которое несколько первый самым своими том тот этого}

*formes de S qui ne sont pas dans Z*

{вкруг даже других какое которого нею ничего о одно своим}

On remarquera que sont dans Z {которая которое}, formes du pronom *lequel* au nominatif singulier pour les genres féminin et neutre; tandis qu'on a dans S {которого}, génitif singulier du même pronom pour les genres masculin ou neutre. Les formes {который которые}, nominatif masculin singulier et nominatif pluriel, sont dans les deux lexiques.

Quant au possessif renvoyant au sujet (sans distinction de personne) on a dans Z {своими}, instrumental pluriel; et dans S {своим}, instrumental singulier (neutre ou masculin); mais de nombreuses formes sont communes à S et Z: {своих свое свои своей свою свой своего...}.

#### 4 Conclusion: comparaisons avec d'autres langues

On a vu au §3.2 que le plan (1, 2) offre du corpus une image globale qui ne dépend pas du choix du lexique. Quant à la classification des fragments, elle produit des subdivisions stables, dont seul varie l'ordre d'agrégation au niveau supérieur (cf. §3.3). S'appliquent donc aux textes russes les méthodes taxinomiques déjà appliquées ailleurs. Ceci nous incite à comparer les diverses langues d'un point de vue statistique général.

##### 4.1 Fréquence relative des outils et longueur des mots

Dans [TEXTES LATINS] (§1.3.3, Vol.XVI, n°4, 1991), on a introduit, à côté des nombres d'occurrences de mots d'un lexique, deux variables:  $\pi_{\text{mot}}$ , nombre de mots (divisé par 10); et  $\pi_{\text{let}}$ , nombre des lettres (divisé par 100). Nous rendrons brièvement compte de l'analyse du tableau croisant, avec l'ensemble J des 62 fragments russes, l'ensemble  $S+ = S + \{\pi_{\text{mot}} \pi_{\text{let}}\}$ .

(146 mots de  $S + \{\pi_{\text{mot}} \pi_{\text{let}}\}$ )  $\times$  62 fragments  
 trace : 1.058e+0  
 rang : 1 2 3 4 5 6 7 8 9 10  
 lambda : 881 756 486 475 448 375 352 330 305 294 e-4  
 taux : 833 714 460 449 423 355 333 312 289 278 e-4  
 cumul : 833 1547 2007 2455 2879 3234 3567 3878 4167 4445 e-4

SIGI	QLT	PDS	INR	F 1	CO2	CTR	F 2	CO2	CTR	F 3	CO2	CTR	F 4	CO2	CTR
$\pi_{\text{ltr}}$	504	114	2	-51	150	3	-74	310	8	-21	25	1	18	19	1
$\pi_{\text{mot}}$	543	211	2	-66	469	10	14	22	1	-22	52	2	0	0	0

À elles deux,  $\{\pi_{\text{mot}} \pi_{\text{let}}\}$  ont un poids (PDS=114+211=325) quasi égal à la moitié de celui de S (PDS=1000-325=675). Et, puisque la contribution de  $\{\pi_{\text{mot}} \pi_{\text{let}}\}$  à l'inertie est voisine de zéro, l'inertie du nuage est diminuée d'un tiers (la trace passe de 1,56 à 1,06); ainsi que l'ordre de grandeur des valeurs propres.

Corrélation entre facteurs de même rang issus de  $S \times J$  et  $S+xJ$ :

rang	1	2	3	4	5	6	7	8
corr	.975	.980	.994	.994	.994	.983	.986	.990

La stabilité des facteurs est remarquable: entre facteurs de même rang issus de  $S \times J$  et de  $S+ \times J$ , considérés comme fonctions sur l'ensemble J des fragments muni du poids maginal de  $S \times J$ , les corrélations dépassent 0,975. L'interprétation des axes et plans n'est donc aucunement modifiée si l'on prend en compte  $\pi_{\text{let}}$  et  $\pi_{\text{mot}}$ .

Quant à l'interprétation, on note que  $\pi_{\text{mot}}$  tend vers la poésie (F1<0), où le nombre des mots pleins est grand, vis à vis du nombre des mots outils. Pour la même raison,  $\pi_{\text{let}}$  tend, lui aussi, vers la poésie (F1<0); mais, de plus,  $\pi_{\text{let}}$

tend vers l'exposé (F2<0): parce que l'exposé, notamment chez LÉNINE, fait grand usage du vocabulaire de l'idéologie, où prédominent les mots longs.

De façon précise, le taux des mots outil (mots de S, dénombrés par nous) relativement à l'ensemble des mots est d'environ 30%: il peut varier dans un rapport de 7 à 10 quand on passe d'une poésie de DERJAVINE à la prose de POUCHKINE; et, dans *Matérialisme et Empirocriticisme*, la longueur moyenne des mots dépasse de 6% celle (≈5,4) calculée sur l'ensemble du corpus.

Chez certains poètes du XX-ème siècle, tel Б. Пастернак, les mots outil sont encore bien plus rares que dans les odes, non dépourvues d'emphase réthorique, qui prédominent dans notre corpus russe.

En latin, on a déjà vu la relative rareté des outils chez les poètes: mais le corpus latin ne montrait pas de distinction interprétable entre la place de πlet et celle de πmot. Transmis de langue en langue au cours des siècles, alourdi, au sein d'une même langue, par composition et suffixation, le vocabulaire idéologique a perdu l'*imperatoria brevitatis* du latin!

### 4.2 Rôle et signification des mots outil les plus fréquents

Phédon	Évangile	Outils les plus fréquents dans six langues d'après des corpus déjà analysés	arabe	
170 ου	368 ου		55 wla	
151 μη	526 μη		57 hy	
.....	.....		60 aw	
254 οκ	1285 εν		70 hda	
331 δε	1366 του		78 anh	
334 η	1957 δε		84 ma	
349 το	2138 οκ		95 la	
1338 καλ	5659 καλ		104 cly	
			107 hw	
			183 fy	
			197 mn	
			205 an	
Latin	Russe		Fr: CAD	
285 ne	35 HET		1194 non	
300 neque			2525 pas	
501 nec	.....		2600 ne	
.....	152 a		.....	
750 quae	154 HO		6820 dans	
760 si	171 K		7012 on	
775 sed	177 OH	7417 du		
812 cum	180 eFO	7633 une		
854 qui	232 KaK	7830 un		
925 ad	288 C	7972 que		
1008 quod	296 CTO	8727 est		
1114 ut	349 Ha	9480 en		
1531 est	452 He	10985 d'		
1593 non	938 B	12657 à		
2306 in	1209 И	13641 et		
3707 et		14168 le		
		497 l'		
		15453 les		
		16206 des		
		17177 l'		
		19716 la		
		33781 de		
			Espagnol	
			336 ni	
			.....	
			1212 los	
			1223 por	
			1328 con	
			1805 no	
			2588 en	
			2614 el	
			2630 a	
			2996 la	
			4973 de	
			5058 y	
			5572 que	

Le sens d'un mot ne se conçoit, en toute précision, que d'après l'usage qui en est fait, concuremment avec celui des autres mots d'une langue. De ce point

de vue, on ne peut traduire aucun mot, si ce n'est un mot technique dont l'usage est imposé à tous. *A fortiori* ne peut on prétendre traduire les mots outil, dont le système est propre à chaque langue. Mais des équivalences approximatives suffisent à fonder une comparaison, entre les langues, quant au rôle et à la signification des outils les plus fréquents.

Des langues déjà étudiées, nous présumons, au §1.1, que la plus proche du russe était le latin: cette présomption se confirme sur le tableau de fréquence des outils. Dans les deux langues, vient en tête la conjonction de coordination {et И}, suivie de la préposition *dans* {in В} et de la négation usuelle (équivalente au *ne pas* du français) {non НѢ}; le poids relatif de la négation apparaît encore plus fort si l'on prend en compte d'autres formes {nec neque не НЕТ}.

Pour le français, on a considéré deux corpus, d'une part les articles de [CAD XII-XVII], d'autre part les textes divers analysés dans [TEXT. VAR. FR.] (in CAD, Vol.XVIII). Dans les deux cas, est en tête la préposition {de} qui, entre autres rôles majeurs, remplit celui dévolu au cas génitif, dans les langues à déclinaison. Après {de}, on trouve {et} ainsi que diverses formes de l'article (absent du latin et du russe); puis {à} dont le rôle est souvent celui du datif; et le verbe {est}, qui manque en russe mais est très fréquent en latin. La préposition {dans} a un poids moindre que celui de {in В}; les négations semblent moins employées qu'en russe et en latin.

En grec {καὶ} ≈ *et* se détache nettement; non seulement dans le *Nouveau Testament*, où l'on peut présumer l'influence de la phrase sémitique, mais aussi dans le *Phédon* de PLATON. La particule {δε} est une particularité du grec: associée à {μεν}, elle joue comme *d'une part avec d'autre part*, mais elle peut s'employer seule, comme pour souligner un mot. Le poids, rapporté à la conjonction de coordination {καὶ}, des principales négations {οὐ μὴ} est moindre que dans les autres langues.

On pourra être surpris que l'espagnol, langue romane, n'ait pas, avec le français, une similitude plus marquée. Au delà des Pyrénées, l'article n'est pas d'un emploi aussi général qu'en deçà; les négations {no ni} sont fréquentes, relativement à la coordination {y}; la première place revient à {que} dont les rôles sont plus étendus que ceux de notre *que*.

Nous ne citons que pour mémoire l'arabe: car (cf. [MOTS ARABES]), les dénombrements ont été faits, dans cette langue, en suivant la convention orthographique qui lie, au mot qui le suit, l'article {ال} ou le coordonnant {و}; mais il est satisfaisant, pour qui considère la tradition andalouse, de trouver au premier rang le {ان} arabe, comme le {que} espagnol; parallélisme qu'on retrouverait dans le tour de plus d'une phrase.