

Le lexique dans l'Œuvre poétique d'Aragon : Aragon, un poète bien de France

Il y a une dizaine d'années, j'ai décidé, avec l'aide de mon frère Jacques, de numériser la totalité de l'*Œuvre poétique* d'Aragon parue au Livre Club Diderot. Pour des raisons pratiques, nous avons dû nous arrêter à l'année 1952, ce que certains regretteront. J'ai ensuite confié notre travail à la banque de données FRANTEXT, qui n'en a traité, selon ses propres normes, que la première partie. Ayant connu divers changements de direction, elle a tout simplement rompu tout contact avec moi. Néanmoins, le public peut bénéficier des traitements lexicaux offerts par le centre de Nancy, pour ce qui concerne le premier tome de l'œuvre poétique, jusqu'au *Paysan de Paris* inclus, outre certains romans et recueils postérieurs, tels *Les Voyageurs de l'impériale* (saisie fournie par le Centre de recherche Hubert de Phalèse), *Le Crève-cœur*, *Le Roman inachevé*. Au total, il s'agit de 35.441 mots ou occurrences, tandis que j'en ai traité, pour ma part, un peu plus de 900.000.

Le corpus de l'Œuvre Poétique

On peut légitimement s'étonner de la composition de l'*Œuvre poétique* telle que l'a voulue Aragon lui-même, encore plus si on la compare aux œuvres poétiques proposées par la Bibliothèque de la Pléiade ! Le mélange des genres s'y pratique à souhait. Y figurent des textes que nous considérons comme des essais (*Une vague de rêves*), des romans (*Le Paysan de Paris*), du théâtre même (*Le Trésor des Jésuites*). Mais le *Traité du style* n'y est pas ! Le plus grave, pour nous qui nous intéressons à la langue poétique d'Aragon, est qu'il a inséré de longs textes de ses amis, qu'il m'a fallu écarter momentanément dans le cadre de ce travail pour maintenir la cohérence de

la voix. Reste que l'ensemble comporte bien le langage poétique d'Aragon, au sens où il l'entendait non seulement dans le temps de l'écriture mais encore sur la fin de sa vie.

Voici un tableau récapitulant les œuvres numérisées par mes soins, dont le premier volume seulement se trouve incorporé à la base de données textuelle FRANTEXT :

OP 1 1917-1926	p. 9-352	Livre I	1917-1920	<i>Feu de joie</i> <i>Écritures automatiques</i> <i>Les Aventures de Télémaque</i>
	p. 357-642	Livre I	1921-1925	<i>Le Mouvement perpétuel</i> <i>Une vague de rêves</i>
	p. 645-940	Livre III	1926	<i>Les destinées de la poésie</i> <i>Le Paysan de Paris</i>
OP II 1927-1935	p. 9-374	Livre IV	1927-1929	<i>La Chasse au snark</i> <i>La défense de l'infini</i> <i>La Grande Gaité</i> <i>Le Trésor des Jésuites</i>
	p. 375-804	Livre V	1930-1933	<i>La Peinture au défi</i> <i>Front rouge</i> <i>Persécuté persécuteur</i> <i>La Littérature internationale</i>
	p. 807-1296	Livre VI	1934-1935	<i>Hourra l'Oural</i> <i>Commune</i>
OP 3 1936-1941	p. 11-480	Livre VII	1936-1937	<i>La poésie soviétique à Paris</i> <i>Réalisme socialiste et réalisme français</i>
	p. 483-862	Livre VIII	1938	<i>Nlles du vaste monde</i>
	p. 867-1321	Livre IX	1939-1941	<i>Le Crève-cœur</i> <i>Les Yeux d'Elsa</i> <i>La Leçon de Ribérac</i>
OP 4 1942-1952	p. 9-217	Livre IX'	2e partie (1942)	<i>En français dans le texte</i> <i>Brocéliande</i>
	p. 219-653	Livre X	1943-1945	<i>Musée Grévin</i> <i>Diane française</i> <i>L'enseigne de Gersaint</i>
	p. 655-1221	Livre XI	1946-1952	<i>Chroniques du Bel canto</i> <i>Nouveau crève-cœur</i> <i>Mes caravanes</i> <i>La Patrie en danger</i>

Tableau 1 — Corpus numérisé

Caractéristiques globales du vocabulaire

Notre corpus poétique comporte un total de 903.672 occurrences (suites de caractères séparées par un blanc ou par une ponctuation) ou de 41.748 *formes* différentes (ou « Vocables » dans la terminologie d'Étienne Brunet, auteur du logiciel HYPERBASE dont je me sers ici). C'est l'un des plus important corpus poétique qui ait été numérisé pour un seul auteur. À titre de comparaison, les poésies de Victor Hugo enregistrées par FRANTEXT comptent un peu plus d'un million d'occurrences pour 34.380 formes.

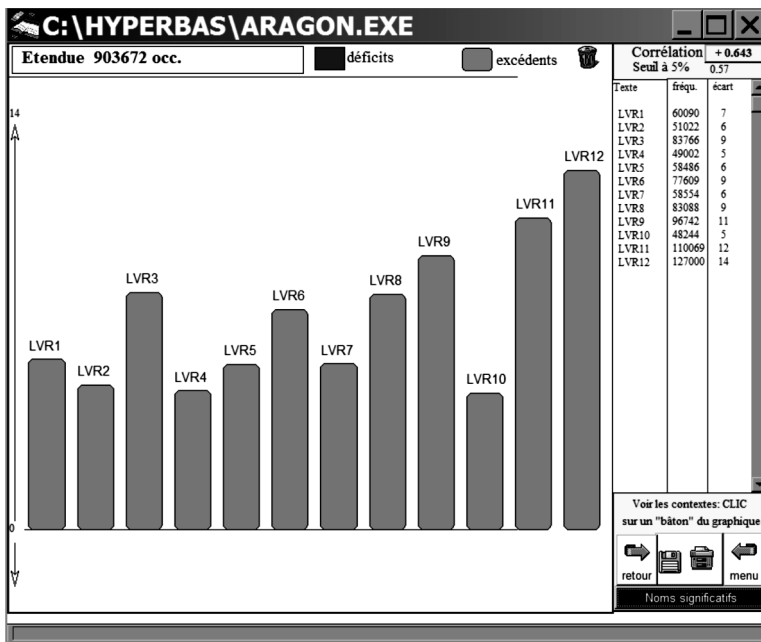


Figure 1 — Caractéristiques du corpus (occ.)

Les logiciels n'acceptant pas un libellé équivoque tel que le IX et le IXbis, j'ai dû procéder à un décalage, de sorte que le Livre IX' devient (uniquement dans les traitements) le 10, et ainsi de suite jusqu'à 12. Cela explique que, dans le graphique ci-dessus, le Livre 10 soit le plus court, avec seulement 48.244 occurrences, puisqu'il procède d'un dédoublement éditorial.

Prenons une première vue de la fréquence du vocabulaire utilisé dans chaque partie :

N°	TITRE	OCC.	VOCABLES
1	LVR1	60 090	9 599
2	LVR2	51 022	7 941
3	LVR3	83 766	1 1140
4	LVR4	49 002	8 039
5	LVR5	58 486	8 916
6	LVR6	77 609	9 530
7	LVR7	58 554	7 984
8	LVR8	83 088	10 358
9	LVR9	96 742	12 225
10	LVR10	48 244	7 128
11	LVR11	110 069	12 396
12	LVR12	127 000	13 103
TOTAL		903 672	41 748

Tableau 2 — Structure du corpus (occ./vocables)

Ces premiers calculs sont indispensables pour la suite des analyses quantitatives, mais ils n'intéressent que les statisticiens ou lexicomètres.

Richesse du vocabulaire et hapax

Un public de poètes s'intéressera davantage à ce que certains nomment la « richesse » du vocabulaire, encore que cette notion soit très controversée, puisqu'elle dépend avant tout de la longueur du corpus considéré. Voici tout d'abord les fréquences observées des vocables (ou formes) et des hapax :

n°	réel	Hapax
1	9599	1625
2	7941	1048
3	11140	1973
4	8039	1155
5	8916	1328
6	9530	1284
7	7984	984
8	10358	1637
9	12225	2138
10	7128	776
11	12396	2155
12	13103	2429
TOTAL	41748	18532

Tableau 3 — Hapax

On observera que le nombre total des hapax (18.532) est inférieur au total des hapax de toutes les parties, pour l'excellente raison qu'une forme peut fort bien n'apparaître qu'une seule fois dans une partie du corpus (ce pourquoi il est compté comme hapax), mais reparaitre ensuite ailleurs, ce qui l'élimine du calcul final. D'autre part, il faut insister sur le fait que nous prenons en considération la seule forme graphique et non le lemme d'un mot, notamment chaque forme d'un verbe, ce qui explique le nombre élevé d'hapax.

L'outil dont nous nous servons calcule les écarts réduits, c'est-à-dire la proportion des fréquences observées par rapport à la longueur relative de chaque partie. Nous représenterons ce tableau sous la forme d'un histogramme :

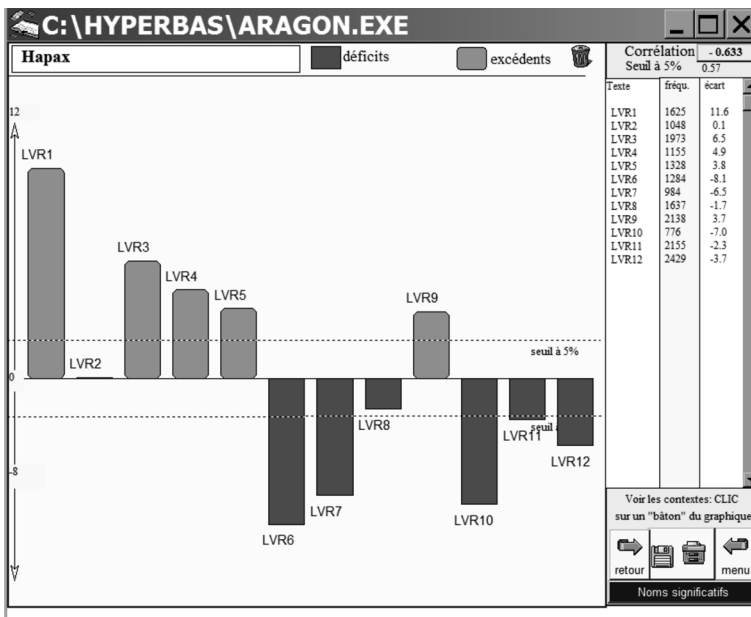


Figure 2 — Richesse/Hapax dans l'ensemble du corpus

S'il est évident à nos yeux que le premier tome de *L'Œuvre poétique* d'Aragon comporte un maximum d'hapax, il peut sembler curieux que ceux-ci soient en quantité relativement décroissante jusqu'en 1935, puis franchement déficitaire ensuite, à l'exception d'une brève période de redressement en 1939-41. Je suppose que chacun aura son explication, liée à l'idée qu'il se fait de l'évolution du poète lui-même. Il conviendra toutefois d'examiner la liste de ces hapax avant de conclure.

On peut mettre cette notion de richesse lexicale en relation avec un autre paramètre, celui de l'accroissement lexical, calcul évaluant le renouvellement progressif des formes :

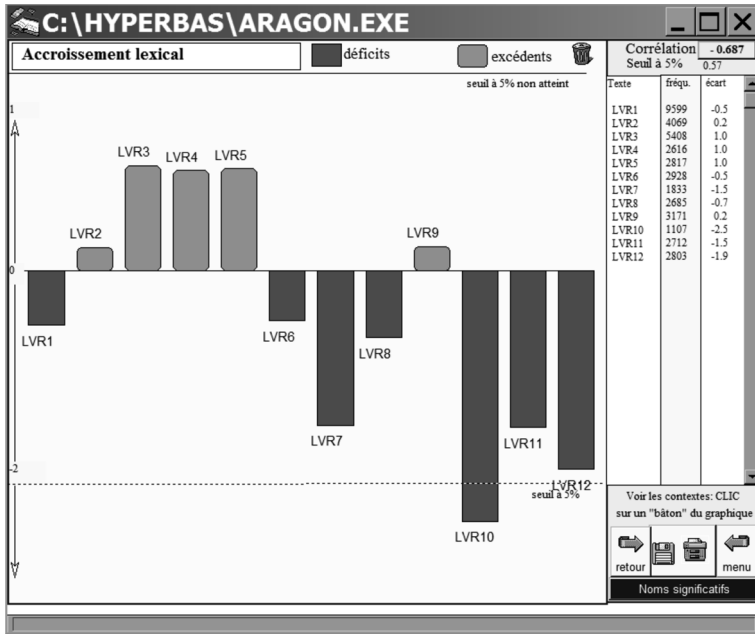


Figure 3 — Accroissement lexical

Ce renouvellement suit normalement la chronologie et se ralentit au cours du temps. On observe ici la quasi symétrie avec l'histogramme précédent.

Pour significatifs qu'ils soient, de tels graphiques ne parlent pas assez aux aragoniens que vous êtes. C'est pourquoi je vous présenterai ici la liste des 25 premiers mots en progression dans l'ensemble de l'œuvre considérée :

France, 1083 ; leçon, 127 ; lorsque, 80 ; ceux, 933 ; français, 665 ; national, 101 ; actuel, 21 ; justice, 142 ; angoisse, 22 ; faite, 108 ; notre, 1083 ; morts, 182 ; fusillé, 35 ; nationale, 86 ; guerres, 43 ; naguère, 29 ; troupeau, 22 ; précisément, 107 ; ont, 1418 ; chant, 184 ; été, 806 ; crains, 26 ; treize, 37 ; camps, 46 ; française, 382 ;

Tableau 4 — Mots en progression

Toujours calculée relativement à l'ensemble du corpus, la progression des formes est particulièrement significative. Ce tableau montre la place de plus en plus grande que prennent les préoccupations nationales face à l'envahisseur dans la poésie d'Aragon, en fonction d'une réalité présente. À tel point que cela constitue un véritable thème, caractéristique de la poésie de la résistance. Ce thème est à mettre en relation avec d'autres formes moins attendues, telles que leçon, qu'un retour au texte montre en progression croissante depuis le début, ou angoisse, qui n'apparaît qu'au livre VI pour croître à son tour.

À l'inverse, et selon le même mode de calcul, certains vocables sont en régression constante. Voici les 25 premiers d'entre eux :

aperçoit, 31 ; voyez, 112 ; besoin, 109 ; avez, 198 ; petits, 201 ; aise, 23 ; nues, 23 ; désirs, 37 ; corps, 223 ; sein, 51 ; milieu, 198 ; doute, 313 ; fruits, 36 ; à, 14002 ; vous, 3063 ; étend, 26 ; personne, 281 ; Dada, 77 ; représentation, 29 ; univers, 137 ; reconnaît, 32 ; petites, 96 ; êtes, 116 ; caresses, 31 ; se, 4005

Tableau 5 — Mots en régression

Ce sont d'abord des formes verbales du champ de la vision qui décroissent : *aperçoit*, à partir du livre VII, jusqu'à disparaître totalement des livres X et XI ; *voyez* de même (sans jamais disparaître, pourtant). Le *besoin*, très présent dans les six premiers livres, décroît ensuite proportionnellement, tout comme la locution à *l'aise*. Même coupure pour les *nues* (femmes ou nuages). Le ou les *corps*, le *sein*, les *désirs*, les *caresses* connaissent le même sort, à partir du si problématique livre V. Certes, il faudrait poursuivre l'analyse de chaque occurrence dans son contexte. On remarquera toutefois que de telles observations, portant sur de grandes quantités, ne seraient pas possibles sans l'outil informatique.

Les hapax ne sont pas des néologismes

Pour ce qui concerne la richesse lexicale, le texte parfait serait celui où chaque forme n'est employée qu'une fois, ce qui, on en conviendra, n'est pas tenable sur la longueur. Quoi qu'il en soit, dans ce corpus, les hapax représentent à eux seuls plus de 44 % des formes employées par Aragon. Ce pourcentage n'est peut-être pas significatif en lui-même. Qu'on le compare au corpus poétique de Victor Hugo, mentionné ci-dessus : 13.000 hapax, soit 37,8 % des formes. Contrairement à ce qu'on imagine, la poésie d'Aragon est incontestablement plus riche !

Il ne m'est pas possible de fournir la liste des hapax au cours de cet exposé. Dans l'ensemble, toutes les formes employées une fois seulement relèvent d'un registre fort commun (à l'exception des noms propres). Curieusement, le vocabulaire rare d'Aragon, s'il est parfois recherché (*hourvari* VIII, 2471, *icôneries* VI, 1510, *mirlitonesque* VI, 1653) ou populaire (*maquarel*, IV, 1129 ; *métallo*, X, 3212) ne comporte ni néologismes, ni formations artificielles. Tout se passe comme si le poète obéissait à la règle scolaire interdisant de répéter un même mot, sous la même forme, sur la longueur d'une page. Il use d'un verbe à tous les temps, tous les modes et toutes les personnes, et ceci une seule fois dans le corpus : *méprisa*, *méprisais*, *méprisait*, *méprisante*, *méprisée*, *méprît*. De même les substantifs sont employés tour à tour au masculin, au féminin, au singulier, au pluriel : *malsain*

(VII 2024), *malsaine* (IX 2667), *malsains* (I 88), *malsaines* (III 573). Une fois chacun, comme s'il voulait abattre un carré d'as, et ceci sur la longueur de son œuvre poétique !

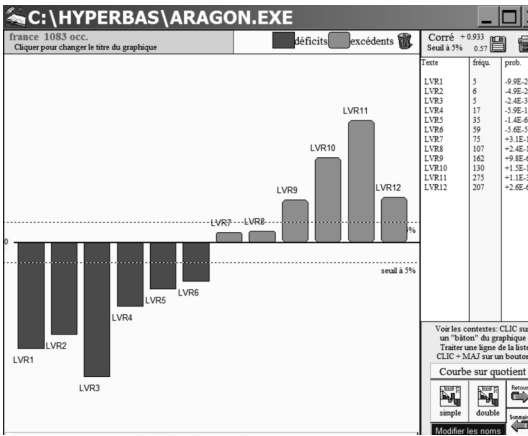
À l'inverse de cet usage discret des mots, observons les 100 vocables qu'Aragon emploie le plus fréquemment, dans l'ordre de fréquence décroissante :

rang	frq	mot	rang	frq	mot	rang	frq	mot
1	49085	,	39	3347	comme	77	1155	tous
2	48343	'	40	3214	par	78	1152	dont
3	33991	de	41	3063	vous	79	1143	mon
4	22867	.	42	2962	sur	80	1116	monde
5	22116	la	43	2857	cette	81	1103	temps
6	16868	et	44	2751	avec	82	1100	quand
7	16790	le	45	2537	où	83	1098	là
8	16177	l	46	2530	mais	84	1094	peut
9	14732	les	47	2512	tout	85	1083	notre
10	14002	à	48	2471	y	86	1083	france
11	11506	des	49	2267	son	87	1077)
12	10850	-	50	2240	aux	88	1045	leurs
13	10779	qui	51	2066	j	89	1045	(
14	10087	que	52	2036	elle	90	1035	;
15	9791	d	53	2033	ils	91	994	encore
16	9738	un	54	2027	...	92	971	moi
17	9676	est	55	2001	ces	93	951	hommes
18	8967	il	56	1956	si	94	946	ici
19	7096	dans	57	1894	ou	95	934	vie
20	6963	en	58	1862	ses	96	933	ceux
21	6830	qu	59	1823	sont	97	929	autre
22	6768	ce	60	1800	?	98	921	deux
23	6674	du	61	1759	sa	99	919	rien
24	6374	une	62	1734	même	100	890	vers
25	5424	pas	63	1728	sans			
26	5256	je	64	1676	fait			
27	5183	ne	65	1670	me			
28	5040	pour	66	1660	leur			
29	4907	a	67	1591	bien			
30	4767	n	68	1562	lui			
31	4663	au	69	1543	m			
32	4402	on	70	1457	était			
33	4005	se	71	1418	ont			
34	3975	nous	72	1418	!			
35	3829	s	73	1330	homme			
36	3480	:	74	1251	ai			
37	3428	plus	75	1228	être			
38	3412	c	76	1189	avait			

Les plus hautes fréquences : Aragon poète national

De la liste des plus hautes fréquences (classées dans l'ordre hiérarchique décroissant), je note que, ponctuation mise à part, c'est, comme dans tous nos corpus, la préposition de qui vient en tête, suivie de l'article défini au singulier. De même, *homme* occupe la première place du vocabulaire lexical (73^e rang), suivi ici de *monde* et de *temps*. Arrivant en 39^e position, la conjonction *comme* (3347 occ.), qu'André Breton tenait pour le plus beau mot de la langue française ; en 67^e position l'adverbe *bien* (1591 occ.),

majoritairement employé comme intensif, ce qui est bien un trait personnel ; et en 86^e position le substantif *France* (1083 occ.), passant de la détestation (les 3 occ. d'Anatole France ne peuvent qu'y contribuer) à l'appropriation, comme on peut le vérifier par le contexte, et selon un emploi massif que résume le graphique ci-joint.



De la poésie d'un seul à la poésie de tous : je/nous

S'agissant des hautes fréquences des formes-outils que l'amateur de poésie ne remarque guère dans un corpus aussi considérable, il est une autre évolution remarquable, celle du couple *Je/Nous*. Aux 5256 occ. de *Je* on pourrait ajouter les 2066 apparitions de *J'* que cela ne changerait pas le propos. En vérité, le locuteur (qu'il soit le poète lui-même ou un personnage, dans le cas des récits tels que *Les Aventures de Télémaque*) s'exprime massivement à la première personne jusqu'au livre IV, et reprend la parole pendant la guerre, dans des recueils tels que *Le Crève-cœur* et *Les Yeux d'Elsa*, pour s'effacer ensuite devant la première personne du pluriel. Celle-ci représente bien un collectif, tour à tour, selon le contexte d'énonciation, Aragon et Breton, les Dadas, comme ils disaient alors, ou l'ensemble du groupe surréaliste, et ensuite les camarades du Parti¹. C'est ce

1. Certes, il faudrait distinguer le nous collectif du nous auctorial, comme l'a joliment fait Michel BERNARD pour Hugo dans son article « Les *je/nous* de Victor Hugo. Pluriel de modestie et pluriel de majesté dans *Les Misérables* », *Revue de littératures française et comparée*, Publications de l'Université de Pau, n° 3, novembre 1994, pp. 129-136. Un bref examen des contextes montre que cet emploi, toujours ambigu, demeure rare chez Aragon.

qu'illustre l'histogramme suivant, où les emplois de ces deux formes ont été superposés (je en rouge, nous en bleu). Alors que la quantité des je dépasse largement celle du nous, le calcul hypergéométrique montre bien le déficit progressif de l'un au profit de l'autre :

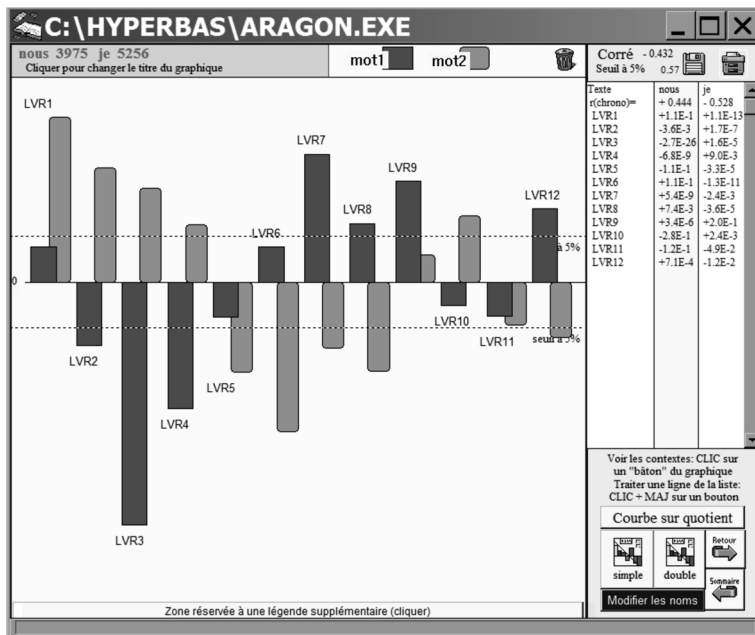


Figure 4 — Histogramme superposé Je/Nous

Pour le dire en des termes empruntés à son ami Paul Éluard, avec qui il s'était retrouvé durant la Deuxième Guerre mondiale, Aragon est bien passé de l'horizon d'un seul à l'horizon de tous. En témoigne clairement son livre V, rassemblant des pièces de 1930 à 1933.

Spécificités internes par parties

Il est temps de s'intéresser à chaque section de l'œuvre poétique telle qu'elle a été constituée par le poète lui-même, et de voir ce qui distingue l'une de l'autre. Pour cela, nous disposons d'un programme calculant les « spécificités ». Il isole les formes particulièrement fréquentes par rapport à leur probabilité d'apparition dans la section du corpus examiné. Soit le mot *cinéma* qui se trouve 31 fois dans le Livre I, alors qu'il est employé 83 fois dans l'ensemble du corpus, le calcul des probabilités montre que cette distribution n'avait qu'une chance sur 10^{16} d'intervenir par hasard. Elle est

donc fortement caractéristique (au passage, observons que ce vocable apparaît ainsi dans ce premier livre puis dans le VI, signe des temps pour Aragon).

Pour des raisons pratiques, une fois de plus, je m'en tiendrai à la seule période surréaliste d'Aragon, entendue au sens large puisque dans notre partition elle part de ses premiers écrits et va jusqu'à son éclatante sortie du groupe.

Aragon poète surréaliste (1917-1932)

Pour de bonnes raisons théoriques et pratiques, je me limiterai à la période surréaliste, qui court des livres I à V, jusqu'à la rupture officielle avec le groupe, en 1932. Et faute de temps pour analyser en détail le tableau des spécificités de chaque livre, je les résumerai ainsi : le livre I est celui de Tzara (8/26) ; le livre II de Breton (28/62, E-19) ; les livres I et II mentionnant Dada à part égale (19/77 E-7 ; 19/77 E-8). Le livre III, dominé par *Le Paysan de Paris*, est celui de la *rue*, de la *métaphysique*, du *concret*, de *l'idée*, du *sentiment*, de *l'imagination* et du *passage*.

Plus composite, le livre IV alterne philosophie (*Héraclite*, *Hegel*) et grossièreté du temps de *La Grande Gaité* (*chie, con, pine, foutre...*). Situé à la période charnière où Aragon est supposé exclu du groupe surréaliste, alors qu'il ne l'a quitté que très progressivement, ce volume demande un examen précis. Rappelons qu'il contient notamment *La Peinture au défi*, *Front rouge*, *Persécution persécuteur*, et les textes parus dans *La Littérature internationale*, dont André Breton prit la défense devant la justice. Les termes spécifiques sont relatifs au *collage* chez Max Ernst, au *Guépéou*, à la *dialectique*, au *surréalisme* et au *matérialisme*, au *prolétariat*, à la *révolution* et aux *révolutionnaires*. C'est dire combien il est complexe et, globalement, voué aux deux pôles de l'imaginaire aragonien.

Spécificités externes de la période surréaliste

La question qui se pose maintenant est de savoir dans quelle mesure ce poète surréaliste use d'un vocabulaire original par rapport à ses confrères, dans la même période 1920-1932.

Pour en juger, j'ai comparé¹, dans la banque de données FRANTEXT, le corpus Aragon que j'y ai déposé à celui des 48 textes poétiques enregistrés pour la même période, dont voici la liste par ordre alphabétique des auteurs :

1. Les requêtes et calculs subséquents ont été effectués par Michel BERNARD, que je remercie de sa disponibilité.

38 *Le lexique dans l'Œuvre poétique d'Aragon...*

n°	Cote	Auteur	Titre	Date	Genre	Edition
1	S671	*Collectif	<i>La Révolution surréaliste.</i> N° 1, première année. 1 ^{er} décembre 1924	1924	critique, essai poétique, poésie	Paris : Jean-Michel Place, 1991.
11	S681	*Collectif	<i>La Révolution surréaliste.</i> N° 12, cinquième année. 15 décembre 1929	1929	critique, essai poétique, poésie	Paris : Jean-Michel Place, 1991.
12	S078	ARAGON Louis	<i>Feu de joie</i>	1920	poésie	Paris : Livre Club Diderot, 1989.
13	S080	ARAGON Louis	<i>Le Mouvement perpétuel</i>	1926	poésie	Paris : Livre Club Diderot, 1989.
14	S082	ARAGON Louis	<i>Les Destinées de la poésie</i>	1926	poésie	Paris : Livre Club Diderot, 1989.
15	K576	CLAUDEL Paul	<i>Feuilles de saints</i>	1925	poésie	Paris : Gallimard, 1957.
16	S635	DESNOS Robert	<i>Corps et biens</i>	1930	poésie	Paris : Gallimard, 1998.
17	S421	ÉLUARD Paul	<i>Les Animaux et leurs hommes, les hommes et leurs animaux</i>	1920	poésie	Paris : Gallimard, 1990.
18	S422	ÉLUARD Paul	<i>Pour vivre ici : Onze haïkai</i>	1920	poésie	Paris : Gallimard, 1990.
19	S423	ÉLUARD Paul	<i>Les Nécessités de la vie et les conséquences des rêves précédé d'Exemples</i>	1921	poésie	Paris : Gallimard, 1990.
20	S424	ÉLUARD Paul	<i>Répétitions</i>	1922	poésie	Paris : Gallimard, 1990.
21	S425	ÉLUARD Paul	<i>Les Malheurs des immortels</i>	1922	poésie	Paris : Gallimard, 1990.
22	S426	ÉLUARD Paul	<i>Mourir de ne pas mourir</i>	1924	poésie	Paris : Gallimard, 1990.
23	S427	ÉLUARD Paul	<i>152 proverbes</i>	1925	poésie	Paris : Gallimard, 1990.
24	S428	ÉLUARD Paul	<i>Au défaut du silence</i>	1925	poésie	Paris : Gallimard, 1990.
25	L394	ÉLUARD Paul	<i>Capitale de la douleur</i>	1926	poésie	Paris : Gallimard, 1990.
26	S430	ÉLUARD Paul	<i>Les Dessous d'une vie ou la Pyramide humaine</i>	1926	poésie	Paris : Gallimard, 1990.
27	S431	ÉLUARD Paul	<i>Défense de savoir</i>	1928	poésie	Paris : Gallimard, 1990.
28	S432	ÉLUARD Paul	<i>L'Amour la poésie</i>	1929	poésie	Paris : Gallimard, 1990.

n°	Cote	Auteur	Titre	Date	Genre	Edition
29	S434	ÉLUARD Paul	<i>À toute épreuve</i>	1930	poésie	Paris : Gallimard, 1990.
30	S436	ÉLUARD Paul	<i>Dors</i>	1931	poésie	Paris : Gallimard, 1990.
31	S437	ÉLUARD Paul	<i>La Vie immédiate</i>	1932	poésie	Paris : Gallimard, 1990.
32	S433	ÉLUARD Paul BRETON André CHAR René	<i>Ralentir travaux</i>	1930	poésie	Paris : Gallimard, 1990.
33	L364	JACOB Max	<i>Le Cornet à dés</i>	1923	poésie	Paris : Gallimard, 1945.
34	L268	JOUVE Pierre-Jean	<i>Tragiques</i>	1922	poésie	Paris : Stock, 1922.
35	L597	PONCHON Raoul	<i>La Muse au cabaret</i>	1920	poésie	Paris : Fasquelle, 1920.
36	R914	REVERDY Pierre	<i>Flaques de verre</i>	1929	poésie	Paris : Flammarion, 1972.
37	R661	SAINT-JOHN PERSE	<i>Anabase</i>	1924	poésie	Paris : Gallimard, 1986.
38	R995	SUPERVIELLE Jules	<i>Débarcadères</i>	1922	poésie	Paris : Gallimard, 1994.
39	R996	SUPERVIELLE Jules	<i>Gravitations</i>	1925	poésie	Paris : Gallimard, 1994.
40	R997	SUPERVIELLE Jules	<i>Le Forçat innocent</i>	1930	poésie	Paris : Gallimard, 1994.
41	L737	TOULET Paul-Jean	<i>Les Contrerimes</i>	1920	poésie	Paris : Émile-Paul, 1949.
42	R585	TZARA Tristan	<i>Poèmes (1912-1924)</i>	1924	poésie	Paris : Flammarion, 1975.
43	R593	TZARA Tristan	<i>Indicateur des chemins de cœur</i>	1928	poésie	Paris : Flammarion, 1977.
44	R594	TZARA Tristan	<i>Poèmes (1925-1929)</i>	1929	poésie	Paris : Flammarion, 1977.
45	R589	TZARA Tristan	<i>L'Arbre des voyageurs</i>	1930	poésie	Paris : Flammarion, 1977.
46	R590	TZARA Tristan	<i>L'Homme approximatif</i>	1931	poésie	Paris : Flammarion, 1977.
47	R591	TZARA Tristan	<i>Où boivent les loups</i>	1932	poésie	Paris : Flammarion, 1977.
48	K492	VALÉRY Paul	<i>Poésies : Album de vers anciens/La Jeune Parque/ Charmes</i>	1922	poésie	Paris : Gallimard, 1959.

On ne sera pas surpris d'y voir un grand nombre d'œuvres surréalistes, à commencer par la revue *La Révolution surréaliste*, puisque j'en ai moi-même fourni les fichiers numérisés à cet organisme national. Si l'on estime qu'il y a là une surreprésentation des surréalistes durant la période considérée, je répondrai que cela rend notre comparaison d'autant plus pertinente. À noter que les œuvres poétiques d'Aragon y ont été éclatées par recueil (et non plus par livres de l'édition collective).

Et voici le tableau des 50 premières spécificités positives qui résulte de cette comparaison :

Forme	Total	Partiel	Mantisse	Exposant
Eucharis	447	447	2,56	297
Télémaque	88	87	1,38	56
:	3562	1161	4,05	51
,	48322	11770	2,22	46
Mentor	67	67	3,68	45
Calypso	46	46	3,11	31
Persienne	44	43	1,05	27
Vous	3468	1017	5,57	26
André	211	115	3,32	25
Dada	98	66	5,09	22
Apollinaire	59	46	5,66	20
Ce	5890	1567	3,15	19
Ça	223	110	1,05	19
.	27683	6578	6,63	17
Rue	354	144	7,97	16
Philippe	69	47	2,08	16
Ici	565	204	4,29	15
Poésie	198	93	3,29	15
sentiment	133	68	9,63	14
On	3649	981	8,48	14
Métaphysique	57	39	5,9	14
Cinéma	59	40	4,53	14
Gens	296	119	6,02	13
Avait	732	243	5,43	13
Passage	162	77	3,38	13
Certa	19	19	2,51	13
Café	111	60	1,08	13
À	14817	3557	9,42	12
Breton	182	81	6,38	12

Forme	Total	Partiel	Mantisse	Exposant
Mariette	17	17	5,32	12
Opéra	46	32	5,29	12
Buttes	20	19	3,98	12
Magnétiques	25	22	2,94	12
Galerie	32	26	1,27	12
Concret	41	29	2,75	11
Moderne	124	61	1,46	11
Coiffeur	23	19	8,83	10
Semblait	60	35	8,65	10
Parler	209	85	5,45	10
Quartier	42	28	5,33	10
Paris	225	90	4,82	10
Reverdy	32	24	1,97	10
Ulysse	34	25	1,68	10
Alors	625	205	1,05	10
Brouet	17	15	9,74	9
Logique	78	40	9,61	9
Petits	213	83	9,3	9
Texte	56	32	9,15	9
Connaissance	110	51	8,59	9
Idée	397	136	6,51	9

Figure 5 — Spécificités positives de la poésie surréaliste d'Aragon par rapport au corpus poétique contemporain

On ne s'étonnera pas davantage de constater que les spécificités constatées pour la période surréaliste d'Aragon à l'intérieur de l'*Œuvre poétique* sont sensiblement les mêmes quand on rapporte ce corpus à celui enregistré par Frantext pour la même période : outre les noms propres de ses maîtres et amis (Apollinaire, Reverdy, André Breton, Philippe Soupault), c'est Dada qui ressort pour la période, comme en témoigne l'usage perturbant d'*Eucharis* ou de *persienne*. Les noms de personnages (de récits ou de théâtre) y sont exclusifs (*Télémaque*, *Mariette*). Walter Benjamin l'avait déjà observé : *Paris*, la *rue*, le *passage*, la *galerie*, le *quartier* sont, avec le *cinéma*, les mots propres d'Aragon qui s'intéresse à la *poésie*, à la *métaphysique* et à la *connaissance*. Cela ne veut pas dire que les autres n'y attachent aucune importance. Il est celui qui fait entrer ce vocabulaire dans sa poésie. À cet égard, le tableau des *spécificités négatives* est tout aussi significatif, puisqu'il signale les vocables en sous-emploi chez lui par rapport à ses contemporains dans le même temps. Pour éviter l'effet vitrine, je relèverai les éléments les

plus remarquables : par rapport à ses contemporains, il n'emploie jamais la forme *encor*, les substantifs *âme, nuit, vin, ciel, terre, rose, feuille, amour*, et même *révolution, sang, mort* ; les pronoms *tu, ton, te, toi, ta, nous...* sont remarquablement sous employés. Qu'en dépit de son admiration pour Maurice Barrès, Aragon ne soit pas le poète de la volupté, du sang et de la mort, cela ne peut nous surprendre, mais qu'il emploie moins que ses contemporains les termes *révolution, révolutionnaire*, cela demande considération ! Serait-ce que ces mots échappent au langage poétique ?

La seconde partie de l'Œuvre poétique

Au-delà de cette période dite surréaliste, il faudrait poursuivre l'examen des spécificités de chaque livre, comme nous l'avons fait précédemment. Nous en avons les moyens, mais le temps manque. S'agissant d'un corpus présenté dans l'ordre chronologique de publication, on y verrait l'évolution particulière du poète dans l'orbite communiste, antifasciste militant pour le Front populaire, puis pour une poésie nationale par un retour à la versification traditionnelle, célébrant, comme on l'a dit, la nation française.

Mieux que tout discours, un tableau fondé sur la différenciation des fréquences nous montrera comment se répartissent les livres. Je terminerai par une analyse des segments de phrases répétés.

L'analyse factorielle des correspondances (AFC)

L'analyse factorielle des correspondances (AFC) est une méthode statistique qui rapproche les différentes parties selon leur contenu, ou plus exactement la proximité et la hiérarchisation des mots entre eux, en formant des constellations, des nuages plus ou moins proches.

Alors qu'il n'est bâti que sur la proximité des formes, le tableau ci-dessous résume l'évolution chronologique du vocabulaire poétique aragonien, le Livre V (1930-1933) formant charnière entre le groupe des œuvres surréalistes et les dites « réalistes ».

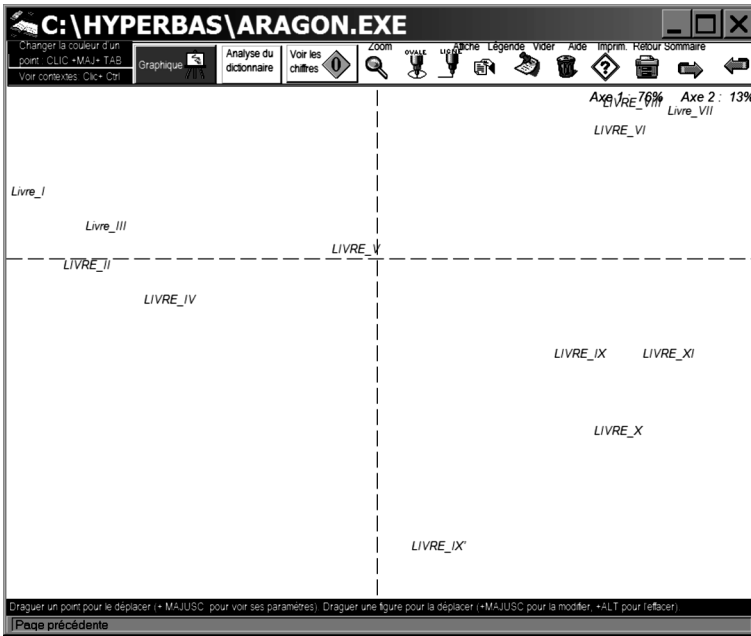


Figure 6 — AFC du corpus

Les segments répétés

Jusqu'à présent, nous avons considéré chaque forme ou vocable en soi. Or la langue poétique, comme toute autre, est une combinatoire de ces formes. Un programme¹ nous permet d'identifier les séquences de mots répétés, à quelque distance qu'elles se trouvent dans le corpus. Ainsi, nous observons 6139 segments de 2 à 11 segments répétés de 1 à 443 fois. Écartons la série des *Eucharis* (livre I, 200) et celle de *Persienne* qui peuvent être pris comme mot ou phrase et faussent la perspective. Dans l'ordre décroissant, c'est la dizaine de *Celui qui croyait au ciel Celui qui n'y croyait pas* de « La rose et le réséda » (livre X) et *Vive le son vive le son du canon* de « La carmagnole des enfants » (livre X) ; 9 segments reviennent 10 fois : *des canons par centaines et des fusils par milliers* ou encore 7 segments : *il n'y a pas d'amour...* En somme, c'est ici la chanson qui prévaut, si l'on écarte les dénominations figées du genre *Congrès des écrivains pour la défense de la culture* ; 5 segments se répètent 17 fois : *la maison de la culture*. Une séquence est plus remarquable : la chaîne de 6 segments revient 20 fois : *à l'heure qu'il est*, marquant une préoccupation de l'instant ; 6 segments 19 fois : *il n'y a pas d'* ; et *il n'y a pas de* 21 fois ; *je n'aime pas les gens qui* 12 fois. On m'accuserait

1. Élaboré par André Salem, professeur à l'université Paris III, Lexico3 est accessible à l'adresse : <http://www.cavi.univ-paris3.fr/ilpga/ilpga/tal/lexicowww/>

de censure si je ne signalais pas la reprise (15 fois) de 5 segments : *la pine et le con* (« Calendrier 1929 ») ; enfin, de même longueur, la dizaine de *dans le domaine de la* : nous situe dans le discursif.

En résumé, ces séquences répétées caractérisent la chanson ou, à l'opposé, le discours théorique.

Conclusion

On conviendra avec moi qu'il a rarement été possible d'examiner « scientifiquement » autant de recueils d'un seul regard. C'est le regroupement effectué par Aragon lui-même que nous avons pris en considération, à la fois en soi et par rapport aux œuvres contemporaines. Une fois de plus ce diable d'homme place ses lecteurs et la postérité en porte à faux : associant des œuvres lyriques, du théâtre, des essais, des compositions de ses amis à sa propre publication, il mêle les genres et pervertit nos habitudes, tant de lecture que d'analyse. Cette singularité ne nous empêche pas d'observer que chez lui, la richesse du vocabulaire, l'usage des hapax, qui est grand, n'est pas source de néologie. Tous les indicateurs numériques convergent pour souligner la bipartition de son œuvre, avec un livre charnière en 1930-32, opposant la période dada-surréaliste d'une part à la période réaliste d'autre part, ce que confirme la comparaison avec les œuvres poétiques de ses contemporains. Statistiquement, le vocabulaire de la deuxième période prend le dessus. Le poète du *Je* devient celui du *Nous*. De sorte que l'opinion publique n'a pas tort qui fait d'Aragon le chantre national. Pour reprendre un mot qu'il affectionnait, c'est la *leçon* que je tirerai de cette longue relecture assistée par mon personnel golem, qui, si elle ne bouleverse pas nos connaissances, a du moins le mérite de nous obliger à une attention mot à mot.

Henri BÉHAR

Université Paris III-Sorbonne Nouvelle

Spécialiste des avant-gardes, éditeur des œuvres complètes de Roger Vitrac et de Tristan Tzara, auteur d'une biographie d'André Breton et de nombreux essais, Henri Béhar, Professeur émérite de Littérature française à la Sorbonne Nouvelle, édite la revue *Mélusine*.

Dès 1963, il a rencontré Aragon, qui s'est empressé de le mettre dans un de ses livres. C'est peut-être pour ne pas subir sa légendaire séduction qu'il s'attache à observer sa poésie à l'aide d'outils incontestablement objectifs.