

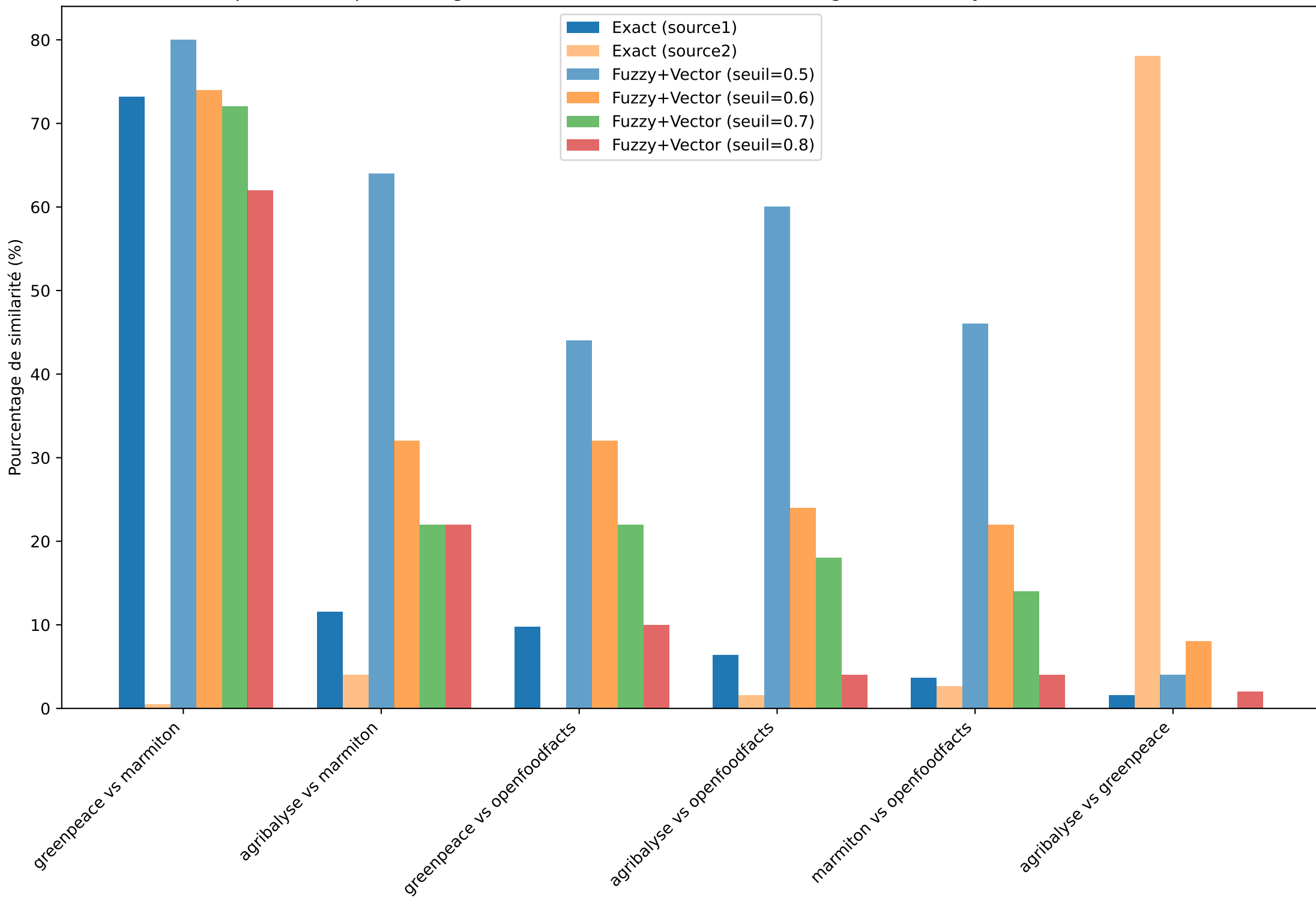
Correspondance exacte entre sources

source1	source2	common_count	total_source1	total_source2	percent_source1	percent_source2
greenpeace	marmiton	30	41	5940	73.17	0.51
agribalyse	marmiton	239	2070	5940	11.55	4.02
greenpeace	openfoodfacts	4	41	8315	9.76	0.05
agribalyse	openfoodfacts	132	2070	8315	6.38	1.59
marmiton	openfoodfacts	218	5940	8315	3.67	2.62
agribalyse	greenpeace	32	2070	41	1.55	78.05

Similarité fuzzy + vector (échantillon, 2 sens, seuil=0.65) entre sources

source1	source2	similar_count_source1	sampled_total_source1	percent_source1	similar_count_source2	sampled_total_source2	percent_source2
agribalyse	greenpeace	18	1000	1.8	36	41	87.8
agribalyse	marmiton	314	1000	31.4	185	1000	18.5
agribalyse	openfoodfacts	213	1000	21.3	71	1000	7.1
greenpeace	marmiton	36	41	87.8	16	1000	1.6
greenpeace	openfoodfacts	12	41	29.27	4	1000	0.4
marmiton	openfoodfacts	163	1000	16.3	84	1000	8.4

Comparaison du pourcentage de similarité entre sources (matching exact et fuzzy+vector, multi-seuils)



Echantillon de paires matchées (global\_score > 0.5)

source1	name1	source2	name2	code_source2	global_score	vector_similarity	fuzzy_similarity	seuil
greenpeace	crosne	agribalyse	crosne	20231	1.0	1.0	1.0	0.5
marmiton	soupe mayonnaise pot	openfoodfacts	mayonnaise	00005466	0.624	0.773	0.524	0.5
marmiton	tomates cerise rouges	agribalyse	tomate cerise	20172	0.651	0.742	0.591	0.5
agribalyse	huile darachide	marmiton	huile darachide		1.0	1.0	1.0	0.5
greenpeace	rutabaga	marmiton	rutabaga		1.0	1.0	1.0	0.5
greenpeace	epinard	marmiton	epinards		0.77	0.876	0.7	0.5
marmiton	pommes terre chair tendre	openfoodfacts	puree pommes terre	0001002974992	0.513	0.587	0.464	0.5
greenpeace	pomme terre conservation pomme terre	agribalyse	conservation sans peau bouillie	4028	0.646	0.742	0.581	0.5
openfoodfacts	bûches viennois chocolat lait	agribalyse	biscuit beurre chocolat	24017	0.551	0.706	0.447	0.5
agribalyse	crèmes vanillee appertisee	marmiton	creme marrons		0.54	0.714	0.424	0.5
openfoodfacts	crepes surgelees	marmiton	feves surgelees		0.558	0.68	0.476	0.5
agribalyse	boeuf collier braise	openfoodfacts	roti boeuf braise	00279567	0.587	0.747	0.48	0.5
openfoodfacts	soupe courgette noisette	marmiton	soupe noisettes		0.682	0.83	0.583	0.5
agribalyse	pates farcies fromage cuites	marmiton	pates fraiches farcies		0.69	0.758	0.645	0.5
marmiton	olives noires grecque	openfoodfacts	olives noires facon grece	0021885002573	0.575	0.734	0.469	0.5

Echantillon de paires matchées (global\_score > 0.6)

source1	name1	source2	name2	code_source2	global_score	vector_similarity	fuzzy_similarity	seuil
marmiton	roquette	agribalyse	roquette	20217	1.0	1.0	1.0	0.6
greenpeace	haricot	openfoodfacts	haricot	0012656200017	1.0	1.0	1.0	0.6
marmiton	noix veau	agribalyse	veau noix	6522	0.993	0.982	1.0	0.6
greenpeace	cardon	agribalyse	cardon	20054	1.0	1.0	1.0	0.6
greenpeace	ail	marmiton	ail		1.0	1.0	1.0	0.6
agribalyse	chocolat sans sucre ajouté	openfoodfacts	chocolat sans sucre ajouté	0011835212001	0.627	0.802	0.51	0.6
openfoodfacts	confiture fraises	agribalyse	confiture fraise	31024	0.871	0.915	0.842	0.6
greenpeace	patate douce	agribalyse	patate douce	4101	1.0	1.0	1.0	0.6
greenpeace	asperge	marmiton	asperges		0.727	0.768	0.7	0.6
marmiton	poivre huile dolive	openfoodfacts	huile dolive	00013024	0.738	0.87	0.65	0.6
agribalyse	pate foie porc	marmiton	pate foie porc		1.0	1.0	1.0	0.6
greenpeace	laitue	marmiton	laitue		1.0	1.0	1.0	0.6
agribalyse	fromage chevre buche	marmiton	fromage chevre buche		1.0	1.0	1.0	0.6
greenpeace	aubergine	agribalyse	aubergine	20053	1.0	1.0	1.0	0.6
agribalyse	tripes mode caen	openfoodfacts	tripes mode caen bigard	0017000004235	0.745	0.801	0.708	0.6

Echantillon de paires matchées (global\_score > 0.7)

source1	name1	source2	name2	code_source2	global_score	vector_similarity	fuzzy_similarity	seuil
greenpeace	chou	marmiton	chou		1.0	1.0	1.0	0.7
greenpeace	brocoli	marmiton	brocoli		1.0	1.0	1.0	0.7
agribalyse	huile sardine	marmiton	sardine lhuile		0.753	0.824	0.706	0.7
greenpeace	betterave	agribalyse	betterave	20091	1.0	1.0	1.0	0.7
greenpeace	endive	agribalyse	endive	20026	1.0	1.0	1.0	0.7
openfoodfacts	frites chips corn tortilla chips	marmiton	chips tortilla		0.701	0.816	0.625	0.7
greenpeace	laitue	marmiton	laitue		1.0	1.0	1.0	0.7
agribalyse	myrtille	marmiton	myrtilles		0.732	0.738	0.727	0.7
greenpeace	chou romanesco	marmiton	chou romanesco		1.0	1.0	1.0	0.7
greenpeace	poivron	marmiton	poivron		1.0	1.0	1.0	0.7
agribalyse	beurre mg	openfoodfacts	beurre	00005484	0.723	0.757	0.7	0.7
marmiton	creme mg	openfoodfacts	creme	0000209534006	0.716	0.79	0.667	0.7
agribalyse	celeri remoulade preemballe	openfoodfacts	celeri remoulade	00013406	0.702	0.845	0.607	0.7
greenpeace	concombre	openfoodfacts	thon concombre	00169066	0.704	0.759	0.667	0.7
greenpeace	cardon	agribalyse	cardon	20054	1.0	1.0	1.0	0.7

## Echantillon de paires matchées (global\_score &gt; 0.8)

source1	name1	source2	name2	code_source2	global_score	vector_similarity	fuzzy_similarity	seuil
greenpeace	brocoli	marmiton	brocoli		1.0	1.0	1.0	0.8
agribalyse	celeri	greenpeace	celeri		1.0	1.0	1.0	0.8
greenpeace	carotte	marmiton	carotte		1.0	1.0	1.0	0.8
greenpeace	mache	agribalyse	mache	20099	1.0	1.0	1.0	0.8
agribalyse	comte	openfoodfacts	comte	0000325039083	1.0	1.0	1.0	0.8
greenpeace	blette	marmiton	blette		1.0	1.0	1.0	0.8
agribalyse	coriandre graine	marmiton	graines coriandre		0.892	0.968	0.842	0.8
greenpeace	courge	agribalyse	courge	20139	1.0	1.0	1.0	0.8
agribalyse	pate feuilletée	marmiton	pate feuilletée		1.0	1.0	1.0	0.8
greenpeace	citrouille	marmiton	citrouille		1.0	1.0	1.0	0.8
greenpeace	haricot	marmiton	haricot		1.0	1.0	1.0	0.8
greenpeace	potiron	marmiton	potiron		1.0	1.0	1.0	0.8
greenpeace	oignon	marmiton	oignon		1.0	1.0	1.0	0.8
greenpeace	aubergine	marmiton	aubergines		0.808	0.894	0.75	0.8
greenpeace	poivron	marmiton	poivron		1.0	1.0	1.0	0.8

## Synthèse matching exact vs fuzzy+vector (toutes sources et hors greenpeace)

Type de matching	Toutes sources	Hors greenpeace
Matching exact (noms identiques)	2	86
Fuzzy+vector (liens)	39	1041

L'utilisation du fuzzy+vector permet d'augmenter le nombre de correspondances multi-sources de 1850.0% par rapport au matching exact.