

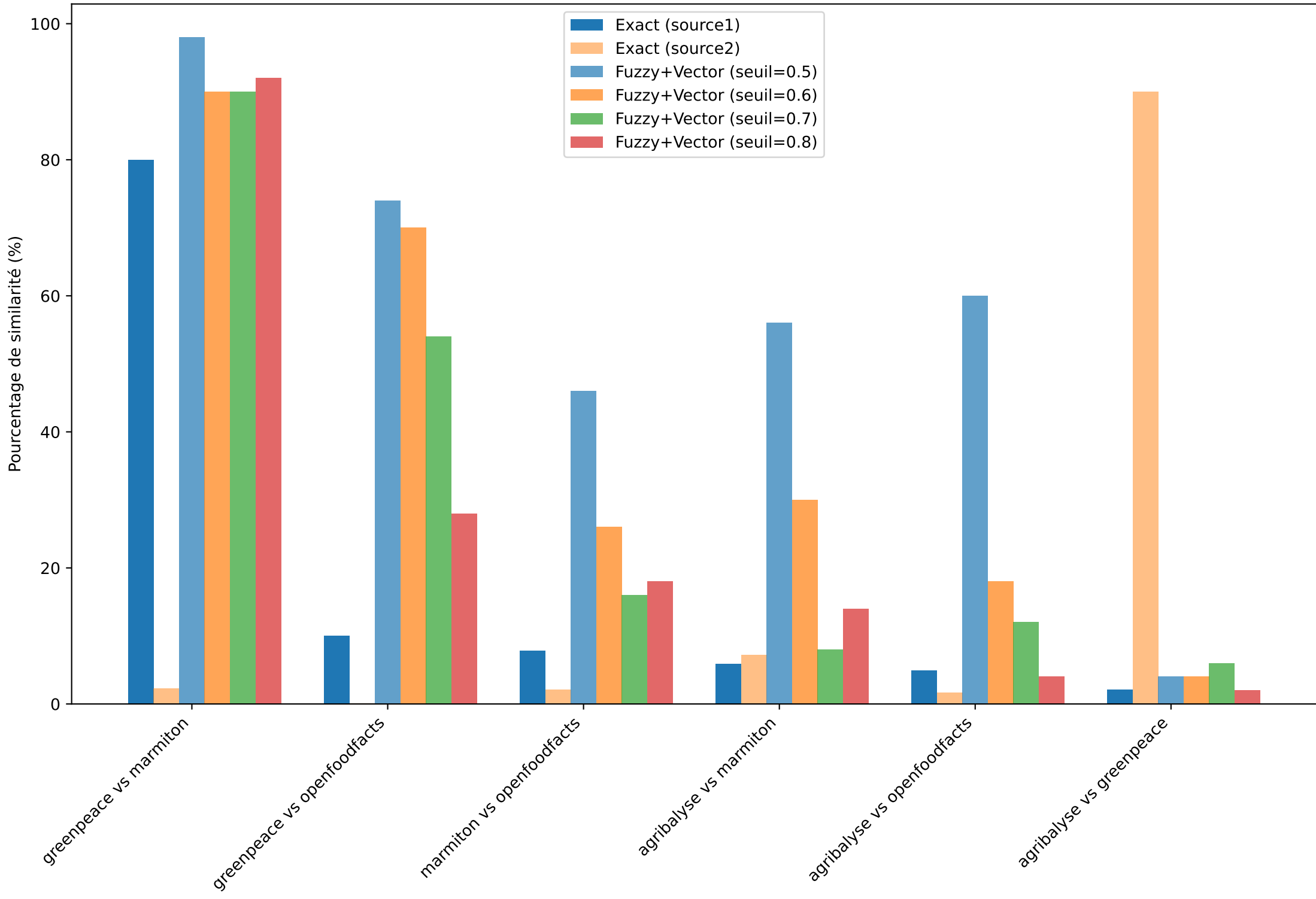
Correspondance exacte entre sources

source1	source2	common_count	total_source1	total_source2	percent_source1	percent_source2
greenpeace	marmiton	8	10	347	80.0	2.31
greenpeace	openfoodfacts	1	10	1294	10.0	0.08
marmiton	openfoodfacts	27	347	1294	7.78	2.09
agribalyse	marmiton	25	425	347	5.88	7.2
agribalyse	openfoodfacts	21	425	1294	4.94	1.62
agribalyse	greenpeace	9	425	10	2.12	90.0

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	25	486	5.14	73	81	90.12
agribalyse	marmiton	82	486	16.87	72	347	20.75
agribalyse	openfoodfacts	91	486	18.72	129	1522	8.48
greenpeace	marmiton	75	81	92.59	10	347	2.88
greenpeace	openfoodfacts	49	81	60.49	6	1522	0.39
marmiton	openfoodfacts	76	347	21.9	150	1522	9.86

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	global_score	vector_similarity	fuzzy_similarity	seuil
agribalyse	boeuf bourguignon	openfoodfacts	bourguignon	0.796	0.865	0.75	0.5
greenpeace	betterave	marmiton	betterave	1.0	1.0	1.0	0.5
greenpeace	ail	agribalyse	ail	1.0	1.0	1.0	0.5
greenpeace	ail	marmiton	ail	1.0	1.0	1.0	0.5
marmiton	cacao amer poudre	agribalyse	cacao non sucre poudre soluble	0.507	0.695	0.382	0.5
greenpeace	betterave	marmiton	betterave	1.0	1.0	1.0	0.5
marmiton	cerneau noix	agribalyse	noix sechee cerneaux	0.678	0.876	0.545	0.5
agribalyse	biscotte	openfoodfacts	biscotti	0.751	0.922	0.636	0.5
greenpeace	carotte	openfoodfacts	carottes	0.789	0.923	0.7	0.5
marmiton	aneth cisele	openfoodfacts	aneth	0.52	0.608	0.462	0.5
greenpeace	carotte	openfoodfacts	carottes	0.789	0.923	0.7	0.5
openfoodfacts	fourres creme marrons	marmiton	creme marrons	0.621	0.774	0.519	0.5
marmiton	carotte rapees	greenpeace	carotte	0.623	0.757	0.533	0.5
greenpeace	cardon	agribalyse	cardon	1.0	1.0	1.0	0.5
greenpeace	celeri	openfoodfacts	celery	0.7	0.918	0.556	0.5

Echantillon de paires matchées (global_score > 0.6)

source1	name1	source2	name2	global_score	vector_similarity	fuzzy_similarity	seuil
greenpeace	ail	agribalyse	ail	1.0	1.0	1.0	0.6
greenpeace	carotte	openfoodfacts	carottes	0.789	0.923	0.7	0.6
greenpeace	aubergine	marmiton	aubergine	1.0	1.0	1.0	0.6
greenpeace	celeri	openfoodfacts	celery	0.7	0.918	0.556	0.6
greenpeace	ail	marmiton	ail	1.0	1.0	1.0	0.6
marmiton	ail	agribalyse	ail	1.0	1.0	1.0	0.6
agribalyse	chevreau	marmiton	chevre	0.697	0.843	0.6	0.6
agribalyse	carotte	openfoodfacts	carottes	0.789	0.923	0.7	0.6
agribalyse	chair saucisse	marmiton	saucisse	0.692	0.831	0.6	0.6
agribalyse	bambou pousse	marmiton	pousse bambou	0.994	0.985	1.0	0.6
greenpeace	aubergine	marmiton	aubergine	1.0	1.0	1.0	0.6
greenpeace	ail	marmiton	ail	1.0	1.0	1.0	0.6
greenpeace	asperge	agribalyse	asperge	1.0	1.0	1.0	0.6
greenpeace	carotte	agribalyse	carotte	1.0	1.0	1.0	0.6
marmiton	brin cerfeuil	agribalyse	cerfeuil	0.729	0.858	0.643	0.6

Echantillon de paires matchées (global_score > 0.7)

source1	name1	source2	name2	global_score	vector_similarity	fuzzy_similarity	seuil
greenpeace	blette	marmiton	blette	1.0	1.0	1.0	0.7
greenpeace	carotte	agribalyse	carotte	1.0	1.0	1.0	0.7
greenpeace	carotte	openfoodfacts	carottes	0.789	0.923	0.7	0.7
greenpeace	carotte	marmiton	carotte	1.0	1.0	1.0	0.7
greenpeace	celeri	marmiton	celeri	1.0	1.0	1.0	0.7
greenpeace	carotte	agribalyse	carotte	1.0	1.0	1.0	0.7
greenpeace	celeri	marmiton	celeri	1.0	1.0	1.0	0.7
marmiton	biscuit	openfoodfacts	biscuits	0.749	0.823	0.7	0.7
greenpeace	celeri	marmiton	celeri	1.0	1.0	1.0	0.7
greenpeace	betterave	marmiton	betterave	1.0	1.0	1.0	0.7
greenpeace	carotte	openfoodfacts	carottes	0.789	0.923	0.7	0.7
agribalyse	biscuit tablette chocolat	openfoodfacts	tablette chocolat	0.77	0.887	0.692	0.7
greenpeace	ail	agribalyse	ail	1.0	1.0	1.0	0.7
agribalyse	barre chocolat lait nougat	openfoodfacts	barre chocolat lait praline	0.707	0.886	0.588	0.7
greenpeace	aubergine	agribalyse	aubergine	1.0	1.0	1.0	0.7

Echantillon de paires matchées (global_score > 0.8)

source1	name1	source2	name2	global_score	vector_similarity	fuzzy_similarity	seuil
greenpeace	artichaut	openfoodfacts	artichauts	0.82	0.926	0.75	0.8
greenpeace	ail	openfoodfacts	ail	1.0	1.0	1.0	0.8
greenpeace	aubergine	agribalyse	aubergine	1.0	1.0	1.0	0.8
greenpeace	celeri	agribalyse	celeri	1.0	1.0	1.0	0.8
marmiton	bouillon boeuf	openfoodfacts	bouillon boeuf	1.0	1.0	1.0	0.8
greenpeace	aubergine	marmiton	aubergine	1.0	1.0	1.0	0.8
greenpeace	celeri	marmiton	celeri	1.0	1.0	1.0	0.8
greenpeace	artichaut	openfoodfacts	artichauts	0.82	0.926	0.75	0.8
openfoodfacts	ail	agribalyse	ail	1.0	1.0	1.0	0.8
greenpeace	brocoli	marmiton	brocoli	1.0	1.0	1.0	0.8
greenpeace	aubergine	marmiton	aubergine	1.0	1.0	1.0	0.8
greenpeace	carotte	agribalyse	carotte	1.0	1.0	1.0	0.8
greenpeace	betterave	marmiton	betterave	1.0	1.0	1.0	0.8
greenpeace	celeri	marmiton	celeri	1.0	1.0	1.0	0.8
greenpeace	celeri	agribalyse	celeri	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	77
Fuzzy+vector (liens)	162	1106

L'utilisation du fuzzy+vector permet d'augmenter le nombre de correspondances multi-sources de 8000.0% par rapport au matching exact.