



HAL
open science

Idiomaticité, constructions et variations : approche contrastive sur corpus

Yvon Keromnes

► **To cite this version:**

Yvon Keromnes. Idiomaticité, constructions et variations : approche contrastive sur corpus. *Langages*, 2022, 225, pp.35-47. 10.3917/lang.225.0033 . hal-04249816

HAL Id: hal-04249816

<https://hal.science/hal-04249816>

Submitted on 19 Oct 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Copyright

Yvon Keromnes
Université de Lorraine & ATILF-CNRS (France)

Idiomaticité, constructions et variations : approche contrastive sur corpus

1. INTRODUCTION ¹

Notre propos est d'étudier les expressions du type « ADJ *comme* SN » à valeur intensive dans leurs différentes variantes en allemand, en anglais et en français ². L'intérêt particulier de ces expressions participant à la fois de la syntaxe et du lexique est de pouvoir illustrer plusieurs questions importantes touchant au traitement linguistique de la phraséologie. L'approche proposée ici est à la fois contrastive et constructionnelle (Boas 2013 ; Legallois & Patard 2017), elle a pour but de montrer en quoi ces expressions, au-delà des variations intra- et inter-linguistiques, procèdent d'un même mécanisme à base de comparaison ³ réunissant des expressions lexicalisées et des combinaisons libres pouvant éventuellement se lexicaliser. Les premières sont en effet connues et étudiées depuis longtemps dans le domaine de la phraséologie, où elles sont généralement classées par les lexicographes parmi ce que l'on appelle assez improprement « expressions idiomatiques » ou « expressions figées » ⁴, mais au-delà des instances décrites dans les dictionnaires de locutions, un traitement rapporté à

1. Nous remercions vivement Rainer Perkuhn (IDS Mannheim) et Benjamin Husson (ATILF, Nancy) pour leur aide précieuse, respectivement sur les corpus allemand (DEREKO) et français (FRANTEXT).

2. En notant qu'elles existent dans de très nombreuses autres langues.

3. Sur ce point, nous nous opposons donc à l'idée qu'il y aurait, d'un côté, une véritable comparaison et, de l'autre, une « fausse » comparaison ; voir Buvet & Gross (1995) ; Leroy (2003).

4. « Expression idiomatique » laisse entendre que les expressions ne correspondant pas aux critères définitionnels de cette catégorie, en particulier la non-compositionnalité ne seraient pas idiomatiques ; une approche aussi restrictive de l'idiomaticité paraît difficilement défendable aujourd'hui (voir Hausmann 1997 ; Hilpert 2019) ; l'inconvénient du second terme, comme le note Moon (1998 : 1), est que nombre d'expressions dites « figées » ne le sont pas vraiment.

Aspects de la recherche actuelle en phraséologie

la notion de *construction* paraît plus apte à rendre compte de cette structure abstraite établissant une correspondance entre un adjectif et un syntagme nominal et qui, dans chaque langue qui la possède, peut être considérée comme faisant partie du répertoire linguistique des locuteurs natifs.

L'unité de cette structure abstraite notée « ADJ_{-comp}-SN » est illustrée par des données extraites d'un corpus de référence pour chacune des trois langues considérées. Une telle approche « continuiste »⁵, plutôt que de poser des oppositions binaires entre catégories discrètes et autonomes, telles que *lexique* et *syntaxe* ou *expression figée* et *expression libre*, considère en effet les termes de chaque opposition comme des valeurs extrêmes sur un même axe. De ce point de vue, cette étude se veut illustrative d'une nouvelle approche de la phraséologie débordant un cadre purement lexical. En même temps, elle se veut contribution à des débats actuels au sein des GCx : d'une part, en ce qui concerne la définition d'une construction et, d'autre part, en ce qui concerne la comparabilité entre les langues. D'abord sur la définition des constructions, puisque c'est précisément une telle structure en anglais que P. Kay (2013) exclut du domaine des constructions pour en faire une catégorie particulière (*patterns of coining* « motifs néologiques »), intermédiaire entre le lexique et des structures productives qui seraient les constructions *stricto sensu*. Et pour ce qui est de la comparabilité entre les langues, en opposition à la vision d'une « grammaire universelle », les constructionnistes ont tendance à insister sur la spécificité de chaque langue ; les similitudes frappantes observées entre les trois langues étudiées ici méritent donc explication. Concernant le premier point, nous verrons qu'au-delà des expressions lexicalisées, on observe dans les corpus de nombreux exemples d'expressions libres attestant à la fois de la *productivité* et de la *créativité* de cette comparaison intensive.

Une précision sur ces deux notions parfois utilisées de façon interchangeable : certes, les deux notions peuvent être liées **mais** le premier terme est à réserver aux phénomènes néologiques. **Mais** la simple instanciation d'un schéma abstrait par insertion d'éléments linguistiques aux emplacements prévus, sans qu'il y ait génération d'un nouveau matériau lexical ou d'une nouvelle construction, relève selon nous de la seule productivité (*type frequency*).

La *productivité* des constructions « ADJ_{-comp}-SN » est attestée par le nombre de *hapax* présents dans les corpus et la *créativité* s'observe en particulier dans deux types d'expressions humoristiques, les phénomènes d'inversion sémantique⁶ (*aimable comme une porte de prison*, qui signifie « très peu aimable ») et les défigements (*fier comme un bar tabac*, à partir de *fier comme Artaban*). Ces mêmes propriétés de productivité et de créativité s'observent en allemand et en anglais ; ainsi, le défigement s'observe en allemand dans des expressions telles que *schlank wie eine Tonne* (« mince comme une barrique »), par paronomase à partir de *schlank*

5. Ce trait des grammaires de construction est plus largement une caractéristique de la linguistique cognitive.

6. Sur ce point, nous nous opposons à l'idée d'une inversion syntaxique proposée par Leroy (2004).

Idiomaticité, constructions et variations : approche contrastive sur corpus

wie eine Tanne («mince comme un pin») et, en anglais, avec des expressions telles que *clear as mud* («clair comme de la boue»), qui répond à *clear as crystal*.

Dans cette étude, nous nous limitons à trois langues pour lesquelles nous sommes à même de juger de l’idiomaticité d’une expression, mais l’on peut aisément proposer des exemples dans d’autres langues. En russe, par exemple :

- (1) ясно как день – jasno kak den’
«clair comme le jour»
- (2) старо как мир – staro kak mir
«vieux comme le monde»
- (3) здоров как бык – zdorov kak byk
«en bonne santé comme un taureau»

Mais s’il s’agit encore de langues indo-européennes, qui possèdent assurément un fonds phraséologique commun⁷, de telles constructions sont aussi communes dans des langues non apparentées telles que le chinois ou le japonais. Ainsi, en chinois :

- (4) 安如磐石 – ān rú pán shí lán
«stable comme un roc»
- (5) 烂醉如泥 – làn zuì rú ní
«saoul comme de la boue»
- (6) 蠢笨如猪 – chǔn bèn rú zhū
«stupide comme un cochon»

et en japonais :

- (7) 石のように硬い – ishi no you ni katai
«dur comme la pierre»
- (8) 氷のように冷たい – kouri no you ni tsumetai
«froid comme la glace»
- (9) 桜のように儂い – sakura no you ni hakanai
«éphémère comme une fleur de cerisier»⁸

Dans ces différentes langues, et certainement dans beaucoup d’autres, on observe donc des expressions lexicalisées comportant un marqueur de comparaison. Ces propriétés partagées et une même variété d’emplois, du libre au figé, permettent de supposer l’existence d’un même mécanisme cognitif exploitant la structure comparative des constructions du type « ADJ_{-comp}-SN » pour produire de l’intensif. Dans une première approche, disons qu’à la base d’un tel emploi se trouve le recours dans le SN à un prototype⁹.

7. À cet égard, voir p. ex. Piirainen (2012).

8. Avec une syntaxe inversée, soit littéralement *pierre-comme-dur*, *glace-comme-froid*, etc.

9. Ce que nous distinguerons de la notion de *parangon* chez Rivara (1990) et Leroy (2004).

Aspects de la recherche actuelle en phraséologie

2. EXPRESSIONS IDIOMATIQUES ET TRAITEMENT LEXICOGRAPHIQUE

Les constructions de ce type sont connues de la phraséologie depuis longtemps. En allemand, nous avons la construction « ADJ *wie* SN », où *wie* constitue le marqueur de comparaison (*klar wie Kloßbrühe* «clair comme de l'eau de roche»¹⁰) ; en anglais, deux formes de comparaison concurrentes, « (as) ADJ *as* SN » et « ADJ *like* SN » (*busy as a bee* «laborieux comme une abeille», *crazy like a fox* «fou comme un renard»).

Les études phraséologiques traditionnelles des constructions de ce type s'appuient régulièrement sur des ressources lexicographiques, lesquelles reflètent à leur tour les *a priori* dont procèdent les premières. Voici quelques exemples représentatifs de ces ressources dans les trois langues étudiées. En français, A. Rey et S. Chantreau (1989), souvent cité¹¹, décrit 10 000 expressions et locutions. En allemand, une ressource connue est le Röhrich¹² (Röhrich 1991), qui compte 15 000 expressions idiomatiques et proverbiales. En anglais, il nous faut mentionner plusieurs ouvrages de référence : d'une part, pour l'anglais britannique, deux dictionnaires : A. P. Cowie, R. Mackin et I. R. McCaig (1991), qui décrit 7 000 expressions idiomatiques, et J. Seidl et W. McMordie (1988), qui ne compte qu'un peu plus de 3 000 entrées ; d'autre part, pour l'américain, R. A. Spears (2007). Ces ouvrages comportent tous un index, mais notons que les mots grammaticaux tels que les marqueurs de comparaison n'y figurent pas toujours (ils en sont exclus chez Cowie, Mackin & McCaig 1991 et chez Spears 2007) ; et même lorsqu'ils y sont, on s'aperçoit rapidement que les renvois sont tout sauf systématiques ; pour le français par exemple, chez A. Rey et S. Chantreau (1989), *comme* figure dans l'index et renvoie à des locutions telles que *comme un dieu*, *comme un pape*, *comme un poisson dans l'eau* ; en revanche, on ne trouve aucun renvoi vers *singe*, bien que le dictionnaire comporte des expressions telles que *agile comme un singe*, *laid comme un singe* et *malin comme un singe*. De même, chez L. Röhrich (1991), *wie* («comme») figure dans l'index, mais il ne renvoie pas systématiquement aux adjectifs concernés par notre construction, par exemple à *dumm* («bête»), alors que *so dumm wie Brot* (lit. «bête comme du pain») figure dans l'ouvrage. En anglais, la variation autour de ces constructions y est également sacrifiée, puisque J. Seidl et W. McMordie (1988) ne présentent que la construction « *as* ADJ *as* SN », alors qu'elle est totalement absente de R. A. Spears (2007) et que la variation entre « *as* ADJ *as* SN » et « ADJ *as* SN » paraît anecdotique

10. Litt. «clair comme un bouillon de boulettes de pâte» (*KlöÙe* ou *Knödel*). Ironiquement, le sens de cette expression n'est pas clair pour tous les germanophones (la présence habituelle d'amidon devrait rendre le bouillon trouble) et, sur différents forums internet, on se demande si elle n'exprime pas plutôt une absence de clarté, comparable au *clear as mud* anglais.

11. Ainsi, Leroy (2004 : 255) annonce baser son étude sur « un corpus de plus de 300 occurrences » tirées de cet ouvrage.

12. La première édition date de 1973, nos observations sont basées sur l'édition de 1991. Son champ est plus vaste que celui des autres ouvrages cités ici, puisqu'il comprend des proverbes et des dictons.

Idiomatité, constructions et variations : approche contrastive sur corpus

chez A. P. Cowie, R. Mackin et I. R. McCaig (1991), où elle apparaît sous la forme d’une mise entre parenthèses aléatoire. Ainsi, après l’entrée (*as*) *black as coal*, dans laquelle le premier *as* est présenté comme optionnel, nous trouvons *as black as ink*, une entrée elle-même illustrée par l’exemple suivant : *the sky had grown black as ink*. Malgré ses dimensions modestes, J. Seidl et W. McMordie (1988) constitue une notable exception, puisque le choix y a été fait de regrouper les *idioms* selon des critères grammaticaux (SN idiomatiques, paires de verbes, de noms, d’adjectifs idiomatiques, etc.) ou sémantiques (*idioms* relevant d’un même sujet spécifique, tel que la santé, ou dont les mots-clés appartiennent à une même catégorie : animaux, couleurs, etc.) ; c’est ainsi que dans le dernier chapitre, *Idioms with comparisons*, nous trouvons les constructions « *as* ADJ *as* SN » et « V *like* SN ». Dans les différents ouvrages cités, le traitement lexicographique des expressions qui nous intéressent paraît donc très incomplet ; il ignore une grande part de la variation existante tout en masquant la généralité du phénomène.

3. ÉTUDES DE CORPUS

3.1. Trois corpus

Sur le fonds lexicographique présenté ici s’effectuent de nombreuses études phraséologiques dans lesquelles on s’intéresse principalement au phénomène du *figement* (Mejri 1994 ; Schapira 2000) et, même si T. Szende (1999) estime que cette construction correspond plutôt à un phénomène collocationnel, tous les auteurs s’accordent sur la nécessité d’apprendre individuellement ces expressions. De fait, les ouvrages lexicographiques précédemment cités s’adressent tout particulièrement aux locuteurs non natifs.

Mais si en consultant ces ressources, on peut avoir l’impression d’un phénomène lexical nettement circonscrit, il en est autrement dès que l’on étend l’observation à un usage effectif de la langue : de nombreuses occurrences ne sont pas figées et toutes ne sont pas intensives. Il faut donc admettre l’existence d’une ambiguïté constructionnelle opposant en français « ADJ *comme* SN »₁, comparaison entre deux entités paraphrasable par « *aussi* ADJ *que* SN », et « ADJ *comme* SN »₂ paraphrasable par « *très* ADJ »¹³. Cette ambiguïté, parmi d’autres facteurs détaillés *infra*, contribue à créer un « bruit » plus ou moins important nécessitant un « nettoyage » conséquent des résultats.

Par ailleurs, les trois corpus de référence – le COCA pour l’anglais, FRANTEXT pour le français et le DEREKO pour l’allemand – présentent des caractéristiques extrêmement différentes :

13. Mais nous verrons *infra* que le problème est plus complexe.

Aspects de la recherche actuelle en phraséologie

- le COCA, représentatif de l’anglais américain entre 1990 et 2017, est de taille importante (560 millions de mots ¹⁴), il est équilibré et annoté, ce qui permet de rechercher directement des constructions abstraites telles que « ADJ *as* SN ». Les résultats sont présentés sous formes de *types* autour des pivots adjectivaux ¹⁵. Parmi les trois langues étudiées, c’est le corpus le plus représentatif d’une variété d’anglais (anglais américain) actuellement écrite et parlée ;
- FRANTEXT, le corpus de référence du français (254 millions de mots) couvre de façon inégale le français écrit du XII^e au XXI^e siècle. Par ailleurs, il n’est pas vraiment représentatif de la langue française, la littérature, plus précisément celle des « grands auteurs », y étant surreprésentée ¹⁶. En revanche, depuis 2018, il est également annoté et permet une recherche directe de « ADJ *comme* SN ». Pour rendre une comparaison avec l’anglais plus pertinente, nous avons créé un sous-corpus sur la base des dates du COCA. Comptant un peu plus de 23 millions de mots pour **372 textes**, ce corpus est d’une taille notablement inférieure à celle du corpus américain, sans pour autant être négligeable. Enfin, compte tenu des caractéristiques évoquées *supra*, ce corpus peut être jugé représentatif d’un usage de la construction qui nous intéresse en français écrit ¹⁷ ;
- le DEREKO, corpus de référence de la langue allemande, est de très grande taille (l’archive « écrit » accessible au public approche les 10 milliards de mots), il n’est pas non plus équilibré, puisque même s’il contient des types de textes variés, la presse, nationale et régionale, y est très majoritaire ¹⁸. À partir de l’archive que nous venons d’évoquer, nous avons créé un sous-corpus reprenant les dates qui encadrent le COCA. Ce corpus compte près de 9 milliards 827 millions de mots ; la quasi-totalité de cette archive est donc située dans cette même période et la réduction de l’empan chronologique est pratiquement sans effet ; ce corpus n’est pas non plus annoté ; une comparaison directe avec FRANTEXT et le COCA n’est pas possible, et il faut pour l’exploiter recourir à une méthodologie particulière, présentée *infra*.

3.2. Trois constructions en anglais

Comparés aux dizaines d’expressions décrites dans les ressources lexicographiques, les résultats obtenus sur les corpus révèlent la très grande richesse des emplois effectifs de ces constructions. Ainsi, le schéma « *as* ADJ *as*

14. Tous les chiffres relatifs à la taille des trois corpus de référence et les fréquences observées portent sur lesdits corpus au moment de la rédaction de cet article, les trois corpus faisant régulièrement l’objet d’ajouts.

15. Les occurrences d’un même adjectif dans la construction étudiée sont regroupées.

16. À l’origine, le projet du *Trésor de la Langue Française*, dictionnaire pour lequel a été constitué FRANTEXT, était de supplanter le dictionnaire de l’Académie française, dont la rédaction était jugée trop lente, mais ce dernier a manifestement inspiré la constitution de FRANTEXT.

17. Alors que FRANTEXT dans son ensemble compte 90 % de textes littéraires, notre sous-corpus n’en compte que 46 % (**172 textes**, dont 17 de poésie), le reste étant constitué de non-fiction.

18. Il s’agit d’un choix délibéré des responsables de la constitution de ce corpus à l’IDS, qui ont estimé que la taille, plus que l’équilibre des genres discursifs, est primordiale pour étudier un phénomène linguistique donné.

Idiomatité, constructions et variations : approche contrastive sur corpus

*NOUN »¹⁹ dans le COCA correspond à 30 428 combinaisons lexicales (*types*) pour 40 300 occurrences distinctes (*tokens*). Il faut bien sûr éliminer un certain nombre de faux positifs, en particulier les comparaisons objectives de deux entités (cf. *infra*), mais le nombre d'exemples pertinents reste malgré tout élevé. Notons également la proportion très importante de *hapax* (dès le rang 2630, soit 91 % des *types*), parmi lesquels le bruit semble plus important que parmi les types plus fréquents, mais qui contiennent également des positifs manifestes : *as bare as an egg* <chauve comme un œuf>, *as beautiful as an angel* <belle comme un ange>, *as big as a garage* <(une bouche) grande comme un garage>, etc. Mais outre le nombre de résultats, l'examen du COCA révèle ce qui semble être l'existence de constructions concurrentes ; alors que J. Seidl et W. McMordie (1988) ne s'intéressaient qu'au schéma « *as ADJ as SN* », P. Kay (2013) et G. Desagulier (2015) ne traitent, pour leur part, que de « *ADJ as SN* ». Le tableau suivant compare les deux schémas pour les 20 expressions les plus fréquentes.

Tableau 1 : Comparaison entre « *as ADJ as SN* » et « *ADJ as SN* »

rang	« <i>as ADJ as SN</i> »	f.	« <i>ADJ as SN</i> »	f.
1	<i>as American as apple pie</i>	53	<i>mad as hell</i>	213
2	<i>as old as time</i>	41	<i>tough as nails</i>	116
3	<i>as good as gold</i>	38	<i>clear as a bell</i>	73
4	<i>as big as a house</i>	29	<i>American as apple pie</i>	68
5	<i>as clear as day</i>	27	<i>white as snow</i>	68
6	<i>as white as snow</i>	25	<i>clear as day</i>	65
7	<i>as clear as a bell</i>	23	<i>cold as ice</i>	62
8	<i>as cold as ice</i>	21	<i>good as gold</i>	60
9	<i>as different as night and day</i>	21	<i>white as a sheet</i>	55
10	<i>as big as a football</i>	19	<i>smooth as silk</i>	54
11	<i>as tall as a man</i>	19	<i>hot as hell</i>	52
12	<i>as old as the hills</i>	18	<i>plain as day</i>	52
13	<i>as big as a man</i>	16	<i>big as a house</i>	42
14	<i>as bright as day</i>	16	<i>old as time</i>	42
15	<i>as bright as the sun</i>	16	<i>black as night</i>	41
16	<i>as light as air</i>	16	<i>free as a bird</i>	41
17	<i>as old as humanity</i>	15	<i>high as a kite</i>	41
18	<i>as smooth as silk</i>	15	<i>smart as a whip</i>	40
19	<i>as plain as the nose on your/my face</i>	14	<i>nervous as hell</i>	38
20	<i>as thick as a man's torso/thigh/body</i>	14	<i>bright as day</i>	36

19. Requête prévoyant la présence d'un élément – le plus souvent d'un déterminant – devant le substantif.

Aspects de la recherche actuelle en phraséologie

Évidemment, le second schéma subsume le premier, mais si nous examinons les occurrences les plus fréquentes dans les deux cas, pour « *as ADJ as SN* », à partir des 200 premiers résultats et après élimination des faux positifs, nous obtenons 102 expressions différentes (*types*) avec des fréquences (*token frequency*) allant de 53 à 4, avec 48 adjectifs différents, le plus productif étant *big*, qui se retrouve dans 22 *types*, suivi de *old* (7), *white* (6) et *clear* (5). Pour « *ADJ as SN* », en procédant de la même façon, nous obtenons 145 expressions différentes, avec des fréquences allant de 213 à 9, et 72 adjectifs différents, les plus productifs étant à nouveau *big* (présent dans 9 *types*), *black* (6), *white* (6), *clear* (5) et *old* (5). Dans les deux cas, les mêmes adjectifs semblent pouvoir participer à plusieurs expressions ; en revanche, le nombre d’expressions différentes est sensiblement supérieur, ce qui semble suggérer une plus grande productivité du deuxième schéma. L’examen du Tableau 1 renforce encore l’hypothèse de deux constructions différentes²⁰ : les expressions les plus fréquentes ne sont globalement pas les mêmes et les fréquences sont également très différentes ; autrement dit, chaque construction attire différents lexèmes, dans des expressions qui possèdent une accessibilité différente. Par exemple, l’expression la plus fréquente de la deuxième construction – *mad as hell* – ne figure même pas dans les 200 expressions les plus fréquentes de la première construction (*as mad as hell* a une fréquence de 4).

Quant au schéma « *ADJ like SN* », très largement absent des travaux de phraséologie, de phraséographie comme des recherches linguistiques, il apparaît pourtant comme une construction à la fois comparable aux deux précédentes et possédant des caractéristiques propres.

Tableau 2 : La construction « *ADJ like SN* »

rang	« <i>ADJ like SN</i> »	f.	rang	« <i>ADJ like SN</i> »	f.
1	<i>crazy like a fox</i>	37	11	<i>sweet like chocolate</i>	3
2	<i>easy like Sunday morning</i>	5	12	<i>white like a ghost</i>	3
3	<i>hot like fire</i>	4	13	<i>yellow like the sun</i>	3
4	<i>hungry like the wolf</i>	4	14	<i>asleep like a baby</i>	2
5	<i>soft like a baby’s hand/mouth/head</i>	3	15	<i>big like mountains</i>	2
6	<i>black like death</i>	3	16	<i>blue like heaven</i>	2
7	<i>cold like ice</i>	3	17	<i>blue like the sky</i>	2
8	<i>heavy like a sack</i>	3	18	<i>bright like a diamond</i>	2
9	<i>quick like a bunny</i>	3	19	<i>bright like christmas lights</i>	2
10	<i>sly like a fox</i>	3	20	<i>bright like the sun</i>	2

20. On trouve peu de remarques sur la différence entre les deux, mais Quirk *et al.* (1985 : 1137) notent que le schéma ne comportant qu’un seul *as* constitue une comparaison moins emphatique (ce qui rend peut-être l’interprétation intensive plus accessible) et que, dans ce schéma, *as* est plus proche de *like*.

Idiomatité, constructions et variations : approche contrastive sur corpus

Pour les fréquences brutes (avant élimination des faux positifs), le COCA indique 4 693 *types* pour 5 182 *tokens*, soit des fréquences très inférieures aux deux constructions précédentes, mais une proportion particulièrement forte de *hapax*, indiquant la productivité de cette construction. Le nombre de faux positifs est plus élevé ; ainsi, après leur élimination des 200 expressions les plus fréquentes, nous n’obtenons que 83 *types*, avec 51 adjectifs différents, le plus fréquent étant *black* (12) – dans des expressions telles que *black like {charcoal / death / ink / ... / beetle backs at midnight}* –, suivi par *soft* (4), *sly* (2) et *white* (2) (*white like {a ghost / paper}*). Le Tableau 2 montre à la fois des expressions spécifiques telles que le proverbial ²¹ *crazy like a fox* (qui possède une fréquence nettement supérieure aux autres expressions) et *cold like ice*, avec des adjectifs et noms qui se retrouvent associés dans les deux constructions précédentes. Nous nous intéressons *infra* au sens de ces constructions ainsi qu’à celui d’expressions spécifiques.

3.3. Comparabilité, sens constructionnel et spécificité

En dépit des différences de contenus et de structures entre les corpus de référence dans les trois langues, et après avoir aligné les dates des sous-corpus dans les trois langues, nous nous sommes efforcé d’établir une base de comparaison la plus rigoureuse possible entre ces langues. Pour le DEREKO, la démarche a consisté à réduire au maximum le nombre des candidats présumés ²². À cette fin ont été extraits les 5-grammes comportant le marqueur de comparaison *wie* en deuxième position. Cette liste a ensuite été lemmatisée et étiquetée avec le programme GLEMM de l’IDS, pour finalement permettre l’obtention d’une liste de 5-grammes correspondant au patron « ADJ *wie* _ _ _ » avec un nom en troisième, quatrième ou cinquième position. Cette liste comporte 758 936 *tokens*, avec 21 864 adjectifs différents, dont 9 929 *hapax*, soit un peu plus de 45 %. Le nombre de faux positifs est cependant particulièrement important. Notons par exemple que les trois termes retenus comme adjectifs les plus fréquents par le lemmatiseur – *ähnlich* (‘semblable’, ‘ressemblant’), *gut* (‘bon’, ‘bien’) et *genau* ²³ (‘exact’), correspondent presque exclusivement à des faux positifs, puisque l’expression *ähnlich wie* est le plus souvent employée pour désigner une

21. Cette expression est mentionnée chez Röhrich (1991) qui, l’interprétant littéralement, y voit une différence culturelle (en Europe, le renard est rusé, chez les Américains, il serait fou). En fait, il s’agit d’un cas apparenté à l’inversion sémantique : le renard, dans sa ruse, est capable de se faire passer pour fou afin de mieux obtenir ce qu’il souhaite.

22. La démarche empruntée ici nécessite une lemmatisation *ad hoc* d’un corpus de travail de très grande taille, elle n’est pas possible pour un utilisateur extérieur du DEREKO.

23. La fréquence de ces adjectifs dans le patron recherché s’élève à plusieurs dizaines de milliers (près de 5 000 pour *gross*, au rang 4). À noter, l’orthographe de *gross*, qui n’est pas celle de l’allemand standard (*groß*) sauf en Suisse, est malgré tout reconnue comme identique à la seconde dans le DEREKO. Dans le fichier d’adjectifs candidats pour cette construction, que nous avons reçu, figurait la première orthographe, c’est donc elle que nous avons conservée. Ces fréquences ne sont pas en soi inintéressantes mais, étant donné que le nombre de faux positifs peut varier d’un adjectif à l’autre, elles ne reflètent pas directement la fréquence de la construction qui nous intéresse, c’est pourquoi nous avons renoncé à les indiquer. Par simplicité, nous limitons

Aspects de la recherche actuelle en phraséologie

comparaison approximative (‘un peu comme...’), *so gut wie* a le plus souvent le sens de ‘pour ainsi dire’ et *genau wie* renvoie généralement à une équivalence exacte (‘tout comme...’). Le sens « très ADJ » ne peut cependant être exclu *a priori*, puisque, par exemple, des expressions telles que *ähnlich wie zwei Tropfen Wasser* (‘aussi semblables que deux gouttes d’eau’) existent. Mais il faut arriver au 4^e adjectif le plus fréquent, *gross* (‘grand’/‘gros’) pour être relativement certain de trouver des occurrences correspondant à ce que nous cherchons, par exemple *gross wie ein Haus* (‘grand/gros comme une maison’) :

- (10) Er scheint ein Herz zu haben **gross wie ein Haus**. (*Neue Zürcher Zeitung Folio*, 03/01/2000)
 ‘Il semble avoir un cœur gros comme une maison.’
- (11) Die Sau kommt auf mich zu, sie ist nun **gross wie ein Haus** [...]. (*Neue Zürcher Zeitung*, 17/08/2015)
 ‘La truie fonce sur moi, là, elle est aussi grosse qu’une maison [...].’

Autour de cette expression se trouvent également des variantes de comparant, *Einfamilienhaus* (‘maison individuelle’), *Mehrfamilienhaus* (‘immeuble’), comme dans l’exemple suivant, où il s’agit de moteurs de bateaux :

- (12) Das sind Riesendinger, so **gross wie Einfamilienhäuser**. (*St. Galler Tagblatt*, 10/06/2000)
 ‘Ce sont des trucs gigantesques, grands comme des maisons individuelles.’

Mais nous notons aussi que l’adjectif *gross* se retrouve dans de nombreux cas d’inversion sémantique, où il sert à exprimer la petitesse :

- (13) Der Speicherchip, so **gross wie ein Daumennagel**, birgt rund 20 000 Songs. (*Neue Zürcher Zeitung*, 05/04/2016)
 ‘La puce-mémoire, grosse comme un ongle de pouce, contient environ 20 000 chansons.’

Parmi les 30 adjectifs les plus fréquents de la liste, nous trouvons, outre *gross*, 14 candidats prometteurs pour notre construction, avec des fréquences allant de 30 000 à 4 000 :

hoch (‘haut’), *stark* (‘fort’), *alt* (‘vieux’), *schnell*²⁴ (‘rapide’), *weit* (‘étendu’, ‘éloigné’), *lang* (‘long’), *teuer* (‘cher’), *einfach* (‘simple’), *bekannt*²⁵ (‘connu’), *sicher* (‘sûr’), *schön* (‘beau’), *klein* (‘petit’), *schwer* (‘lourd’), *hart* (‘dur’, ‘difficile’).

Notons également que la douzaine d’adjectifs illustrant chez L. Röhrich (1991) notre construction se retrouve dans notre liste :

arm (‘pauvre’), *bitter* (‘amer’), *frech* (‘insolent’), *geschmückt* (‘décoré’), *hell* (‘clair’), *keusch* (‘chaste’), *platt* (‘plat’, ‘étonné’), *pünktlich* (‘ponctuel’), *schlank* (‘mince’), *stolz* (‘fier’).

le nombre d’équivalents et nous en tenons à la forme masculine, mais tous ces adjectifs peuvent également correspondre à des féminins.

24. Ce mot s’emploie cependant aussi fréquemment en tant qu’adverbe (*vite*).

25. L’expression *bekannt wie der bunte Hund* (‘connu comme le chien multicolore’) est probablement présente dans tous les recueils d’expressions idiomatiques.

Idiomatité, constructions et variations : approche contrastive sur corpus

Les fréquences, à deux exceptions près ²⁶, sont de l'ordre de plusieurs centaines, et *stolz* approche même les 2 000 occurrences. Mais même sans aller jusqu'aux *hapax*, dont le grand nombre suggère une importante productivité de cette construction, on trouve jusque dans les adjectifs peu fréquents de nombreux exemples probants qui ont jusqu'ici échappé à l'attention des lexicographes. Ainsi, *zugeknöpft* (litt. «boutonné jusqu'en haut», c.-à-d. «peu communicatif», «fermé»), avec en particulier plusieurs occurrences de l'expression *zugeknöpft wie eine Auster* («fermé comme une huître»).

Pour FRANTEXT et le COCA, les résultats peuvent être directement alignés : dans FRANTEXT, nous obtenons 2 207 occurrences du schéma « ADJ *comme* SN », regroupées manuellement, sur la même base que le COCA (adjectif pivot), en 1936 *types* : ici davantage que pour l'anglais, nous observons une forte proportion de *hapax*, avec 95 % des *types*. Et à nouveau, nous avons affaire à un certain nombre de faux positifs, dus par exemple à l'ambiguïté du marqueur *comme*, qui peut avoir la valeur de «tel que», introduisant un exemple (14), ou de «en tant que», spécifiant une catégorie (15) :

- (14) Chaque fois que l'on qualifie de « cancer » une réalité collective **menaçante comme le chômage**, la délinquance, l'alcoolisme [...]. (Forrest, *Tous les enfants sauf un*, 2007)
- (15) T'es pas **normale comme fille** et si t'étais un minimum honnête, t'irais te faire soigner [...]. (Gavalda, *Ensemble, c'est tout*, 2004)

Mais une véritable base de comparaison demande l'élimination des faux positifs, ce qui s'avère rapidement impossible à réaliser sur la totalité des corpus, cette tâche nécessitant l'examen attentif et individuel de la plupart des occurrences. Même en procédant à un échantillonnage aléatoire pour prélever dans chaque corpus 500 *types* « ADJ-comp-SN », l'élimination des faux positifs se révèle impossible à réaliser, en particulier à cause de cas d'indécidabilité face à l'ambiguïté constructionnelle mentionnée *supra* et qui oppose en français deux « ADJ *comme* SN », paraphrasables par « *aussi* ADJ *que* SN » ou « *très* ADJ ». Cette ambiguïté, qui existe aussi en allemand et en anglais, nous conduit à nous interroger sur le sens constructionnel de la construction intensive, souvent hyperbolique, appelée dans plusieurs travaux « comparaison à parangon » (Rivara 1990 ; Leroy 2004). Et si nous considérons ces comparaisons comme instances d'une même construction, celle-ci aurait pour signification le haut degré de l'adjectif qui s'y trouve. L'hypothèse d'un tel sens constructionnel semble confortée par l'existence d'expressions relativement peu transparentes telles que

26. Les deux adjectifs aux fréquences notablement basses dans le patron recherché sont *keusch* (43) et *geschmückt* (*hapax*). Si les fréquences dans notre liste sont relativement imprévisibles dans leur ensemble, ces basses fréquences dénotent peut-être le caractère daté de certaines expressions, liées à des mœurs qui ont évolué et à des coutumes rurales pour le *hapax* : chez Röhrich (1991), on trouve en effet *keusch wie eine Braut* («chaste comme une fiancée») et *geschmückt wie ein Pflingstochse* («décoré comme un bœuf – que l'on parade – pour la Pentecôte»).

Aspects de la recherche actuelle en phraséologie

*bête comme ses pieds*²⁷ ou *dumm wie Brot* (‘bête comme du pain’) en allemand, ainsi que des segments relativement flexibles tels que *comme tout* en français (*bête/facile/mignon comme tout*) ou *as hell* en anglais (*crazy/funny/sexy as hell*, parmi beaucoup d’autres expressions présentes dans le COCA), qui ne peuvent guère s’interpréter qu’à partir de ce sens constructionnel. *A priori*, il semble facile de distinguer les deux interprétations, *égalité* et *intensité* :

(16) Jane is **as cool as Sarah**.

(17) Jane is **as cool as a cucumber**.

En (16), je compare Jane à Sarah mais, en (17), j’exprime le haut degré auquel Jane est « cool ». Ce qui fait dire à J.-C. Milner (1973 : 39)²⁸ que, dans de tels cas, « l’assertion porte sur l’adjectif [...] et non sur l’égalité ». Pourtant, l’adjectif n’est pas nécessairement seul en cause : d’une part, la comparaison est détournée mais n’a pas nécessairement disparu et, d’autre part, il ne s’agit pas seulement d’intensité. Si je dis que *Claude est lent comme une tortue* ou que *Daniel est lent comme un escargot*, je veux évidemment dire qu’ils sont très lents, mais je les compare en même temps aux animaux nommés, si bien que j’ai affaire à ce que G. Fauconnier et M. Turner (2002) nomment une « intégration conceptuelle » (*blending*) : *Quelle tortue, ce Claude ! Et Daniel, quel escargot !* Par ailleurs, le choix du comparant ne se limite pas à l’expression du haut degré : *maigre comme un clou* n’évoque pas simplement la maigreur mais également le caractère décharné d’une personne « toute en os » ; de même que *schlank wie eine Tanne* (‘mince comme un pin’) n’évoque pas simplement la sveltesse mais également une grande taille et une posture droite. Qui plus est, les variations de comparants, qui semblent s’opposer à l’idée d’un parangon, supposé unique, correspondent souvent à une spécialisation sémantique : *blanc comme neige* évoquera souvent l’innocence, alors que *blanc comme un linge* évoquera plutôt la peur ou la maladie. De même, une banane ne renvoie sans doute pas intrinsèquement à l’expression d’une grande taille mais, dans les exemples suivants, c’est bien le cas :

(18) Die Zähne der Raubsaurier waren **so groß wie Bananen**, die Tiere wogen mehr als vier Mittelklasse-Autos und waren dreieinhalbmal so lang. (*Braunschweiger Zeitung*, 17/12/2005)

‘Les dents de ces sauriens prédateurs étaient grosses comme des bananes, ces animaux pesaient quatre fois le poids d’une voiture de classe moyenne et étaient trois fois et demie plus longs.’

(19) Sein Netz verhedderte sich in einem Felsen, und als er ins Wasser stieg, um es loszuzerren, wurde er von einer grossen Hand am Knöchel gepackt, von einer riesigen Hand mit Fingern **so gross wie Bananen**. (*Weltwoche*, 01/12/2005)

‘Son filet s’est coincé sur un rocher, et lorsqu’il est allé dans l’eau pour le

27. Comme le note Mejri (1994), le substantif *pied* ne contient pas de sème de la bêtise. C’est à notre sens un argument contre une explication du sens constructionnel en termes de « redondance sémique ».

28. Cité par Leroy (2007 : 72).

Idiomatité, constructions et variations : approche contrastive sur corpus

décoincer, il a été saisi à la cheville par une grosse main, par une main gigantesque, avec des doigts gros comme des bananes.»

Des dents de prédateurs ou des doigts comparés à des bananes sont bien des objets de grande taille, mais des exemples comme le suivant sont plus difficiles à traiter :

- (20) Parfois, entre deux lampées de potage, il relevait un peu la tête et laissait filer entre ses cils un regard pointu, étincelant et **vif comme une truite qui fuit**. (Chandernagor, *L'Enfant des lumières*, 1995)

Dans ce type d'exemple finalement assez fréquent, la comparaison entre deux entités est patente, avec la double qualification, et, en même temps, il semble bien que la valeur intensive soit actualisée, au moins pour l'adjectif *vif*, la truite qui fuit illustrant de façon prototypique cette vivacité. Ce genre d'exemples rencontrés dans notre sous-corpus nous suggère que, si certains cas d'ambiguïté constructionnelle peuvent être clairement tranchés, d'autres sont plus délicats à évaluer, et nous pouvons supposer qu'il est possible de passer graduellement d'une interprétation à une autre.

4. CONCLUSION

Une approche purement quantitative de la question traitée ici s'avère insuffisante : sans un examen attentif de chaque occurrence, on ne peut vraiment savoir à quoi renvoient les chiffres indiqués. Par ailleurs, une telle approche laisse de côté de nombreux phénomènes liés à la variation inter- et intralinguistique de cette famille de constructions, qui nécessitent une approche qualitative. En revanche, le recours aux corpus est indispensable en ce qu'il permet la mise en évidence de ce que les constructions du type « ADJ *comme* SN » à valeur intensive en allemand, en anglais et en français, existent très majoritairement dans des emplois parfaitement compositionnels, à côté des emplois lexicalisés. Ces familles de constructions constituent des catégories radiales (Lakoff 1987), dont les membres peuvent posséder des propriétés sémantiques plus ou moins idiosyncratiques. Les corpus permettent également de mettre en évidence la productivité²⁹ et la créativité dont ces constructions sont la source. Cette réalité est masquée dans les études phraséographiques, qui traitent généralement chaque expression indépendamment, sans mise en évidence de la construction qu'elle instancie. Il semble pourtant bien, dans les trois langues, s'agir d'une même catégorie au sein de laquelle l'emploi intensif présuppose une valeur prototypique du syntagme nominal quant à la propriété de l'adjectif auquel il est associé. L'existence de ces constructions dans différentes langues associant par l'intermédiaire d'un marqueur de comparaison un adjectif et un syntagme nominal, avec

29. Certes, l'instanciation d'une construction « ADJ-comp-SN » n'est pas aussi productive qu'une construction « ADJ N » en allemand, en anglais ou en français, mais cela ne justifie pas à notre sens d'en faire un schéma à part, comme le propose Kay (2013). Comme le note Goldberg (2019), la productivité des constructions n'est pas absolue, elle varie d'une construction à l'autre.

Aspects de la recherche actuelle en phraséologie

pour fonction d’exprimer un haut degré de l’adjectif (ou parfois au contraire un très bas degré), ne doit pas s’expliquer par une analogie de forme mais, **sans doute**, plutôt par une identité fonctionnelle : l’utilisation du mécanisme comparatif pour l’obtention d’un effet intensif peut **sans doute** s’interpréter comme un processus figuratif qui permet la représentation d’une qualité sous la forme d’une réalité tangible possédant de façon exemplaire cette qualité. À l’instar de ce qui se passe dans les métaphores conceptuelles décrites initialement par G. Lakoff et M. Johnson (1980), nous observons une « réification conceptuelle » (Langacker 2008) indirecte, l’adjectif étant en quelque sorte réifié « par procuration ». Et dès lors que la construction, dans cet emploi intensif, est suffisamment présente dans la langue, elle acquiert elle-même cette valeur indépendamment de ses constituants, ce qui explique à la fois l’existence et la longévité d’expressions comportant des comparants parfaitement opaques et apparemment non motivés, parfois des noms propres n’évoquant aujourd’hui plus rien pour la plupart des locuteurs (*fier comme Artaban*, *frech wie Oskar*, etc.). De telles expressions, nous l’avons vu, peuvent ainsi constituer des schémas partiellement productifs – « ADJ *comme tout* », « ADJ *comme choux* », « ADJ *like hell* » – qui illustrent le principe de « coercition sémantique » décrit par L. A. Michaelis (2004).

Références

- [COCA] *Corpus of Contemporary American English* (1990-2017), Mark Davies, Brigham Young University. [<https://www.english-corpora.org/coca/>]
- [DEREKO] *Das Deutsche Referenzkorpus* (1964-), Institut für Deutsche Sprache – Leibniz. [<http://www1.ids-mannheim.de/kl/projekte/korpora/>]
- [FRANTEXT] *Base textuelle Frantext*, ATILF (CNRS & Université de Lorraine). [www.frantext.fr]
- BOAS H. (2013), “Cognitive construction grammar”, in T. Hoffmann & G. Trousdale (eds.), *The Oxford Handbook of Construction Grammar*, Oxford, Oxford University Press, 233-254.
- BUVET P.-A. & GROSS G. (1995), « Comparaison et expression du haut degré dans le groupe nominal », *Faits de langues* 5, 83-88.
- COWIE A. P., MACKIN R. & MCCAIG I. R. (1991⁷), *Oxford Dictionary of Current Idiomatic English*, vol. 2: *Phrase, Clause & Sentence Idioms*, Oxford, Oxford University Press.
- DESAGULIER G. (2015), « Le statut de la fréquence dans les grammaires de construction : *simple comme bonjour* ? », *Langages* 197, 99-128.
- DOBROVOL'SKIJ D. (2011), “Phraseologie und Konstruktionsgrammatik”, in A. Lasch & A. Ziem (eds), *Konstruktionsgrammatik III: Aktuelle Fragen und Lösungsansätze*, Tübingen, Stauffenburg, 111-130.
- FAUCONNIER G. & TURNER M. (2002), *The Way We Think: Conceptual Blending and the Mind's Hidden Complexities*, New York, Basic Books.
- GOLDBERG A. E. (2019), *Explain Me This: Creativity, Competition, and the Partial Productivity of Constructions*, Princeton, Princeton University Press.
- HAUSMANN F. J. (1997), « Tout est idiomatique dans les langues », dans M. Martins-Baltar (éd.), *La locution entre langue et usages*, Fontenay/Saint-Cloud, ENS Éditions, 277-290.
- HILPERT M. ([2014] 2019²), *Construction Grammar and its Application to English*, Edinburgh, Edinburgh University Press.

Idiomatité, constructions et variations : approche contrastive sur corpus

- KAY P. (2013), "The limits of construction grammar", in T. Hoffmann & G. Trousdale (eds.), *The Oxford Handbook of Construction Grammar*, Oxford, Oxford University Press, 32-48.
- LAKOFF G. (1987), *Women, Fire, and Dangerous Things: What Categories Reveal about the Mind*, Chicago, The University of Chicago Press.
- LAKOFF G. & JOHNSON M. (1980), *Metaphors We Live By*, Chicago, The University of Chicago Press.
- LANGACKER R. W. (2008), *Cognitive Grammar: A Basic Introduction*, Oxford, Oxford University Press.
- LEGALLOIS D. & PATARD A. (2017), « Les constructions comme unités de la langue : illustrations, évaluation, critique », *Langue française* 194, 5-14.
- LEROY S. (2003), « Les constructions comparatives intensives dans *Illusions perdues* », *L'Information grammaticale* 99, 39-42.
- LEROY S. (2004), « *Sale comme un peigne et méchant comme une teigne*. Quelques remarques sur les comparaisons à parangon », *Travaux linguistiques du CerLiCO* 17, 255-267.
- LEROY S. (2007), « Les comparaisons *comme SN* exprimant le plus haut degré », *Travaux de linguistique* 54, 69-82.
- MEJRI S. (1994), « Séquences figées et expression de l'intensité : essai de description sémantique », *Cahiers de lexicologie* 65, 111-122.
- MICHAELIS L. A. (2004), "Type shifting in construction grammar: An integrated approach to aspectual coercion", *Cognitive Linguistics* 15 (1), 1-67.
- MILNER J.-C. (1973),
- MOON R. (1998), *Fixed Expressions and Idioms in English: A Corpus-Based Approach*, Oxford, Oxford University Press.
- PIIRAINEN E. (2012), *Widespread Idioms in Europe and Beyond: Toward a Lexicon of Common Figurative Units*, New York, Peter Lang.
- QUIRK R. et alii (1985), *A Comprehensive Grammar of the English Language*, London, Longman.
- REY A. & CHANTREAU S. (1989), *Dictionnaire des expressions et locutions figurées*, Paris, Dictionnaires Le Robert.
- RIVARA R. (1990), *Le système de la comparaison : sur la construction du sens dans les langues naturelles*, Paris, Les Éditions de Minuit.
- RÖHRICH L. ([1973] 1991), *Lexikon der sprichwörtlichen Redensarten*, Freiburg, Herder.
- SCHAPIRA C. (2000), « Du prototype au stéréotype, et inversement : le cliché *comme + SN* », *Cahiers de lexicologie* 76, 27-40.
- SEIDL J. & McMORDIE W. (1988⁵), *English Idioms*, Oxford, Oxford University Press.
- SPEARS R. A. (2007⁷), *McGraw-Hill's American Idioms Dictionary: Contemporary American English Idioms and Their Meanings*, New York, McGraw-Hill.
- SZENDE T. (1999), « À propos des séquences intensives stéréotypées : plaidoyer pour une description lexicographique », *Cahiers de lexicologie* 74, 61-77.
- TOURNIER J. (2004), *Précis de lexicologie anglaise*, Paris, Ellipses.

ABSTRACTS

Yvon Keromnes, *Idiomatcity, Constructions and Variations – A Corpus-Based Contrastive Approach*

Being part of both the lexicon and syntax, idioms constitute a privileged object on which phraseology and construction grammar converge. But the latter tends to emphasize the specificity not only of each construction, but of each language, which would seem to make comparisons between languages pointless. In this article, we propose a contrastive constructionist approach to the structure “ADJ as SN” (*cute as a button, free as a bird*, etc.) and related structures in French and German. After considering the phraseographic treatment of these expressions, we take a look at them in corpora. These observations seem to show the same mechanism with an undeniable productivity, the systematic nature of which is ignored by phraseography.

Keywords : construction grammar, corpora, English-French-German comparison, phraseology

RÉSUMÉS

Yvon Keromnes, *Idiomatcité, constructions et variations : approche contrastive sur corpus*

Participant du lexique et de la syntaxe, les expressions idiomatiques constituent un point de rencontre privilégié entre la phraséologie et la grammaire de construction. Cette dernière, en revanche, tend à souligner la spécificité non seulement de chaque construction mais de chaque langue, ce qui semblerait rendre vaines les comparaisons interlangues. Dans cet article, nous proposons une approche constructionniste contrastive de la structure « ADJ *comme* SN » (*malin comme un singe, bête comme ses pieds*, etc.) et structures apparentées en allemand et en anglais. Après avoir considéré le traitement phraséographique de ces expressions, nous les observons sur corpus. Ces observations semblent montrer un même mécanisme présentant une indéniable productivité et dont le caractère systématique est ignoré par la phraséographie.

Mots-clés : comparaison allemand-anglais-français, corpus, phraséologie, grammaire de construction