



HAL
open science

CHAÎNES DE RÉFÉRENCE DANS LES TEXTES ENCYCLOPÉDIQUES : ÉTUDE CONTRASTIVE EN FRANÇAIS ET EN CHINOIS

Chang Guo

► **To cite this version:**

Chang Guo. CHAÎNES DE RÉFÉRENCE DANS LES TEXTES ENCYCLOPÉDIQUES : ÉTUDE CONTRASTIVE EN FRANÇAIS ET EN CHINOIS. Travaux de Linguistique : Revue Internationale de Linguistique Française, 2021. hal-04417070

HAL Id: hal-04417070

<https://hal.science/hal-04417070>

Submitted on 25 Jan 2024

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

CHAÎNES DE RÉFÉRENCE DANS LES TEXTES ENCYCLOPÉDIQUES : ÉTUDE CONTRASTIVE EN FRANÇAIS ET EN CHINOIS

Chang Guo*

Introduction

Une *chaîne de référence* (désormais CR) est caractérisée par une suite de trois ou plus de trois expressions coréférentielles, appelées *maillons* (cf. notamment Chastain, 1975 ; Corblin, 1985 ; Charolles, 1988 et Schnedecker, 1997, 2021). Pour le chinois, à notre connaissance, seul Xu (2010 : 333) propose de définir la coréférence comme « deux signes linguistiques renvoyant à une même personne ou un même objet »¹ (notre traduction), définition qui coïncide avec celle proposée pour d'autres langues comme l'anglais et le français. Par exemple, les maillons des CR *la girafe* et 骆驼 ('chameau') sont balisés entre crochets dans les extraits [1] et [2]².

[1] *[La girafe]_{M1} est chassée pour [sa]_{M2} viande et [sa]_{M3} peau. [Elle]_{M4} fut tuée aussi pour [sa]_{M5} queue, utilisée comme monnaie primitive ou symbole d'autorité.*

[2] *[骆驼]_{M1} 的 役用 性能 是 [其]_{M2}
[chameau]_{M1} AUX labour fonction être [son]_{M2}
主要的 生产 性能 之一 [它]_{M3} 可
principal production fonction un-des [il]_{M3} pouvoir
用 作 骑乘、 [骆驼]_{M4}
utiliser comme équitation [chameau]_{M4}
运、 拉 车、 犁地 等。
transporter tirer voiture charrette etc.*

* LILPA, Université de Strasbourg, France et Département de français, Beijing Language and Culture University, Chine ; Chang.Guo@etu.unistra.fr. Ce travail a bénéficié du soutien de l'ANR dans le cadre du projet Democrat (ANR-15-CE38-0008).

‘Être laborieux est une des principales caractéristiques des chameaux. Ceux-ci peuvent être utilisés pour l’équitation, le transport, la traction de voitures, de charrettes, etc.’

Les *textes encyclopédiques* (désormais TE) sont définis comme un type de texte informatif, dont l’objectif est de « classer les connaissances humaines » (Pinault-Sorensen, 1993) et de « diffuser un état de connaissances scientifiques » (Oberlé *et al.*, 2018 : 124). Il s’agit d’un genre de texte « relevant du genre documentaire visant à transmettre et construire des savoirs, tournés vers ‘l’extérieur’ des communautés scientifiques qui les produisent » (Beacco, 2000 : 17). À notre connaissance, peu d’auteurs mentionnent les caractéristiques des CR dans les TE en français et peu d’études comparent les CR *a fortiori* dans les TE du point de vue contrastif (cf. Schnedecker, 2014 ; Oberlé *et al.*, 2018 et Kim, à paraître).

D’un côté, la composition des CR dans les TE attirent notre attention pour leur structure textuelle différente d’autres genres discursifs. Les TE comportent des *séquences de description et d’information* plutôt que des séquences de *narration* (Oberlé *et al.*, 2018 : 125). Par exemple, pour ce qui a trait à la composition, les CR dans les textes scientifiques présentent « plus de disparité » et sont « de taille plus petite » par rapport aux textes de fiction (Schnedecker, 2021 : 164). Par ailleurs, les TE se distinguent d’autres genres informatifs. Par exemple, Schnedecker (2014) signale que les CR dans l’introduction des TE sont plus longues que celles rencontrées dans les *recettes de cuisine* et que ces premières CR ne se succèdent pas les unes aux autres.

D’un autre côté, les caractéristiques typologiques respectives du français et du chinois sont susceptibles d’amener à des variations dans la composition et dans des modes de cohabitation des CR. Ce sont ces différences qui nous conduisent aux questions suivantes. Quelles variations observe-t-on entre les CR du français et du chinois dans les TE ? Quelles en sont les conséquences pour la cohésion textuelle ?

Dans cet article, nous cherchons donc à observer les CR dans les TE par le biais des modes construits par Schnedecker (1997, 2014, 2021), selon divers critères morpho-syntaxiques et sémantiques. Nous aborderons les caractéristiques typologiques du français et du chinois en relation avec les caractéristiques discursives des TE. Nous poursuivrons par les questions méthodologiques en précisant le choix des textes dans le corpus et le traitement des *expressions référentielles* (désormais ER) et des CR dans les textes choisis. Nous terminerons par la présentation des analyses contrastives et des résultats sur les variations de la composition et de la cohabitation des CR en français et en chinois.

1. Chaînes de référence et genre encyclopédique : caractéristiques

1.1. Caractéristiques typologiques du français et du chinois et leurs conséquences sur la composition des chaînes de référence

Le français et le chinois³ appartiennent à deux familles linguistiques très différentes l'une de l'autre. Dans la plupart des classifications, la première fait partie de la famille des langues romanes, tandis que la seconde constitue à elle seule une famille linguistique sans parenté avec d'autres langues (Dubois *et al.*, 2012 : 199). Cette différence typologique rejaille notamment sur la catégorie grammaticale des ER, à savoir notamment les *groupes nominaux* (désormais GN) et les pronoms, qui composent les CR. Par exemple, à la différence du français, le chinois n'a pas d'article défini. Les référents génériques et définis sont désignés majoritairement par les GN sans déterminant (cf. Liu, 2002 ; Chen, 2015, entre autres), comme l'illustrent les expressions entre crochets dans les exemples [3] et [4].

[3] *Si [l'Humanité] permettait [aux baleines] de revenir à leur population avant [la chasse à [la baleine]], cette population serait capable de séquestrer environ 1,7 milliards de CO₂ par an. Si [les forêts françaises] séquestrent 70 millions de CO₂ par an, alors [la séquestration [du carbone] par cette population en 1 an] équivaut à environ 25 ans de séquestration [du carbone] par [les forêts françaises].*

[4] [鲸类] 拥有 [敏锐的 声呐 系统],
 [cétacé] avoir [intelligent sonar système]
 能 利用 [回声定位 功能], 发出
 pouvoir utiliser [écholocation fonction] émettre
 声波 信号 觅食 和 通讯,
 acoustique-onde signal se-nourrir et communiquer
 [[人类] 发明的 声呐] 就 是 通过
 [[humain] inventé sonar] donc être par
 [[鲸] 和 [海豚] 的 声呐 原理]
 [[baleine] et [dauphin] AUX sonar principe]
 发明的。 [鲸类] 依靠 敏锐的 听觉 在
 inventé [cétacé] compter-sur aigu ouïe à
 水下 生存, 但 [须鲸] 和 [齿鲸]
 eau-sous vivre mais [baleine à fanons] et [baleine à dents]
 在 [听觉能力] 上 却 展现出
 à [auditif capacité] sur mais montrer
 巨大的 差别。
 énorme différence

‘Les cétacés ont un système de sonar intelligent. Ils utilisent l'écholocation qui consiste à émettre des signaux acoustiques, afin de trouver de la nourriture et de communiquer. Le sonar inventé par les humains est inspiré par celui des baleines et des dauphins. Les cétacés comptent sur leur ouïe aigüe pour vivre dans l'eau, mais les baleines à fanons et les baleines à dents montrent des différences énormes en ce qui concerne leur capacité auditive.’

Une autre différence concerne la proportion d'emploi des pronoms zéro anaphoriques, beaucoup plus fréquents en chinois qu'en anglais (cf. Xu, 1990 ; Huang, 2000, entre autres) et en français, comme le révèle l'extrait [5], où aucune forme lexicale ne renvoie au référent *autruche*, thème du texte.

- [5] $[\emptyset]_{M1}$ 生活 于 非洲的 沙漠 草地 和 稀树草原
 $[\emptyset]_{M1}$ vivre à africain désert prairie et savane
 地带, 因 其 羽、皮 及 肉
 zone car son plume peau et viande
 等 都 有 很 高 的 经 济 价 值, $[\emptyset]_{M2}$
 etc. tous avoir haut économique valeur $[\emptyset]_{M2}$
 有 生 长 快、繁 殖 力 强、易
 avoir grandir rapide fécondité fort aisé
 饲 养 和 抗 病 力 强 等 优 点, $[\emptyset]_{M3}$
 s'élever et résistance-maladie fort etc. avantage $[\emptyset]_{M3}$
 在 许 多 国 家 被 广 泛 驯 养。
 à beaucoup pays par large élever

‘(Les autruches) Vivent dans les prairies désertiques et les savanes en Afrique. Leurs plumes, leur peau et leur viande se vendent bien sur le marché. En raison d'une croissance rapide, d'une forte fécondité, d'une reproduction aisée et d'une forte résistance aux maladies, (elles) sont élevées dans beaucoup de pays.’

La troisième caractéristique morpho-syntaxique des ER en chinois concerne les modifieurs des GN. Dans *l'échelle de l'accessibilité référentielle* (Ariel, 1990 : 73), le modifieur constitue un facteur qui signale un degré d'accessibilité moindre que les GN qui en sont dotés. Or, la place du modifieur est variable dans chaque langue. Ceux d'un GN en français peuvent être à la fois postposés ou antéposés au nom, comme dans *les forêts françaises* ou *une forte fécondité*. En revanche, les modifieurs des GN en chinois sont toujours antéposés au nom et cet ordre est invariable, comme dans les exemples de l'énoncé [4] : 人类发明的-声呐 *humain-inventé-sonar*, ‘le sonar inventé par les humains’, 敏锐的-听觉 *sensible-audition*, ‘l'audition sensible’ et 巨大的-差别 *énormes-différences*, ‘des différences énormes’.

Concernant l'aspect intra-phrastique, le français et le chinois sont souvent analysés sur deux plans : au niveau syntaxique et au niveau pragmatique. Dans la définition typologique de Li et Thompson (1989), le français est une langue dominée par le sujet syntaxique, tandis que le chinois est dominé par le topique pragmatique. En français, la variation casuelle du pronom sert à distinguer le sujet, le complément d'objet et d'autres fonctions syntaxiques, comme *je*, *me* et *moi*. La position avant ou après le verbe marque également la fonction. En chinois, langue dépourvue de marquage casuel, les fonctions syntaxiques sont moins identifiables. En effet, le verbe en chinois se présente souvent avec la même forme lexicale que celle du nom, comme 通讯 pour 'communication' et 'communiquer' et 发明 pour 'invention' et 'inventer', raison pour laquelle il est souvent difficile de le repérer dans la phrase. En outre, le verbe n'est pas un élément obligatoire d'une phrase en chinois. Comme l'illustre l'énoncé [6], une phrase peut comporter zéro verbe, un seul verbe ou plusieurs verbes. C'est, entre autres, ce qui explique que de nombreux auteurs utilisent les notions pragmatiques de *topique* (ce dont on parle) et *commentaire* (ce que l'on dit par rapport au topique), pour analyser les composants dans le cadre phrastique en chinois (cf. Li et Thompson, 1989 ; Xu, 2010 ; Fang, 2018, entre autres).

- [6] 躯体 粗 圆, 四肢 短, 脚 有
 corps gros rond membre court pied avoir
- 4 趾, 头 硕大, 眼、耳 较 小, 嘴
 quatre orteil tête gros yeux oreilles plutôt petit bouche
- 特别 大, 尾 较 小, 下
 particulièrement grand queue plutôt petit inférieur
- 犬齿 巨大, 长 50-60厘米,
 canine énorme **longueur/mesurer** 50-60 cm
- 重 2.5 千克, 皮 较 厚, 约
poids/peser 2,5 kg peau plutôt épais environ
 40-50毫米。
 40-50 mm

'Le corps (des hippopotames) (est) massif et de forme ronde. Les membres (sont) courts. Les pieds comportent quatre orteils. La tête (est) grande. Les yeux et les oreilles (sont) petits. La bouche (est) particulièrement grande. La queue (est) petite. Les canines inférieures (sont) énormes. (Elles) Peuvent mesurer 50-60 cm et peser 2,5 kg. La peau (est) épaisse d'environ 40-50 mm.'

Les caractéristiques morpho-syntaxiques des ER des deux langues présentées dans cette section sont synthétisées dans le Tableau 1 suivant.

Tableau 1 : Récapitulatif des caractéristiques typologiques du français et du chinois

Langue	Français	Chinois
Référent générique / défini	GN défini	GN sans déterminant
Pronom zéro anaphorique	Moindre fréquence	Forte fréquence
Modifieur de GN	Antéposé ou postposé	Antéposé
Composants phrastiques	Sujet, verbe, complément d'objet	Topique, commentaire

1.2. Caractéristiques discursives des textes encyclopédiques et leurs conséquences pour la composition et la cohabitation de chaînes de référence

À la différence des textes narratifs, les TE sont rédigés de façon impersonnelle avec une intention d'objectivité. Ils contiennent peu d'expressions personnelles, sentimentales ou subjectives comportant des commentaires tels que *quelqu'un trouve que, selon l'auteur*, par exemple. Le rédacteur tente de tenir une position neutre et anonyme, visant à partager des connaissances scientifiques avec un public ouvert, général et non-expert. La progression de l'information ne dépend pas de connecteurs logiques ou spatio-temporels (*d'abord, ensuite, au début, à l'intérieur de*, par exemple). Il y en a proportionnellement moins dans les TE que dans les textes narratifs.

Pour ce qui a trait à la structure, un titre général, placé au début de chaque TE, en indique le thème (*aigle, autruche*, par exemple) et plusieurs sous-titres divisent le texte en sous-thèmes (*étymologie, description, répartition géographique, écologie et comportement*, etc.) dérivant du thème général. Le titre et les sous-titres sont souvent composés d'un nom ou d'un GN sans déterminant et sans modifieur.

À part le titre et les sous-titres, les TE « sont découpés en paragraphes dédiés à des thèmes récurrents, ordonnés de manière assez libre » (Schneidecker, 2014 : 33). Les paragraphes sont souvent nombreux et brefs, constituant des « plans d'organisation textuelle » (Charolles, 1988 : 3). Un paragraphe peut contenir un seul ou plusieurs *référents saillants* (référents qui portent un haut degré d'accessibilité dans l'échelle référentielle d'Ariel, voir aussi Landragin, 2004). Les paragraphes qui sont dominés par un seul référent saillant sont appelés *environnement mono-référentiel*, comme c'est le cas de la CR *les aigles* dans le fragment [7]. D'autres référents se trouvent également dans le paragraphe (*pattes, proies, vue*, par exemple), mais ils ne sont pas des concurrents référentiels du référent dominant *les aigles*.

[7] *[Les aigles]_R sont de grands rapaces planeurs diurnes [qui] possèdent des pattes puissantes et de grandes serres qui*

[leur]_R permettent de saisir [leurs]_R proies. [Ils]_R ont une vue perçante [leur]_R permettant de repérer celles-ci à distance.

Quand un paragraphe comporte plusieurs référents saillants, environnements dits pluri-référentiels, la question se pose de savoir comment cohabitent les CR. Cinq modes de cohabitation sont distingués par Schnedecker (1997, 2021) et Schnedecker et Landragin (2014) : la *succession*, l'*entrecroisement*, la *dérivation*, la *partition/fusion* et le *déroulement en parallèle*.

La *succession* décrit les cas où « une chaîne disparaît dès l'instant où celle qui lui succède apparaît dans le texte » (Schnedecker et Landragin, 2014 : 14). Ce mode de cohabitation suppose que le référent dominant change à l'intérieur d'un paragraphe. Autrement dit, le référent saillant au début du paragraphe ne l'est pas à la fin du paragraphe. Ce mode concernerait quelques cas rares dépendant de la thématique des TE :

- [8] [...]而[花鼠]_{R1}也是[中国温带地区常见的松鼠]_{R1}。
 et [tamia]_{R1} aussi être [Chine tempéré région souvent écureuil]_{R1}
 [这种背部长有
 [ce type dos avoir
 黑白相间纵纹的小动物]_{R1}会花更多的
 noir-blanc rayure AUX petit animal]_{R1} aller passer plus
 时间在地面觅食,并[Ø]_{R1}
 temps à terre se-nourrir et [Ø]_{R1}
 常捕食小动物,[Ø]_{R1}还嗜啃食
 souvent attaquer petit proie [Ø]_{R1} aussi vorace ronger
 树皮。[赤腹松鼠]_{R2}甚为
 écorce [rouge ventre écureuil]_{R2} encore
 胆大,且[Ø]_{R2}可适应受到干扰的林缘
 téméraire et [Ø]_{R2} pouvoir s'adapter dérégulé lisière
 生境,因此[Ø]_{R2}也能很好地
 habitat donc [Ø]_{R2} aussi pouvoir bien
 生活在城市公园绿地中。[...]
 vivre à urbain parc pelouse dans

'[...] Les tamias sont également des écureuils communs dans les régions tempérées de Chine. Ces animaux de petite taille aux rayures noires et blanches passent beaucoup de temps au sol pour se nourrir et s'attaquent souvent à de petites proies. Ils sont également friands d'écorces. Les écureuils à ventre roux, pour leur part, sont très téméraires et peuvent s'adapter aux habitats déréglés de la lisière des forêts, de sorte qu'ils peuvent également bien vivre dans les parcs urbains. [...]'

L'entrecroisement des CR concerne les cas où deux ou plusieurs référents sont « instanciés, consécutivement et/ou simultanément, avec des modifications de leur statut syntaxique » (Schneidecker et Landragin, 2014 : 14), comme le montrent les CR référant aux *baleines* et *carbone* dans le fragment [9]. Le GN *les baleines* passe de la fonction syntaxique de sujet (*les baleines, elles*), puis ce référent est réinstancié par le biais d'un déterminant possessif dans un GN évoquant un autre référent (*leur corps, leur mort*) ; le GN *le carbone* apparaît d'abord en position de complément d'objet (*stocker le carbone*) et ensuite en position sujet (*le carbone libéré ...*) à la fin du paragraphe. Les deux référents sont aussi accessibles l'un que l'autre dans le paragraphe, mais ils ont des degrés d'accessibilité différents dans chaque phrase.

- [9] *[Les baleines]_{R1} sont considérées comme des puits de [carbone naturels]_{R2'} [elles]_{R1} stockent [le carbone]_{R2} dans [leur]_{R1} corps riche en graisse et en protéine. De plus, après [leur]_{R1} mort, [les baleines]_{R1} emmènent avec [elles]_{R1} [le stock de carbone qu']_{R1} [elles]_{R1} ont séquestré]_{R2'}. [Le carbone libéré par la décomposition des squelettes]_{R2} en font un environnement propice au développement d'une biodiversité spécifique (hot spot de biodiversité).*

La dérivation des CR est illustrée par les CR *les rhinocéros* et *leurs cornes* dans le fragment [10], où « l'introduction d'un nouveau référent dans un texte n'opère pas par le biais d'un GN indéfini ou d'un nom propre, mais peut dériver d'une autre CR par des moyens anaphoriques » (Schneidecker et Landragin, 2014 : 14).

- [10] *Dans certaines cultures de l'Asie orientale, [les rhinocéros]_{R1} sont tués uniquement pour [[leurs]_{R2} cornes]_{R1} car [elles]_{R2} sont utilisées pour faire des sculptures, des coupes libatoires notamment. Les supposés effets thérapeutiques et aphrodisiaques attribués à [la corne broyée]_{R2} et la mode des poignards en [[corne]_{R2} de rhinocéros]_{R1} dans les classes supérieures du Yémen, comme marque de standing et symbole de virilité, ont favorisé [leur]_{R2} trafic sur le marché noir et le braconnage d'espèces pourtant en voie de disparition.*

La *partition* et la *fusion* constituent deux procédures symétriques : la partition extrait des référents individués d'un ensemble, tandis que la fusion rassemble les référents isolés dans un même ensemble. L'extrait [11] illustre une partition de la CR *les singes du nouveau monde* aux CR *sapajous* et *sagouins*. Le référent saillant se différencie ainsi du premier référent *singes du nouveau monde* à leurs deux sous-groupes.

- [11] *[Les Singes du Nouveau Monde]_{R1} se séparent quant à [eux]_{R1} entre [« sapajous »]_{R2} et [« sagouins »]_{R3'}. [Les premiers]_{R2} « ont*

la queue prenante » et $[\emptyset]_{R2}$ regroupent l'Ouarine (et l'Alouate), le Coaïta, le Sajou, le Saï et le Saïmiri. $[\text{Les seconds}]_{R3}$ l'ont « entièrement velue, lâche et droite » et $[\emptyset]_{R3}$ comprennent le Saki, le Tamarin, le Ouistiti, le Marikina, le Pinche et le Mico.

Le déroulement en parallèle des CR correspond « aux cas de topiques multiples où deux référents aussi saillants l'un que l'autre sont repris en parallèle » (Schneidecker et Landragin, 2014 : 14), comme l'illustrent les CR *sapajous* et *sajouins* dans [11] et les CR *les hippopotames* et *les baleines* dans [12]. Le paragraphe comporte deux référents aussi saillants l'un que l'autre, constituant deux topiques paragraphiques parallèles.

- [12] *[Les hippopotames]_{R1} ont plus de points communs avec [les baleines]_{R2} qu'[ils]_{R1} n'en ont avec d'autres ongulés comme les cochons. Ainsi, l'ancêtre commun [des hippopotames]_{R1} et [des baleines]_{R2} a existé après que la lignée s'est séparée des ruminants, séparation qui s'est elle-même produite après la divergence du reste des ongulés artiodactyles, cochons compris. Alors que [la baleine]_{R1} et [l'hippopotame]_{R2} sont les plus proches parents vivants [l'un]_{R1} de [l'autre]_{R2}.*

Pour conclure, les CR dans les TE se distinguent à la fois de celles des textes narratifs et de celles d'autres textes informatifs. Le découpage en paragraphes et la présence des sous-titres proposent des environnements discursifs mono-référentiels et pluri-référentiels, permettant d'observer les influences du nombre de référents saillants sur la composition des CR et sur la cohabitation des CR.

2. Méthodologie

2.1. Sélection des textes du corpus

Nous avons constitué un *corpus comparable échantillonné* (Teubert et Čermáková, 2004), correspondant à « un ensemble de textes dans deux langues qui ne sont pas en rapport de traduction mutuelle mais qui traitent du même domaine étroit et qui relèvent du même registre linguistique » (Habert, 2005 : 91). Les encyclopédies ouvertes en ligne, comme *Wikipédia*⁴ et *Baidubaïke*⁵, proposent des TE en diverses langues.

Les TE traitant du thème des animaux ont été choisis pour deux raisons. En premier lieu, ces textes portent sur des thèmes faciles à comprendre. Cela évite des erreurs sur l'identification du référent que pourrait provoquer la terminologie de certains autres domaines, comme la chimie, le droit, etc. En second lieu, les TE animaliers sont assez nombreux, permettant ainsi de constituer un corpus de textes de longueur similaire dans les langues qui nous concernent. Après avoir passé en revue plus d'une

centaine d'articles consacrés aux animaux, nous avons finalement retenu quarante articles d'une longueur comparable portant sur vingt animaux identiques dans les deux langues. L'ensemble des quarante articles représente en tout 245 000 signes en français et 25 3000 signes en chinois, correspondant à 36 000 mots en français et 57 000 mots en chinois⁶, comme le synthétise le Tableau 2 *infra*. En tout, 6 639 maillons ont été repérés dans l'ensemble du corpus, 3 892 maillons en chinois et 2 737 maillons en français. Ils renvoient à 722 référents, correspondant donc au même nombre de CR.

Les CR sont classées en *CR principales* (i.e. CR du thème de texte) et *CR secondaires* (i.e. toutes les CR autres que les CR principales) (cf. Oberlé *et al.*, 2018) en fonction de leur accessibilité discursive dans chacun des textes. Les CR dites *principales* renvoient au référent du titre de l'article. Les CR dites *secondaires* renvoient quant à elles aux référents autres que le thème du texte. Par exemple, dans l'article sur la girafe, la CR principale correspond au thème *girafe*. Les autres CR, comme *son long cou, sa nourriture et le zoo*, constituent des CR secondaires. Les vingt CR principales correspondent aux vingt thèmes d'article. Les CR secondaires sont plus nombreuses, 306 en français et 376 en chinois.

Tableau 2 : Présentation du corpus

Langue		Français	Chinois	Total
Nombre de textes		20	20	40
Nombre de mots	Total	36 102	56 964	93 066
	Moyenne par texte	1 805,1	2 848,2	-
Nombre de paragraphes	Total	650	591	1 241
	Moyenne par texte	32,5	29,6	-
Nombre de maillons	Total	2 737	3 892	6 639
	Dans les CR principales	1 194	1 446	2 640
	Dans les CR secondaires	1 543	2 450	3 993
Nombre de CR	Total	326	396	722
	CR principales	20	20	40
	CR secondaires	306	376	682

2.2. Traitement et annotation du corpus

Une fois les articles choisis, nous avons délimité les expressions coréférentielles à l'aide de l'outil d'annotation SACR⁷, comme l'illustre la Figure 1. Les maillons sont d'abord nommés par le nom du référent, comme *autruche, élevage, plumes*, etc. Ils sont ensuite annotés pour trois caractéristiques morpho-syntaxiques et sémantiques : la catégorie morpho-syntaxique, la fonction syntaxique et la caractéristique de référent.

Figure 1 : Capture d'écran montrant les modalités d'annotation de l'outil SACR

[#25] **AUTRUCHE** Les autruches sont des oiseaux robustes, d'**Eleveage** élevage facile , domestiqués depuis longtemps pour
Plumes **AUTRUCHE** leurs plumes , **Viande** **AUTRUCHE** leur viande , **Cuir** **AUTRUCHE** leur cuir et
Oeuf **AUTRUCHE** leurs œufs .

Concernant les caractéristiques morpho-syntaxiques, les ER sont annotées selon cinq types : les expressions nominales se divisent en *GN+modifieur* et *GN-modifieur*, et les expressions non nominales sont annotées comme *pronom*, *déterminant possessif* et *pronom zéro*, comme l'indiquent les annotations figurant entre crochets dans les énoncés [13]-[22].

- [13] *[La girafe]_{GN-modifieur} vit en [Afrique]_{GN-modifieur}, dans [la savane]_{GN-modifieur}.*
- [14] *[长颈鹿]_{GN-modifieur} 是 [非洲]_{GN-modifieur} 特有的动物。
[girafe]_{GN-modifieur} être [Afrique]_{GN-modifieur} particulier animal
'La girafe est un animal qui vit uniquement en Afrique.'*
- [15] *[Son galop particulier]_{GN+modifieur} est facilité par [son long cou]_{GN+modifieur} [...].*
- [16] *它 不 能 做 [长 距离 奔跑]_{GN+modifieur}
Il NEG pouvoir faire [long distance course]_{GN+modifieur}
'Elle ne peut pas courir de longues distances.'*
- [17] *[Elle]_{Pronom} court à 15 km/h mais peut accélérer à 56 km/h en prenant un curieux galop.*
- [18] *[这]_{Pronom} 是 [...] 常用的 运动 方式。
[cela]_{Pronom} être habituel se-déplacer façon
'Cela constitue sa façon habituelle de se déplacer.'*
- [19] *[Ses]_{Déterminant} ossicônes sont massifs et durs comme de l'ivoire et sur [son]_{Déterminant} front pousse une excroissance osseuse, la corne médiane.*
- [20] *[它的]_{Déterminant} 耳后 和 眼后 还有
[son]_{Déterminant} oreille-derrrière et yeux-derrrière encore avoir
两 对 角, 但是 不 很 明显。
deux paire corne mais NEG très évident
'Derrrière ses oreilles et ses yeux poussent encore deux paires de cornes qui ne sont pas très visibles.'*
- [21] *La girafe se repose et [Ø]_{Pronom zéro} dort debout.*

- [22] 长颈鹿 是 非洲 特有的 动物, [Ø]_{Pronom zéro}
 girafe être Afrique particulier animal [Ø]_{Pronom zéro}
 与 鹿 和 牛 有 亲缘 关系。
 avec cerf et taureau avoir parenté relation

‘La girafe est un animal qui vit uniquement en Afrique, (elle) a une relation de parenté avec le cerf et le taureau.’

Ensuite, sur le plan phrastique, la fonction syntaxique est annotée pour chaque maillon. Nous avons distingué entre *sujet* et *autre*, comme l’illustrent les annotations figurant entre crochets dans [23] et [24]. Les sujets des deux langues partagent certains points communs. Ainsi, le sujet apparaît sous la forme d’un GN, d’un pronom ou d’un pronom zéro en début de phrase ; il porte la fonction référentielle ; il est souvent antéposé au verbe. Néanmoins, le français et le chinois présentent aussi des différences quant au marquage du sujet. Par exemple, le verbe en français s’accorde au sujet, tandis que celui en chinois ne s’accorde pas.

- [23] [*Elle*]_{Sujet} saisit donc [*les pousses d’acacias*]_{Autres} avec
 [*sa langue*]_{Autres}.

- [24] [*长颈鹿*]_{Sujet} [...] 卷食 [*隐藏在里层的树叶*]_{Autres}^o
 [*girafe*]_{Sujet} manger [*caché à intérieur feuilles*]_{Autres}

‘La girafe mange les jeunes feuilles situées à proximité du tronc de l’arbre.’

Enfin, le troisième facteur annoté concerne les caractéristiques référentielles du nominal. De nombreux auteurs proposent des classifications des ER, comme la référence définie et indéfinie, la référence spécifique et non spécifique, la référence générique et non générique (cf. notamment Schwarz, 1979 ; Kleiber, 2001 ; Charolles, 2002 ; Chen, 2015). Sachant que les TE ont pour objectif de présenter les types et les classes d’objets sous les mêmes nominations linguistiques, les TE proposent de bons contextes discursifs pour observer les ER génériques. Ainsi, deux caractéristiques référentielles sont distinguées : *réfèrent générique* et *réfèrent non générique*, comme l’illustrent les annotations entre crochets dans [25] et [26]. Les sous-types de référence non génériques sont nombreux, y compris la référence singulière, quantitative, définie, indéfinie, spécifique, etc.

- [25] En septembre 1984, [*une spectaculaire noyade de*
 [*9 604 caribous*]_{Non générique}] Non générique à
 [*la chute [du Calcaire]*]_{Non générique} Non générique sur
 [*la rivière Caniapiscou*]_{Non générique} a fait le tour [*du monde*]_{Générique}.

- [26] 如今 在 [北欧]^{Générique} 只有
 aujourd'hui à [nord Europe]^{Générique} juste avoir
 [大约2800名萨米族人]^{Non générique} 专职 以
 [environ 2 800 Samien]^{Non générique} professionnel de
 放牧 [驯鹿]^{Générique} 为生, 过着 [半游牧的生活]^{Générique°}
 élever [renne]^{Générique} pour vivre vivant [semi nomade vie]^{Générique}
 'Aujourd'hui au nord de l'Europe, on ne compte plus qu'environ
 2 800 Samiens qui sont partiellement nomades et vivent de l'éle-
 vage des rennes.'

La grille d'annotation est synthétisée dans le Tableau 3 *infra*.

Tableau 3 : Propriétés morpho-syntaxiques et référentielles des maillons annotés

Forme morpho-syntaxique	Fonction syntaxique	Caractéristique référentielle
GN+modifieur GN-modifieur Pronom Déterminant possessif Pronom zéro	Sujet Autres fonctions	Référent générique Référent non générique

2.3. Corpus construit

À partir des annotations (cf. Tableau 2, *supra*), les paramètres suivants⁸ sont calculés pour décrire les CR : (i) *le nombre moyen de CR par texte*, (ii) *le nombre moyen de maillons par texte*, (iii) *la longueur moyenne de CR*, (iv) *la longueur moyenne de la CR principale*, (v) *la longueur moyenne de la CR secondaire* et (vi) *la densité référentielle*⁹, comme le montre le Tableau 4, *infra*.

Certaines données présentent des décalages importants entre les deux langues tandis que d'autres sont plutôt similaires. La moyenne du nombre de CR par texte est de 16,3 en français et de 19,8 en chinois. Le nombre moyen de maillons par texte est de 136,9 maillons en français contre 194,6 en chinois, ce qui est beaucoup plus élevé. On observe une grande différence au niveau de la longueur moyenne de la CR principale : 59,7 maillons en français contre 72,9 en chinois. La longueur moyenne des CR, toutes CR confondues, est presque identique : 9,4 maillons en français et 10,1 en chinois. Par rapport aux CR principales, les CR secondaires sont beaucoup plus courtes et comportent un nombre de maillons similaire en moyenne, à savoir 5,0 maillons en français et 6,5 maillons en chinois.

Comme les chiffres dépendent de la longueur des textes, la densité référentielle, qui indique le nombre de maillons pour cent mots, est plus représentative. Elle atteint 6,83 % en chinois, chiffre proche du français quoiqu'un peu moins élevé que celui du français (7,58 %). Cela signifie que les passages de cent mots contiennent 7,58 maillons en français, alors qu'en chinois ils ne contiennent que 6,83 maillons.

Tableau 4 : Présentation chiffrée des chaînes de référence dans le corpus

	Français	Chinois
(i) Nombre moyen de CR par texte	16,3	19,8
(ii) Nombre moyen de maillons par texte	136,9	194,6
(iii) Nombre moyen de maillons par CR	9,4	10,1
(iv) Nombre moyen de maillons par CR principale	59,7	72,3
(v) Nombre moyen de maillons par CR secondaires	5,0	6,5
(vi) Densité référentielle	7,58 %	6,83 %

3. Résultats

3.1. Composition des chaînes de référence selon trois volets

À l'aide de l'outil d'analyse CRViewer¹⁰, trois types de paramètres ont été appliqués aux CR dans le corpus. Dans cette section la composition des CR sera présentée ainsi selon ces trois paramètres.

À l'issue de l'annotation des formes morpho-syntaxiques des maillons, divers écarts entre les deux langues sont observés à un double niveau : celui de la totalité des CR et celui des CR principales. Les pourcentages de chaque forme sont indiqués dans le Tableau 5, *infra*.

En premier lieu, dans la totalité des CR du corpus, le GN-modifieur est la forme la plus fréquente, quelle que soit la langue, totalisant 55,8 % en français et 62,3 % en chinois. En français, les pronoms et les déterminants possessifs se trouvent respectivement à la deuxième et à la troisième place, dans des proportions assez proches, les premiers à 18,1 % et les seconds à 14,1 %. Or, ces deux formes sont beaucoup moins fréquentes en chinois, qui comporte seulement 5,1 % de pronoms et 4,1 % de déterminants possessifs, qui sont les deux formes les moins utilisées en chinois. Les GN+modifieur et les pronoms zéro en français se trouvent à la quatrième et à la cinquième place. Les GN+modifieur sont moins nombreux, occupant des proportions similaires dans les deux langues, 9,3 % en français et 10,7 % en chinois. Pour les pronoms zéro, une grande variation est toutefois observée d'une langue à l'autre. Les pronoms zéro constituent la catégorie la moins fréquente en français (2,7 %), alors qu'en chinois ils

représentent une proportion beaucoup plus élevée, soit 17,8 %. Ils occupent la deuxième place en chinois, juste après celle des GN-modifieur.

Les proportions similaires des GN-modifieur et des GN+modifieur dans les deux langues suggèrent que la position de modifieur, qu'il soit antéposé ou postposé, n'influence pas la saillance référentielle des GN en français et en chinois. Les écarts observés entre les deux langues pour les pronoms, déterminants possessifs et pronoms zéro montrent que l'emploi des pronoms et des déterminants possessifs est plus fréquent en français par rapport à celui des pronoms zéro en chinois.

En second lieu, pour ce qui est des CR principales, la plupart des formes donnent des proportions similaires par rapport à celles de la totalité des CR du corpus, comme le GN-modifieur (55,9 % en français et 56,1 % en chinois), le déterminant possessif en français (13,8 %) et le pronom en chinois (6,9 %). Certaines formes présentent une proportion plus élevée ou plus basse qui est assez remarquable par rapport à celles de la totalité des CR du corpus. Par exemple, les pronoms en français augmentent de 18,1 % à 25,6 %, alors que les pronoms zéro en chinois s'élèvent encore plus fortement de 17,8 % à 32,6 %. Ces écarts attestent un usage plus fréquent des pronoms en français et des pronoms zéro en chinois pour les référents saillants. Les déterminants possessifs en chinois diminuent de 4,1 % à seulement 2,8 %. Les pronoms zéro en français baissent légèrement de 2,7 % à 1,6 %. Les GN+modifieur réduisent très fortement : seulement 3,1 % en français et 1,6 % en chinois. Ce sont les formes les moins utilisées en ce qui concerne les référents saillants.

Tableau 5 : Distribution des formes morpho-syntaxiques de maillons en français et en chinois

Langue	Français		Chinois	
	Dans le corpus	Dans la CR principale	Dans le corpus	Dans la CR principale
GN-modifieur	55,8 %	55,9 %	62,3 %	56,1 %
Pronom	18,1 %	25,6 %	5,1 %	6,9 %
Déterminant possessif	14,1 %	13,8 %	4,1 %	2,8 %
GN+modifieur	9,3 %	3,1 %	10,7 %	1,6 %
Pronom zéro	2,7 %	1,6 %	17,8 %	32,6 %

D'après les chiffres absolus (cf. Tableau 4, *supra*), les CR principales sont en moyenne plus longues en chinois (72,3 maillons) que celles du français (59,7 maillons), ce qui suggère que le thème de l'article est repris plus souvent en chinois qu'en français afin de permettre aux lecteurs d'accéder au référent thématique. Or, 32,6 % de maillons en chinois sont constitués par des pronoms zéro, tandis que seulement 1,6 % le sont en

français, ce qui nous amène à conclure que les CR principales en chinois sont plus longues mais contiennent en grande partie des maillons non exprimés.

Les proportions des fonctions syntaxiques des maillons sont enregistrées dans le Tableau 6, *infra*. En français, le pourcentage des maillons sujets s'élève à 39,0 % dans l'ensemble des CR et à 45,1 % dans les CR principales. L'augmentation de la proportion des sujets dans les CR principales montre une tendance à exprimer le référent dominant en position sujet. En chinois, les maillons sujets représentent une grande moitié des maillons dans la totalité des CR (54,5 %) et une majorité dans les CR principales (71,8 %). Ce dernier pourcentage est beaucoup plus élevé que celui de la totalité des CR, suggérant la même tendance à l'encodage du référent saillant comme sujet en chinois, tendance qui est encore plus frappante qu'en français.

Tableau 6 : Fonctions syntaxiques des maillons

Langue	Français		Chinois	
	Dans le corpus	Dans les CR principales	Dans le corpus	Dans les CR principales
Sujet	39,0 %	42,1 %	54,5 %	71,8 %
Autres fonctions	61,0 %	57,9 %	45,5 %	28,2 %

En ce qui concerne les caractéristiques sémantiques des maillons, les référents génériques sont beaucoup plus nombreux que les référents non génériques en raison du contexte encyclopédique. Les premiers offrent une proportion majoritaire : 85,2 % en français et 91,6 % en chinois, comme le révèle le Tableau 7 ci-après. La proportion majoritaire des référents génériques pourrait résulter du nombre de maillons des CR principales, qui renvoient toujours à des référents génériques. Pourtant, dans chaque langue, 14,8 % (en français) et 8,4 % (en chinois) de maillons renvoient à des référents non génériques. Ceux-ci semblent concerner dans la plupart des situations des référents définis, et plus précisément des toponymes, comme *France, Afrique, Tibet*, etc.

Tableau 7 : Caractéristiques sémantiques des maillons dans les chaînes de référence

Pourcentage	Français	Chinois
Référents génériques	85,2 %	91,6 %
Référents non génériques	14,8 %	8,4 %

3.2. Cohabitation des chaînes de référence dans deux types d'environnement textuel

Comme le montre le Tableau 8 *infra*, nous recensons 283 paragraphes en français et 382 paragraphes en chinois qui comportent soit une seule CR, à savoir les paragraphes mono-référentiels, soit plusieurs (paragraphes dits *pluri-référentiels*). Les paragraphes mono-référentiels sont majoritaires : 59,93 % en français et 56,96 % en chinois.

Les cas où plusieurs CR cohabitent présentent un grand écart quant au nombre et à la proportion de chaque mode. Les cinq modes de cohabitation évoqués *supra* (cf. Schnedecker et Landragin, 2014) sont présents dans les deux langues étudiées. Certaines différences plus ou moins significatives s'observent entre les proportions du même mode.

L'entrecroisement des CR occupe la première place dans les TE : 13,83 % de modes de cohabitation en français et 14,70 % en chinois. La dérivation occupe la deuxième place : 12,41 % en français et 14,44 % en chinois. Une proportion moins importante est trouvée pour les CR qui se déroulent en parallèle : 4,26 % de cas en français et 4,72 % en chinois. La partition et la fusion, en tant que procédures symétriques, représentent respectivement 4,96 % et 2,84 % en français et 4,72 % et 1,57 % en chinois. Le mode le moins utilisé est la succession des CR qui représente seulement 1,42 % en français et 3,67 % en chinois, ce qui montre que le référent dominant change assez rarement à l'intérieur du paragraphe.

Tableau 8 : Modes de cohabitation des chaînes de référence dans les textes encyclopédiques

Nombre (pourcentage)		Français	Chinois
(i) Paragraphe mono-référentiel		169 (59,93 %)	217 (56,96 %)
Paragraphes pluri-référentiels	(ii) Entrecroisement	39 (13,83 %)	56 (14,70 %)
	(iii) Dérivation	35 (12,41 %)	55 (14,44 %)
	(iv) Partition	14 (4,96 %)	18 (4,72 %)
	(v) Déroulement en parallèle	12 (4,26 %)	18 (4,72 %)
	(vi) Fusion	8 (2,84 %)	6 (1,57 %)
	(vii) Succession	4 (1,42 %)	14 (3,67 %)
Total		282	381

Le recensement des CR dans les paragraphes mono-référentiels révèle trois patrons de CR qui semblent représentatifs, comme le montre le Tableau 9 *infra*. Dans le premier, les maillons nominaux et non nominaux s'entrecroisent à hauteur de 76,92 % des cas en français et 74,28 % en chinois. Le deuxième patron ne contient que des maillons nominaux, avec ou sans modifieur, ce qui correspond à 14,20 % des CR en français et 20,21 % en chinois. Le troisième modèle est moins utilisé par rapport

aux deux premiers. Ce sont des CR initiées par un des trois marqueurs non nominaux d'accessibilité forte : pronom, déterminant possessif et pronom zéro, représentant 8,88 % des CR en français et 5,51 % en chinois.

Tableau 9 : Trois patrons de chaînes de référence dans les textes encyclopédiques en français et en chinois

Modèle	Français	Chinois
(i) GN-modifieur – pronom / déterminant possessif / pronom zéro ... (GN-modifieur)	76,92 %	74,28 %
(ii) GN (+/-modifieur) – GN (-modifieur) – GN (-modifieur) ...	14,20 %	20,21 %
(iii) Pronom / déterminant possessif / pronom zéro – Pronom / déterminant possessif / pronom zéro ...	8,88 %	5,51 %

Dans le premier patron, chaque GN (la plupart sans modifieur) est repris par un à quatre marqueurs non nominaux, à savoir les pronoms, les déterminants possessifs ou les pronoms zéro. Ce patron correspond à tous les types de CR, principales comme secondaires. Il apparaît alors une seule fois dans les paragraphes courts et peut être repris plusieurs fois dans les paragraphes longs, comme le révèlent les énoncés [27] et [28]. Dans des cas rares, la CR est terminée par un maillon nominal.

[27] [*Les autruches*]_{M1} sont essentiellement herbivores, mais [*leur*]_{M2} régime est varié ; [*elles*]_{M3} sont capables d'ingérer tout ce qui passe à portée de [*leur*]_{M4} bec.

[28] [*吉娃娃犬*]_{M1} 不 适合 在 户外 饲养着, 因为 [*chihuahua*]_{M1} NEG s'adapter à plein-air élevant car 太 热 或是 太 冷的 环境 trop chaud ou trop froid environnement 都 容易 让 [*它们*]_{M2} 生病。作为 主人 最好 tous facile rendre [*ils*]_{M2} malade comme éleveur préférable 在 天气 较好的 时候 就 带 à temps mieux temps donc ramener [*它们*]_{M3} 出 去 玩, 散散步, 顺便 享受 [*ils*]_{M3} sortir s'amuser se-promener en-même-temps profiter 一 下 日 光 浴, 晒 晒 太 阳。 un coup soleil bain s'exposer soleil [*吉娃娃*]_{M4} 天 生 怕 冷, [*Ø*]_{M5} 受 [*chihuahua*]_{M4} naturellement avoir-peur froid [*Ø*]_{M5} subir 寒 后 易 患 肺 炎 和 froid après facile avoir pneumonie et

风湿性关节炎, 到了 冬天 的 时候
polyarthrite-rhumatoïde arrivée hiver AUX temps

一定要注意 要给 [它]_{M6} 保温。
devoir faire-attention donner [il]_{M6} garder-chaud

‘Les chihuahuas ne peuvent pas être élevés en plein air, car ils tombent facilement malades à cause de la température trop chaude ou trop froide. Il est préférable que l’éleveur les emmène jouer, se promener et profiter du soleil quand il fait beau. Naturellement, les chihuahuas ont peur du froid et sont sujets à la pneumonie et à la polyarthrite rhumatoïde. Il faut faire très attention en hiver de les garder au chaud.’

Certaines CR, principales comme secondaires, ne comportent que des maillons nominaux, avec des variations de modifieur entre les maillons. Ce patron est plus souvent utilisé en chinois qu’en français. Les maillons nominaux en chinois se révèlent plus variables, comme l’illustrent les exemples [29] et [30]. Dans la CR *l’oie*, tous les maillons sont du type GN-modifieur avec la même tête lexicale. Or, dans la CR *le chiot* (de chihuahua), aucun maillon nominal n’a la même tête lexicale : 刚进入新环境的幼犬 ‘chiots qui viennent d’entrer dans un nouvel environnement’ – 狗仔 ‘chiot’ – 幼犬 ‘jeune chien’ – 狗狗 ‘toutou’.

[29] [L’oie]_{M1} est un des esprits totems de la roue de la vie des Amérindiens. En Chine, [l’oie]_{M2} est considérée comme un principe yang qui illumine la nature. Pour les Tchéous [l’oie]_{M3} symbolise le mariage, le mari doit offrir [une oie]_{M4} lors de la signature du contrat de fiançailles.

[30] 对于 [刚 进入 新 环境 的 幼犬]_{M1},
pour [venir-de entrer nouveau environnement AUX chiot]_{M1}
饲养者 要 了解 原
propriétaire devoir savoir ancien
主人 给 [狗仔]_{M2} 的 食物 种类、份量和
propriétaire donner [chiot]_{M2} AUX nourriture type quantité et
饲喂 的 时间, [...] 在
alimenter AUX temps à
[幼犬]_{M3} 足够 适应 外界 环境
[jeune-chien]_{M3} assez s’adapter extérieur environnement
的 基础上, 慢慢的 调节 [狗狗]_{M4}
AUX base sur doucement modifier [toutou]_{M4}
的 饮食. [...]
AUX nourriture

‘Pour les chiots qui viennent d’entrer dans un nouvel environnement, le propriétaire doit se renseigner sur le type de nourriture, sur la quantité et le temps d’alimentation chez l’ancien propriétaire, et ajuster lentement l’alimentation du toutou en fonction de son adaptation.’

Le troisième patron est réservé aux CR principales. Le sujet exprimé par un pronom, un déterminant possessif ou le sujet non exprimé, est coréférentiel au titre du texte ou au sous-titre de la section. La différence entre les deux langues concerne l’usage des pronoms et des déterminants possessifs, plus fréquents en français [31], ou des pronoms zéro, plus fréquents en chinois [32].

[31] [*Son*]_{M1} galop particulier est facilité par [*son*]_{M2} long cou qui balance et crée l’équilibre, grâce à un petit muscle spécial qui [*le*]_{M3} tire en avant.

[32] [\emptyset]_{M1} 栖息于江河流速较缓的水域或湖泊。
 [\emptyset]_{M1} habiter à rivière vitesse lent eau ou lac

[\emptyset]_{M2} 为杂食性鱼类。采收和

[\emptyset]_{M2} être omnivore poisson pêche et

储藏: [\emptyset]_{M3} 常年均可捕捞, 捕

conservation [\emptyset]_{M3} toute-année toujours pouvoir pêcher pêcher

后, 除去鳞片及骨脏,

après enlever écaille et arête

洗净鲜用。

laver frais manger

‘(Les truites) Habitent dans les rivières et les lacs. C’est un poisson omnivore. Pêche et conservation : (elles) peuvent être pêchées pendant toute l’année. Après la pêche, enlevez les écailles et les arêtes et lavez-les pour les manger fraîches.’

Dans les paragraphes pluri-référentiels, les patrons des CR sont beaucoup plus complexes, surtout quand le même paragraphe contient trois CR ou plus.

Quand deux ou plusieurs CR cohabitent, les maillons en chinois ne sont pas repris par des pronoms zéro, comme l’illustre le paragraphe [33] où cohabitent deux chaînes référant respectivement à *singe* et à *cheveux*. La répétition du GN *les cheveux* est utilisée dans l’intention de maintenir la continuité de la CR et d’éviter des ambiguïtés sémantiques.

- [33] [猴子]_{R1} 梳理 [毛发]_{R2} 不是 为了 找 虱子,
 [singe]_{R1} se-toiletter [cheveux]_{R2} NEG pour chercher pou
 而是 [猴子]_{R1} 吃的 食物
 mais [singe]_{R1} mangeant nourriture
 中 缺少 电解质的 东西, [[它们]_{R1} 毛发]_{R2}
 dans manquer électrolyte chose [[leur]_{R1} cheveux]_{R2}
 产生 盐的 结晶, [它们]_{R1}
 produire sel AUX cristal [ils]_{R1}
 互相梳理 [毛发]_{R2} 是 为了 找 盐粒 吃. [...]
 se-toiletter [cheveux]_{R2} être pour chercher sel manger

‘Les singes ne se toilettent pas les cheveux pour trouver des poux, mais pour compléter les électrolytes qui manquent dans leur nourriture quotidienne. Leurs cheveux produisent des cristaux de sel. Ils se toilettent mutuellement pour trouver des grains de sel. [...]’

Les reprises en français s’avèrent plus variables dans les environnements pluri-référentiels. La reprise au moyen d’un pronom ou d’un déterminant possessif permet toujours des interprétations univoques, comme le montre le fragment [34], où les maillons des CR *la mère* et *le petit* varient plus aisément en comparaison avec l’exemple [33] en chinois.

- [34] [*La mère*]_{R1} se fait comprendre de [[*son*]_{R1} *petit*]_{R2} en [*le*]_{R2} cares-
 sant avec le bout de [*son*]_{R1} museau ou avec un langage sonore
 d’une fréquence trop élevée pour les humains. [*Elle*]_{R1} [*l’*]_{R2} incite
 ainsi à [*la*]_{R1} suivre et à [*la*]_{R1} téter, créant ainsi le lien maternel.

Conclusion et perspectives

Notre étude a abordé un contraste entre les CR en français et en chinois dans un genre spécifique, à savoir le genre encyclopédique. L’objectif de notre étude, basée sur un corpus franco-chinois comparable contenant quarante textes, était d’observer les caractéristiques de la composition et de la cohabitation des CR dans les deux langues en question et dans le genre encyclopédique.

En ce qui concerne la composition des CR, nous avons comparé les formes morpho-syntaxiques, les fonctions syntaxiques et les caractéristiques sémantiques dans l’intégralité du corpus et dans les CR principales. Les GN-modifieur sont les composants principaux des CR des deux langues. Les pronoms et les déterminants possessifs sont plus souvent utilisés que les pronoms zéro en français, tandis que le phénomène inverse est observé en chinois. Pour les deux langues, les environnements mono-référentiels l’emportent sur les situations de cohabitation des CR dans les

TE. Trois patrons morpho-syntaxiques des maillons sont construits dans les environnements mono-référentiels. Les maillons nominaux en chinois sont plus variables en raison du changement de la tête lexicale. Les pronoms zéro en chinois prennent la place des pronoms exprimés et des déterminants possessifs en français qui donnent le plus haut degré d'accessibilité référentielle.

Dans le prolongement de cette étude, deux pistes de recherches sont envisageables. Tout d'abord, l'annotation des fonctions syntaxiques des maillons s'est limitée à distinguer le sujet de toutes les autres fonctions. Il serait toutefois intéressant de distinguer d'autres fonctions syntaxiques pour développer leur influence sur la saillance référentielle. Par ailleurs, le genre encyclopédique ne comporte pas seulement des articles animaliers. Il conviendrait donc d'explorer des textes développant des thèmes différents pour vérifier les résultats obtenus sur la composition des CR et les modes de cohabitation.

NOTES

1. 同指 (co-référence) 回指指的是, 两个语言成分是指同一个人或者物。
2. Les exemples de cet article proviennent de notre corpus, textes encyclopédiques issus de *Wikipédia* et *Baidubaiké*, collectés en 2020. Les exemples en chinois sont accompagnés des gloses et de notre traduction.
3. Notre étude porte sur le français proprement dit et le chinois mandarin, langues officielles de chacun des pays. Les variations dialectales ne sont pas discutées dans cet article.
4. Wikipédia (https://fr.wikipedia.org/wiki/Wikip%C3%A9dia:Accueil_principal, dernière consultation : janvier 2021) est un projet d'encyclopédie « collective en ligne, universelle et multilingue ». En mars 2018, la version française de Wikipédia comptait presque deux millions d'articles.
5. Baidubaiké (百度百科, <https://baike.baidu.com/>, dernière consultation : janvier 2021) est reconnu comme la plus grande encyclopédie collective chinoise en ligne. Le projet a officiellement démarré en 2008. En octobre 2020, Baidubaiké comptait plus de 21 millions d'articles.
6. Le nombre de mots en chinois est calculé par AiHanYu selon les algorithmes des règles lexico-morphologiques (<http://www.aihanyu.org/cncorpus/index.aspx>, dernière consultation : juillet 2021).
7. SACR (Script d'Annotation des Chaînes de Référence) : <http://boberle.com/projects/sacr/index.html>, dernière consultation : janvier 2021.
8. Cf. Schnedecker (1997, 2014, 2017, 2021), Longo et Todirascu (2014), Todirascu *et al.* (2017), Oberlé *et al.* (2018).
9. *Densité référentielle* : Nombre de maillons / Nombre de mots.
10. CRViewer : <http://boberle.com/projects/crviewer/index.html>, dernière consultation : janvier 2021.

RÉFÉRENCES BIBLIOGRAPHIQUES

- ARIEL M., 1990, *Accessing noun-phrase antecedents*, London, Routledge.
- BEACCO J.-C., 2000, « Écritures de la science dans les médias », *Les Carnets du Cediscor*, 6, p. 15-24.
- CHAROLLES M., 1988, « Les plans d'organisation textuelle : périodes, chaînes, portées et séquences », *Pratiques*, 57, p. 3-13.
- CHAROLLES M., 2002, *La référence et les expressions référentielles en français*, Paris, Éditions Ophrys.
- CHASTAIN C., 1975, « Reference and Context », in GUNDERSON K., *Language, Mind and Knowledge. Minnesota Studies in the Philosophy of Science*, 7, Minneapolis, University of Minnesota Press, p. 194-269.
- CHEN P. (陈平), 2015, « Référence, une notion essentielle pour la linguistique » (语言学的一个核心概念“指称”问题研究), *Rhétorique contemporaine (当代修辞学)*, 189, p. 1-15.
- CORBLIN F., 1985, « Les chaînes de référence : analyse linguistique et traitement automatique », *Intellectica : Revue de l'Association pour la recherche cognitive*, 1, p. 123-143.
- DUBOIS J., GIACOMO M., GUESPIN L., MARCELLESI C., MARCELLESI J.-B. et MEVEL J.-P., 2012, *Le Dictionnaire de linguistique et des sciences du langage*, Paris, Larousse.
- FANG M. (方梅), 2018, « Phrases au sujet zéro dans le genre narratif en chinois » (汉语叙事语体中零形主语小句的句法整合功能), in WANG X.-L. et HU Q. (王秀丽 & 胡庆), *Analyse textuelle : Études sur la répartition et la méthodologie d'annotation des chaînes de référence en chinois et en français (篇章分析—对汉语、法语指称链条分布规律的实证研究及其标注)*, Pékin, Beijing Language and Culture University Press (北京语言大学出版社), p. 15-36.
- HABERT B., 2005, *Instruments et ressources électroniques pour le français*, Paris, Éditions Ophrys.
- HUANG Y., 2000, *Anaphora: A Cross-linguistic Study*, New York, Oxford University Press.
- KIM S.-Y., à par., *Étude contrastive des chaînes de référence et problèmes de la coréférence : en français et en coréen*, thèse de doctorat, Université de Strasbourg.
- KLEIBER G., 2001, « Remarques sur la dénomination », *Cahiers de praxématique*, 36, p. 21-41.
- LANDRAGIN F., 2004, « Saillance physique et saillance cognitive », *Corela, Cognition, Représentation, Langage*, 2/2, p. 1-30.
- LI C. N. et THOMPSON S. A., 1989, *Mandarin Chinese: A functional reference grammar*, Vol. 3, Berkeley & Los Angeles, University of California Press.
- LIU D.-Q. (刘丹青), 2002, « La référence générique en chinois : aspect sémantique et aspect syntaxique » (汉语类指成分的语义属性和句法属性), *Studies of Chinese Language (中国语文)*, 290, p. 411-422.
- LONGO L. et TODIRASCU A., 2014, « Vers une typologie des chaînes de référence dans des textes administratifs et juridiques », *Langages*, 195, p. 79-98.

- OBERLÉ B., SCHNEDECKER C., BAUMER E., CAPIN D., GLIKMAN J., GUO C., REVOL T., TODIRASCU A. et TUSHKOVA J., 2018, « Les chaînes de référence dans les textes encyclopédiques du 12^e au 21^e siècle : étude longitudinale », *Travaux de linguistique*, p. 67-141.
- PINAULT-SORENSEN M., 1993, *L'encyclopédie*, Paris, Presses Universitaires de France.
- SCHNEDECKER C., 1997, *Nom propre et chaînes de référence*, Paris, Klincksieck.
- SCHNEDECKER C., 2014, « Chaîne de référence et variation selon le genre », *Langages*, 195, p. 23-42.
- SCHNEDECKER C., 2017, « Les chaînes de référence : une configuration d'indices pour distinguer et identifier les genres textuels », *Langue française*, 195, p. 53-72.
- SCHNEDECKER C., 2021, *Les chaînes de référence en français*, Paris, Éditions Ophrys.
- SCHNEDECKER C. et LANDRAGIN F., 2014, « Les chaînes de référence : présentation », *Langages*, 195, p. 3-22.
- SCHWARZ D., 1979, *Naming and Referring: the Semantics and Pragmatics of Singular Terms*, Berlin, Walter de Gruyter.
- TEUBERT W. et ČERMAKOVA A., 2004, « Directions in corpus linguistics », *Lexicology and Corpus Linguistics*, p. 113-166.
- TODIRASCU A., FRANÇOIS T., BERNHARD D., GALA N., LIGOZAT A.-L. et KHOBZI R., 2017, « Chaînes de référence et lisibilité des textes : le projet ALLuSIF », *Langue française*, 195, p. 35-52.
- XU J.-J. (徐赳赳), 1990, « Analyse discursive de *ta* (il) dans les textes narratifs » (叙述文中“他”的话语分析), *Studies of Chinese Language* (中国语文), 218, p. 325-337.
- XU J.-J. (徐赳赳), 2010, *Text Linguistics in Contemporary Chinese* (现代汉语篇章语言学), Pékin, The Commercial Press (商务印书馆).