e-ISSN 2385-3042

Annali di Ca' Foscari. Serie orientale
Vol. 61 – Giugno 2025

Mapping Referents in Chinese and Italian: A Corpus-Based Analysis of Anaphora, Deixis, and Other Referential Relations

Anna Morbiato
Università Ca' Foscari Venezia, Italia; The University of Sydney, Australia

Abstract This paper provides an overview of different types of referential expressions, including *deixis* and *anaphora*, and explores the various linguistic forms used to encode these relationships in Chinese and Italian – two languages that exhibit distinct reference-tracking systems: the former largely inference-based, the latter primarily morpho-syntactic. To this end, the study examines empirical data from both large and small corpora, spanning a range of genres and text types. The contrastive analysis includes both qualitative and quantitative observations, highlighting similarities and differences, and revealing novel or previously understudied patterns. The results have implications not only from a typological perspective but also for applied fields such as natural language processing and second/foreign language acquisition.

Keywords Referential relationships. Anaphora. Deixis. Corpus study. Chinese. Italian.

Summary 1 Introduction. – 2 Laying the Background: Review of the Literature. – 2.1 Reference, Deixis, and Anaphora. – 2.2 Typological variation. – 3 The Study: Objectives, Methodology, and Data. – 4 The Analysis: Referential Mapping and Forms. –4.1 Exophora. – 4.2 Anaphora. – 5 Conclusions.

Edizioni
Ca'Foscari

Peer review

Submitted 2025-04-23
Accepted 2025-06-30
Published 2025-07-31

Open access

© 2025 Morbiato | 🅭🅯 4.0

Citation Morbiato, A. (2025). "Mapping Referents in Chinese and Italian: A Corpus-Based Analysis of Anaphora, Deixis, and Other Referential Relations". *Annali di Ca' Foscari. Serie orientale*, 61, 385-424.

DOI 10.30687/AnnOr/2385-3042/2025/01/012

1 Introduction

Referential relationships are a fundamental yet complex component of linguistic organisation.[1] They are crucial for ensuring textual coherence and cohesion, as they form the backbone of a text. For discourse to be intelligible, reference and cohesion are essential – particularly for tracking participants accurately across sequences of clauses. This means that in a text describing situations involving the same participants or entities, these must be referred to in ways that allow readers or listeners to identify them as either the same as or distinct from those mentioned earlier (Foley, Van Valin 1984, 1). Crucially, languages achieve this through different, often more or less overt, means. In other words, the mechanisms for encoding reference and maintaining cohesion vary cross-linguistically, ranging from morphosyntactic markers to pragmatic inference. This makes referential relationships and reference tracking among the most intricate areas of linguistic inquiry. Their complexity stems not only from the remarkable variation observed across languages but also from the inherently multifaceted nature of reference itself, which intersects with semantics, semiotics, syntax, and pragmatics. Over the past few decades, linguistic research has increasingly focused on reference and anaphora, with efforts aimed at explaining anaphoric phenomena within various syntactic theories and at exploring the significant cross-linguistic variation in reference-tracking mechanisms across typologically distant language families (cf. next section). However, substantial gaps remain – especially in the systematic classification of referential phenomena and in cross-linguistic, corpus-informed comparisons involving Chinese and Romance languages.

This paper aims to provide an overall sketch of the various types of referential relationships, outlining both primary categories and subtypes, and to examine how these are realised in Chinese in comparison with Italian and English. This work builds on the scheme of referential relationships outlined by Korzen (2023), and investigates its applicability to account for the forms and phenomena observed in the research corpus. The analysis is based on natural linguistic data, including large-scale, generalised corpora such as the PKU CCL corpus (Peking University) and the BCC corpus (Beijing

---

[1] Simplified Chinese characters and the Pinyin romanisation system have been used throughout the article. The glosses follow the general guidelines of the Leipzig Glossing Rules. Additional glosses include: BA = ‘Chinese 把 *bǎ* patient marker’; BEI = ‘Chinese 被 *bèi* marker’; COS = ‘change of state’; EXP= ‘experiential aspect marker’; SFP = ‘sentence final particle’; SP = ‘structural particle’. List of abbreviations: C. = Chinese; E. = English; I. = Italian. In this paper, the term ‘Chinese’ refers to *Pǔtōnghuà*, the standard language of the PRC.

Language and Culture University). Additionally, specific cases are examined where significant differences emerge among the languages under investigation. To this end, the analysis is conducted on small, self-compiled corpora tailored to each case study, consisting of texts in genres where these specific referential relationships are particularly prominent. When necessary, native Chinese speakers were also consulted. This is a preliminary study and does not aim for exhaustiveness, but rather seeks to lay the groundwork for a practical taxonomy of referential relationships in Chinese and beyond, while encouraging further comparative research – particularly through larger corpus-based studies.

The rest of the paper is structured as follows: Section 2 provides the background and a brief review of key studies in the literature on referential phenomena, though a detailed discussion of terms like anaphora and deixis falls outside the scope of this work. Section 3 outlines the study's objectives and framework and describes the data and corpora used. Section 4 presents the analysis and discusses the main findings. Section 5 concludes with a summary of the results and their implications.

2 Laying the Background: Review of the Literature

Referential relationships, reference tracking, and anaphora constitute particularly complex linguistic phenomena. This complexity is closely tied to the remarkable variation across languages in how referential relationships are encoded, as well as the inherently intricate nature of reference itself. In Chomsky's words,

> There are certain phenomena, like anaphora, which have just been extremely good probes; they've raised questions that have to be answered, and there are other things that also do, but I haven't seen many. […] It seems to me to make good sense to work intensively in those subareas where sharp questions seem to arise that can be answered in ways that have an explanatory character. (Chomsky 1982, 83)

Since Chomsky's observation, research on anaphora and reference has grown into a substantial and multifaceted area of linguistic investigation. Scholars have pursued two main lines of inquiry: on the one hand, explaining coreference and anaphoric phenomena within various syntactic frameworks; on the other, investigating the cross-linguistic diversity of reference-tracking mechanisms across typologically distant language families – work that has led to the development of typologies of reference-tracking systems. This has resulted in a diverse and heterogeneous body of literature, rooted in different theoretical frameworks, which in turn has contributed to terminological confusion. Terms like reference, deixis, and anaphora are not easy to define, given the vast literature on these matters that spans multiple fields and disciplines, including semiotics, philosophy, syntax, semantics, pragmatics, natural language processing, and psycholinguistics – cf. the related entries in reference books such as Horn and Ward (2004) and Mey (2009). Hence, no unified consensus exists on how key notions such as *anaphora* and *deixis* should be defined (Cornish 2010; Huang 2000a; Levinson 1983), nor on how a taxonomy of referential relationships should be structured. The next sections will attempt to provide a brief sketch of this complex landscape.

2.1 Reference, Deixis, and Anaphora

Reference is a fundamental concept in linguistics that deals with how language users identify and point to entities in the world or in discourse. It encompasses the ways in which words, phrases, and expressions link to people, objects, events, and abstract concepts, allowing for meaningful communication. The scholarship on the concept of reference itself is highly complex and multifaceted, as it intersects various disciplines, from philosophy and psycholinguistics to semiotics and linguistics. Given this breadth, a comprehensive account lies far beyond the scope of the present paper. However, borrowing Sidnell and Enfield's words,

> if human cognition is fundamentally intentional in the sense of being about or directed towards something, reference is a form of shared intentionality in which the cognitive focus of two or more persons is aligned and jointly focused […] reference involves directing the attention of some other person to something, [which] may or may not be present in the immediate context of interaction. (Sidnell, Enfield 2017, 217)

In the latter part of their sentence, some scholars locate the distinction between deixis and exophoric reference on one hand, and anaphora and endophoric reference on the other – an issue that the following paragraphs will attempt to explore. However, there remains limited consensus in the literature regarding the precise definitions of these terms (see Green 2009 for discussion).

In traditional linguistics, deixis (from Ancient Greek: δεῖξις, lit. 'display, demonstration, or reference') is generally understood to be the linguistic encoding of references pointing to the spatiotemporal context of the utterance (Stapleton 2017). Typical examples include terms such as *I*, *here*, *now*, and *this* – Bühler's so-called 'pure deictic terms'. These terms are heavily context-dependent and represent a kind of cognitive centre of orientation for the speaker. More in general, deixis refers to the phenomenon wherein understanding the meaning of certain words and phrases in an utterance requires contextual information (Stapleton 2017). In other words, it concerns the ways in which languages encode or grammaticalise features of the context of utterance or speech event (Levinson 1983). For instance, saying "this one" while pointing to a book anchors the reference to the speaker's location, indicating relative proximity (see Levinson 1983; 2006) for a more detailed overview of the concept of deixis). Following Fillmore, Levinson (1983, 65) conveniently distinguishes between the interpretation of deictic forms through either gestural or symbolic usage. Gestural deixis requires direct sensory monitoring of the speech event, often involving a selecting gesture, as in *This one's genuine, but this one is a fake*. Symbolic deixis, on the other hand, relies on contextual knowledge, such as spatial or participant parameters, to be understood, as in *This city is really beautiful*. The complexity of this notion is nonetheless suggested by the vastness of the literature on the matter, which reveals diverse interpretations.[2]

The term anaphora (from Ancient Greek: αναφόρα, 'carrying back upstream'), in modern linguistics, has received various definitions, as well (Huang 2015). Generally, it refers to a relation between two or more linguistic expressions, where the interpretation of one (the anaphor or anaphoric expression) is determined by the interpretation of the other (the antecedent), as in (1) (Huang 2000a; 2015). This conception of anaphora is generally accepted by linguists both within formal and functional approaches, including several pragmaticians (cf. Cornish 2010), and lies at the basis of psycholinguistic experiments on anaphor interpretation as well as computational approaches to automatic anaphora resolution (see, e.g. Mitkov 2022). In the Chomskyan generative tradition, the term is instead used to distinguish between different types of NPs: an NP with the features [+anaphoric, -pronominal], such as a reflexive, is an anaphor, while an NP with the features [-anaphoric, +pronominal], like the pronoun *tā* 'he' in (2), is pronominal. Hence, the focus is more on the notion of coreference, with anaphors defined as inherently bindable expressions, in accordance with the Binding Theory. Often, the term anaphora is presented in a dichotomy (as e.g. Halliday, Hasan 1976; Korzen 2023), referring to expressions whose antecedent appears earlier in the discourse, in contrast to *cataphora*, where the antecedent appears later, as in (3). In this sense, both anaphora and cataphora can be subsumed under the term endophora, which refers to the relation where the anaphoric or cataphoric expression and its antecedent are within the same text or conversation. In contrast, the term exophora refers to cases where the antecedent of a deictic or indexical expression lies outside the text or conversation, as in (4) (cf. also **fig. 1**).

---

[2] See e.g. Green 2009; Levinson 1983; Sidnell, Enfield 2017; Verschueren 2009.

(1) (examples from Huang 2015)小明$_1$說ø$_1$/自己$_1$困了。

Xiǎo	*Míng$_1$*	*shuō*	*ø$_1$/zìjǐ$_1$*	*kùn-le*
Xiao	Ming	say	3SG/SELF	sleepy-PFV/COS

‘Xiǎo Míng said that he was sleepy’.

(2) 小明$_1$喜歡他$_2$。

Xiǎo	*Míng$_1$*	*xǐhuan*	*tā$_2$*
Xiao	Ming	like	3M.SG

‘Xiǎo Míng likes him’.

(3) ø$_1$ 越说,小明$_1$越高兴。

ø$_1$	*yuè*	*shuō*	*Xiǎo*	*Míng$_1$*	*yuè*	*gāoxìng*
	more	say	Xiao	Ming	more	happy

‘The more he talks, the happier Xiǎo Míng becomes’.

(4) [Referent in the physical environment and with selecting gesture]
他不是校长。

tā	*bù*	*shì*	*xiàozhǎng*
3SG	NEG	be	principal

‘He’s not the principal’.

In some other approaches (Cornish 2010; 2014), deixis and anaphora are conceived differently. The former serves to orient the addressee’s attention toward a new discourse entity, while the latter occurs when an anaphoric form – together with the clause in which it appears – signals the continuation of a previously established focus of attention. This is seen within discourse, understood as a hierarchically structured and dynamically evolving sequence of communicative acts aimed at achieving a specific goal, and shaped by the context in which it unfolds. In other words, anaphoric references maintain the discourse focus on an entity, situation, or proposition already introduced, thereby ensuring topic continuity. Deictic references, by contrast, signal topic shifts – that is, they redirect the addressee’s or reader’s attention to a new, discourse-initial entity, situation, or proposition. This approach aligns – albeit with differences in terminology – with observations found in several studies on the matter, e.g. topic continuity (Givón 1983), discourse topics and activation costs (Chafe 1994), pragmatic accounts of discourse anaphora (Huang 2000b), and accessibility theory (Ariel

2009), among others. These scholars focus more on activation states and processing costs, as well as topic continuity, rather than directly addressing the specific mechanisms of anaphoric relationships themselves. While these theories emphasise the cognitive processes involved in tracking referents, they provide valuable insights into how the accessibility of antecedents and the continuity of discourse topics influence anaphoric form choice and interpretation. For example, Huang (2000b) outlines the following distributional pattern of anaphora in conversation: (i) Establishment of reference tends to be achieved through the use of an elaborated form, notably a lexical NP. (ii) Shift of reference tends to be achieved through the use of an elaborated form, notably a lexical NP. (iii) Maintenance of reference tends to be achieved through the use of an attenuated form, notably a pronoun or a zero anaphor. Certainly, as Green (2009, 181) observes, "pragmatic approaches have in general attempted to blur the line between deixis and anaphora", contributing to terminological confusion and controversies. This can be seen, for example, in sentence (4) above, where Huang's analysis does not fully clarify whether the pronoun *tā* 'he' is deictic or anaphoric. Nonetheless, as Green (2009, 181) continues, "the tendency to see one element (deixis) as essentially exophoric and the other (anaphora) as intralinguistic remains". This study aligns with this tendency.

2.2 Typological Variation

Another major factor that makes anaphoric phenomena difficult to account for within linguistic formalisms and theories is the cross-linguistic diversity reference tracking exhibits. Languages, and language families, vary significantly in their mechanisms for signalling and tracking anaphoric relationships. In some, morphosyntactic structures play a prominent role in disambiguation, whereas in others, resolution relies heavily on inferential processes informed by context, shared knowledge, and world understanding. This has motivated scholars to develop typologies of reference-tracking systems, identifying gender, switch-reference systems, switch-function, and inference-based reference tracking systems (Comrie 1989; Foley, Van Valin 1984). These systems reflect a spectrum of linguistic strategies, with gender systems relying primarily on lexical and morphological cues, switch-reference and switch-function systems being predominantly grammatical, while inference systems being largely pragmatic and inference-based. A later typology outlined by Huang (2000) differentiates only between two types: syntactic and pragmatic, to which Italian and Chinese belong, respectively. A preliminary comparison between Chinese and Italian is offered by Huang (2016). Italian marks gender and number on nouns, including common nouns, and exhibits rich verbal morphology signalling number and person, which is absent in Chinese.

(5) (Huang 2016)
I. cammin-*o* ‘walk-1SG’, cammin-*i* ‘walk-2SG’, cammin-*a* ‘walk-3SG’, cammin-*iamo* ‘walk-1PL’, cammin-*ate*, ‘walk-2PL’, cammin-*ano* ‘walk-3PL’
C. 走 *zǒu* ‘walk’

This morphological marking plays a crucial role in referential tracking by establishing agreement between nouns, pronouns, and verbs. This enables several important functions. First, finite verbs agree in number and person with the subject, which facilitates the retrieval of omitted (dropped) pronominal subjects in unmarked, non-contrastive sentences (5I). These subjects can thus be systematically recovered through verbal inflection. Second, agreement allows for flexible word order – even highly ambiguous permutations such as OVS – without necessitating changes to verbal diathesis, such as active–passive alternations. Despite these variations, the syntactic and semantic roles of noun phrases can still be recovered, as illustrated in (6B) (note that English does not permit such word order without significant rephrasing in translation):

(6)

A.	Dove	*sono*	fini-*ti*	*gli*	spaghett-*i*?
	where	be.PRS.3PL	end-PTCP.3PL.M	the.M.PL	spaghetti-M.PL

‘Where did the spaghetti go?’

B.	*Li*	*ha*	mangia-*ti*	Alberto.
	3PL.M.ACC	have.PRS.3SG	eat-PTCP.3PL.M	Alberto(M.SG)
	O	V		S

‘Alberto ate them’.

More broadly, agreement facilitates the interpretation of referents across sentences. In certain constructions – particularly those involving past participles – object agreement in gender and number is also available, as seen in ***Li*** *ha mangia**ti*** in (6B) and *Non* ***li*** *avrebbe infastidi**ti*** in (7d) below:

(7) (Gombrich, *Breve storia del mondo*; cf. https://www.treccani.it/enciclopedia/anafora_(Enciclopedia-dell'Italiano)/)

(a) Si chiam-*ava*(3SG) *Rodolfo d'Asburgo*(M.SG),

(b) e *i*(M.PL) princip-*i*(M.PL) tedesc-*hi*(M.PL) *lo*(3M.SG) *avevano*(3PL) elett-*o*(M.SG) re nel 1273 nella speranza che,

(c) essendo ø *un*(M.SG) cavalier-*e*(M.SG) pover-*o*(M.SG) e sconosciut-*o*(3M.SG),

(d) ø non *li*(3PL) *avrebbe*(3SG) infastidit-*i*(PL) troppo.

(e) Ma ø non *avevano*(3PL) fatto i conti con la *sua*(SG) abilità e la *sua*(SG) giustizia.

'His name was Rudolf of Habsburg, and the German princes had elected him king in 1273 in the hope that, being a poor and unknown knight, he would not bother them too much. But they had not reckoned with his ability and his justice'.

A very different picture emerges with Chinese, which notoriously lacks gender or number marking on nouns or verbal morphology. Rather, Chinese, along with many other East and Southeast Asian languages, relies on an inference-based reference-tracking system, where disambiguation is mainly achieved through pragmatics and world knowledge (Huang 1994; 2000a; 2015). These systems are characterised by the extensive use of zero anaphora, which can appear in various syntactic positions (8) and is capable of recalling antecedents across adjacent sentences, even in the presence of intervening referents (cf. Tao, Healy 1996 and Xu 2015 for examples). When more referents are present, both salience and reference play crucial roles in antecedent retrieval (Huang 1994; Xu 2015). Huang argues this is evident in (9), where the control verb *shuofu* 'persuade' can prefer either agent/subject or patient/object control depending on context and real-world knowledge. In (9a), *shuofu* typically expresses a request for action, favouring patient/object control, but in (9b), inference/real-world knowledge overrides other constraints, as it more likely that a surgeon will operate on a patient, not the other way around.

(8) (Huang 2016, 30)

A: 老师$_1$有没有表扬过小明$_2$?

B: Ø$_1$表扬过Ø$_2$。

A:	*lǎoshī$_1$*	*yǒu*	*méi*	*yǒu*	*biǎoyáng-guo*	*Xiǎomíng$_2$?*
	teacher	have	NEG	have	praise-EXP	Xiaoming
B:	*Ø$_1$*	*biǎoyáng-guò*	*Ø$_2$*			
B:	Ø	praise-EXP	Ø			

A: 'Has the teacher$_1$ ever praised Xiaoming?' B: '(She$_1$) has praised (him)'.

(9) (Huang 2016, 31)

a. 病人$_1$说服医生$_2$Ø$_2$明天给他$_1$开刀。

bìngrén$_1$	*shuōfú*	*yīshēng*$_2$	Ø$_2$	*míngtiān*	*gěi*	*tā*$_1$	*kāidāo*
patient	persuade	doctor		tomorrow	for	3SG	operate

‘The patient$_1$ persuaded the surgeon$_2$ that (he$_2$) will operate on him$_1$ tomorrow’.

b. 医生$_1$说服病人$_2$Ø$_1$明天给他$_2$开刀。

yīshēng$_1$	*shuōfú*	*bìngrén*$_2$	Ø$_1$	*míngtiān*	*gěi*	*tā*$_2$	*kāidāo*
doctor	persuade	patient		tomorrow	for	3SG	operate

‘The surgeon$_1$ persuades the patient$_2$ that (he$_1$) will operate on him$_2$ tomorrow’.

These peculiarities have attracted considerable attention from linguists regarding how Chinese, unlike English, uses fewer overt markers while still allowing for accurate disambiguation, as well as on the mechanisms and reasons behind the licensing of zero anaphora.[3] This is why the majority of existing studies on Chinese focused on the linguistic forms and mechanisms employed to encode anaphora as well as the challenges of incorporating these phenomena into various syntactic frameworks (cf. Huang 1994; 2000a; 2015, among others). Huang (1994), for example, aimed to demonstrate the inadequacy of a purely syntactic approach to explaining Chinese anaphora. Through an in-depth analysis of phenomena such as long-distance reflexivisation, zero anaphora, and discourse anaphora, he developed a pragmatic theory of anaphora grounded in the neo-Gricean framework of conversational implicatures. His later cross-linguistic study (2000) explored various formalisms, both syntactic and semantic, including Optimality Theory (OT) and, to a lesser extent, Generalized Phrase Structure Grammar (GPSG) and Lexical Functional Grammar (LFG). Despite this broader scope, his work maintained its primary focus on the role of pragmatics in licensing and disambiguating anaphora.

Despite the increasingly extensive body of literature – of which the above sections provide only a brief outline – many aspects of referential relationships remain underexplored. Few corpus-informed studies have examined in detail the differences in the linguistic forms that encode various types of referential relationships across two (or more) typologically diverse languages. In particular, few corpus-informed studies have examined in detail how different linguistic forms encode various types of referential relationships across two (or more) typologically diverse languages. In the case of Chinese, while considerable attention has been devoted to anaphora – especially zero anaphora – most existing work has concentrated on mechanisms for choosing among anaphoric forms, the cognitive processes underlying antecedent retrieval, and the challenges of accounting for certain phenomena (such as zero anaphora and long-distance reflexivisation) within formal syntactic frameworks, as e.g. zero anaphora and long-distance reflexivisation (Huang 1994; 2000a). Conversely, to my knowledge, few studies have attempted to develop a systematic taxonomy of the different types of relationships and corresponding forms through which reference and anaphora are realised. Furthermore, with some exceptions primarily focused on English, little research has explored the function-form mappings of anaphoric phenomena in Chinese in a comparative perspective. Lastly, considerable confusion persists in the literature on Chinese regarding the boundaries between anaphora, deixis, and other related categories. Yet these concepts are essential – not only for theoretical linguistic analysis but also from an applied perspective. For instance, they are central to current research in natural language processing (see Morbiato 2024 for an overview), and they remain largely overlooked in language pedagogy despite their clear relevance for communicative competence.

---

[3] Huang 1984; Li, Thompson 1979; Pu, Pu 2014; Sun 2020; Tao, Healy 1996; 2005; Tu 2010.

3 The Study: Objectives, Methodology, and Data

Given the background outlined so far, this study aims to explore how different types of referential relationships are expressed in Chinese, in comparison with Italian and, where relevant, English. Specifically, it proposes an overall scheme of referential relationships based on Korzen (2023), outlining both major categories and subcategories. Korzen sketches a taxonomy of referential relationships – what he describes as a text’s referential skeleton – based on Italian, illustrating various types of anaphoric, cataphoric, and exophoric reference, with particular attention to the relatively underexplored phenomenon of so-called *unfaithful anaphors*. His work is adopted as a baseline here, as it offers one of the most detailed analyses and taxonomies of Italian data, thus enabling a more nuanced comparison with Chinese. The present study aims to assess the applicability of Korzen’s taxonomy to Chinese data in comparison with Italian, and, where relevant, English. It also seeks to identify any new patterns or forms of reference that emerge from the corpus analysis. To this end, examples that closely resemble those discussed in Korzen (2023) were identified in the corpora to ensure greater comparability. This allows for a more controlled evaluation of whether the same referential categories and distinctions hold across languages, or whether language-specific patterns emerge in the Chinese and Italian data.

This study is broadly situated within the field of cross-cultural pragmatics (House, Kádár 2021). As Alcón and Safont Jordà (2008,

193) note, pragmatic competence requires “knowledge of those rules and conventions underlying appropriate language use in particular communicative situations and on the part of members of specific speech communities”. While similarities in the use of anaphoric forms are attested crosslinguistically, languages do not necessarily employ the same set of referring expressions to encode a given relationship. Even when forms appear superficially similar, they may differ in terms of coding principles such as informativity, rigidity, and attenuation (Ariel 2009, 3). This approach builds on contrastive pragmatic research – namely, the systematic comparison of language use across comparable texts or corpora – and employs ancillary methods that precede or follow contrastive analysis, serving as pilot studies to identify relevant features for comparison. It relies on several key criteria: the use of corpora, the application of both qualitative and quantitative methods, the inclusion of variation or multiple languages to ensure comparability, and the consideration of different data types – both naturally occurring and elicited. For the present study, a variety of linguistic data were gathered from both large-scale, generalised corpora and small, self-compiled corpora. Data collection utilised two primary large-scale corpora: the PKU CCL Corpus, Peking University, and the BCC Corpus, Beijing Language and Culture University. In addition, the study also consists of three case studies on self-compiled corpora consisting of texts from specific genres, including news reports on car accidents, procedural texts such as recipes, and political discourse. These genres were selected because they specifically exhibit the target phenomena, as will be explained in each dedicated section. As preliminary steps, native Chinese speakers were also consulted to provide translations for specific examples from the literature and judgments regarding the naturalness or acceptability of various referential forms in particular contexts. The scope of the data collected reflects the exploratory nature of this study, which is based on preliminary qualitative and quantitative observations. Accordingly, the study does not aim for representativeness or exhaustiveness. Instead, it seeks to offer initial insights that may guide and inspire further comparative research – particularly through larger, corpus-based investigations focusing on specific genres.

4 The Analysis: Referential Mapping and Forms

The sections that follow explore the mapping of referential relationships and the different forms they display in the analysed dataset. Figure 1 illustrates Korzen’s (2023, 107) scheme of referential relationships: exophoric reference points to an element outside the text or discourse itself, which could consist of an element from the immediate discourse context, external world knowledge, or an individual’s mental representation of a concept or object **[fig. 1]**. In contrast, endophoric reference refers to elements contained within the text or discourse in which the reference occurs, pointing either to a previously mentioned or to an upcoming element within the same communicative exchange. Specifically, anaphora happens when the referential expression points back to something mentioned earlier in the discourse, while cataphora occurs when the reference anticipates something that will be introduced later. The sections that follow will explore each of these types of reference in detail, examining their specific characteristics and providing an overview of their related subcategories. Korzen’s Italian examples are compared with similar examples drawn from Chinese corpora (mainly BCC and CCL), their English translations, and additional data collected by the author.

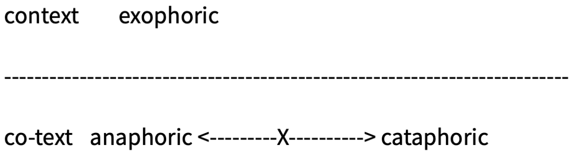

Figure 1 Korzen’s (2023) referential relationships (*valenze di rinvio*)

4.1 Exophora

As previously mentioned, exophora is broadly intended as a type of reference in which a linguistic expression points to an entity or information external to the text or conversation, yet mutually accessible to both the speaker and the listener. Korzen’s scheme of exophoric relationships predicts two broad types of identification, individual and categorial identification, with possible subcategorisations as illustrated in figure 2 and in subsections below **[fig. 2]**. Typically, the entity being exophorically referred to is present in the context of the discourse (*deixis*), which makes it cognitively salient for both interlocutors (Chafe 1976). In other cases, it points to an entity which is identifiable by both interlocutors in that accessible / present in the knowledge they share (*common knowledge*), although it is not available in the immediate context – cf. Prince’s definition (1981) of “New Unused”. Another case is *generic* reference, which points to a category of things that is mentally identifiable by the hearer (for an overview of the notions of accessibility, identifiability, and shared knowledge, I refer the reader to the seminal works by Chafe (1976), Prince (1981), Lambrecht (1994), *inter alia*.

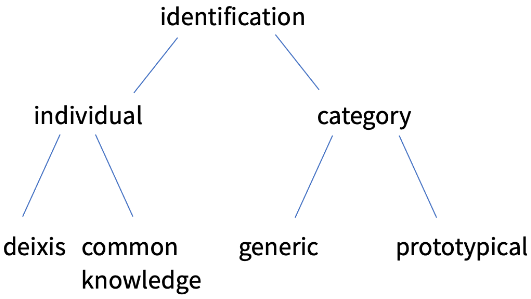

Figure 2 Exophoric reference (Korzen 2023, 108)

The following sections are devoted to identifying these subtypes and the forms they can be encoded with across the three languages, as well as exploring any additional patterns.

4.1.1 Deictic Reference

In deictic reference, individual entities (either animate or inanimate) are identifiable because they are present in the context of the discourse. Korzen's deictic reference broadly corresponds to what Levinson (2006) calls gestural usage, which necessarily involves understanding the context and the spatial-temporal positioning of the interlocutors. Deictic reference can be encoded through various linguistic means across all languages. Here are some Chinese examples drawn from corpora that are very similar to the Italian sentences discussed by Korzen (2023), to ease the comparison:

(10) C. 瞧你的狗, 咬我的新鞋! (BCC)

qiáo	*nǐ*	*de*	*gǒu*	*yǎo*	*wǒ*	*de*	*xīn*	*xié*
look.at	2SG	SP	dog	bite	1SG	SP	new	shoe

I. Guarda, il tuo cane sta morsicando le mie scarpe nuove!

E. Look, your dog is biting my new shoes!

(11) C. 瞧瞧那个。Ø太美妙了 (BCC)

qiáo-qiáo	*nà*	*ge*	*Ø*	*tài*	*měimiào*	*le*
look.at-look.at	that	CFL	Ø	too	beautiful	SFP

I. Guarda quello. Ø (È) davvero bello!

E. Look at that. It's wonderful!

(12) C. 那只猫很漂亮 (BCC)

nà	*zhǐ*	*māo*	*hěn*	*piàoliang*
that	CLF	cat	very	beautiful

I. Molto bello quel gatto.

E. That cat is beautiful.

(13) C. 你瞧我。(CCL)

nǐ	*qiáo*	*wǒ*
2SG	look.at	1SG

E. Look at me!

I. a. Guarda me!
   b. Guardami!

(14) C. 你要Ø吗? (BCC)

nǐ	*yào*	*Ø*	*ma*
2SG	want	Ø	Q

I. Ne vuoi?

E. Do you want some?

(15) C. Ø好漂亮! (BCC)

Ø	*hǎo*	*piàoliang*
	very	beautiful

I. Ø Molto bello!

E. Ø So beautiful!

All three languages make use of full NPs determined by demonstratives, possessive pronouns, articles/number-classifier constructions,[4] demonstrative headless NPs/pro forms (including pronouns), and zero forms. A first difference can be seen in possessive modifiers, which in Italian are determined by the article – absent in both Chinese and English (10). Moreover, English differs from Italian and Chinese in that pronominal subjects cannot be omitted across sentences (11E), whereas such omission is permissible in both Chinese and Italian (11C-11I), although for different reasons (see Section 4.2.1). Another substantial difference lies in the distinction between weak and strong pronouns in Italian – a feature that is absent in both English and Chinese. Weak pronouns include, for example, accusative-case suffixes such as -mi, -ti, -lo, and partitive pronouns like ne. The weak pronoun in (13Ia) conveys unmarked deictic reference, while the strong pronoun in (13Ib) feels marked, often carrying additional emphasis, such as contrast – (13Ib) means ‘look at me, not at someone else’. Lastly, partitivity is conveyed differently in the three languages: in (14), Chinese employs a zero, while Italian and English use pronouns, the former the clitic *ne*, the latter the pronoun *some*.

---

[4] A well-known difference between the three languages concerns the article versus classifier system. Italian and English exhibit an article system, a feature not shared by Chinese, which instead exhibits a demonstrative/numeral plus classifier (CLF) noun modification.

Korzen notes that deixis can also be *indirect* (or *associative*), i.e. refer to entities that are not directly perceivable but are linked to present and visible referents or contexts. Typical examples include signs or labels. For instance, in the following examples, the underlined phrases indicate any potential entities within the temporal and spatial context of the utterance. These may be specific (as in (16), where the dog owner can identify the specific dog) or nonspecific (as in (17), referring to any wild birds which visitors reading the sign might encounter). Indirect or associative deixis is characterised by limited linguistic realisations. According to Korzen (2023), Italian only allows noun phrases with a definite article (which includes contracted prepositions).[5] In fact, plural bare nouns are also allowed (18). Chinese, conversely, employs bare nouns or nouns modified by descriptive adjectives (16-17C), while English allows all the above means:

(16) C.当心恶犬 (CCL)
dāngxīn *è-quǎn*
beware vicious-dog
I. Attenti al cane!
E. Watch out for the dog!

(17) C. 不要喂饲野鸟 (BCC)
bùyào *wèisì* *yě-niǎo*
NEG.IMP feed wild-bird
I. Non dare da mangiare agli uccelli selvatici.
E. Don’t feed wild birds.

---

5 In Italian, contracted prepositions, known as *preposizioni articolate*, occur when a preposition combines with the definite article to form a single contracted word. For example, *in* + *il* = *nel* ‘in the’, *di* + *il* = *del* ‘of the’, and *a* + *il* = *al* ‘to the’. This phenomenon is, in Italian, obligatory, standing in contrast to English, where prepositions and articles remain distinct, though some instances of contraction may appear in casual speech, such as informal reductions (e.g. *to the* becoming *ta*).

(18) (source: https://hinterland.camp/it/)

Silenzio	notturno	da-lle	22.00	non	d-are	da	mangi-are
silence	night.time	from-the	10 p.m.	NEG	give-INF	to	eat-INF
a	can-i	e	uccelli.				
to	dog-PL	and	bird-PL				

‘Nighttime silence from 10 p.m.; do not feed dogs or birds’.

4.1.2 Common Knowledge-Driven Identification

Common knowledge-driven identification occurs when a noun phrase points to a referent that is neither directly nor indirectly present in the context but is nonetheless identifiable for both interlocutors, as it is unique within a specific referential universe and cognitively salient to both of them (Levinson 2006 refers to this as symbolic usage). The referential universe may vary in scope/breadth and may consist of a shared context between the interlocutors (e.g. a specific family, city, job context or environment). In this case, while Chinese employs bare noun phrases, English and Italian typically rely on the definite article, although bare nouns may also occur in specific cases (21).

(19) C. 我看见了太阳! (BCC)

wǒ	*kàn-jiàn-le*	*tàiyáng*
1SG	see-RES-PVF	sun

I. Ho visto il sole!
E. I saw the sun!

(20) C. 我今天见到校长了。 (Source: book 暗之匙)

wǒ	*jīntiān*	*jiàn-dào*	*xiào-zhǎng*	*le*
1SG	today	see-RES	school-principal	PFV/COS

I. Oggi ho visto il preside.
E. Today I saw the school principal.

(21) C. 我真想立刻见到妈妈 (BCC)

wǒ	*zhēn*	*xiǎng*	*lìkè*	*jiàn-dào*	*māma*
1SG	really	want	immediately	see- RES	mother

E. I so want to see mom right away.
I. Voglio vedere mamma/la mamma sùbito.

One last case, not covered in Korzen’s outline, involves indefinites introducing previously unmentioned referents into discourse; in this case, the referent may be either specific (i.e. identifiable by the speaker only) or not (unidentifiable for both). These are generally encoded as indefinite noun phrases (with indefinite articles/adjectives in Italian and English and with the numeral *yī* 'one' + CLF or else as postverbal bare NPs in Chinese, although see Morbiato 2020 for a quantitative and qualitative corpus-based account of sentence-initial indefinites (so-called 无定NP主语句 *wúdìng* NP *zhǔyǔ jù*) in Chinese).

(22) C.在窗边他看见了一个鸟笼养着一只金丝雀 (BCC)

zài	*chuāng-biān*	*tā*	*kàn-jiàn-le*	*yī*	*ge*	*niǎolóng*
at	window-side	3SG	see-RES-PFV	one	CLF	birdcage
yǎng-zhe	*yī*	*zhī*	*jīnsīquè*			
raise-PROG	one	CLF	canary			

E. At the window he saw a cage with a canary.

I. Alla finestra vide una gabbia con un canarino.

In Chinese, a key difference between indefinite noun phrases and (proper) names/bare NPs lies in the importance of the entities they introduce as potential discourse topics. Xu's (2010) study on narrative Chinese discourse shows that 53.8% of entities introduced by names or bare noun phrases are never mentioned again, compared to only 4.8% for those introduced by indefinite noun phrases. Entities introduced by indefinite noun phrases are referred to far more frequently later in the text, with an average of 15.8 mentions, rising to 22.4 for those introduced by existential-presentative constructions, such as *cóngqián, yǒu* 'long ago, there was' in (23). These findings suggest that indefinite noun phrases, particularly in existential-presentative constructions, play a crucial role in introducing major discourse topics in Chinese. Overt pronouns, such as *tā* 'he' in (23b), along with zero pronouns, demonstrative expressions, and bare noun phrases referring to entities already introduced into discourse seem then to perform more the function of topic maintenance and reference tracking (anaphora), which we will examine in greater in detail in Section 4.2.

(23) (Xu 2015) (a) 从前, 有个勤劳的铁匠, (b) 他有个儿子$_j$,(c) Ø$_j$ 快满二十岁了…

(a)	*cóngqián*	*yǒu*	*ge*	*qínláo*	*de*	*tiějiàng*
	long.ago	there.be	CLF	hard.work	SP	blacksmith

(b)	*tā*	*yǒu*	*ge*	*érzi*$_j$
	3SG	have	CLF	son$_j$

(c)	*Ø*$_j$	*kuài*	*mǎn*	*èrshí*	*suì*	*le*
	Ø$_j$	soon	reach	twenty	year.old	COS

'Long ago, there was a hardworking blacksmith. He had a son who was nearly twenty'.

4.1.3 Case Study 1: Exophoric Coreference in Political News Reports

Korzen (2023, 109) observes that, in Italian, a further exophoric phenomenon is also attested, which he calls exophoric coreference. Here, the same referent is pointed to exophorically by two distinct noun phrases with lexically different heads. In the example below, the two noun phrases il Presidente della Repubblica and il Capo dello Stato refer to the same specific referent. In Korzen's view, they each identify the referent exophorically and independently of one another: the latter would, in other words, not constitute an anaphoric form.

(24) (Korzen 2023, 109)

Oggi il Presidente della Repubblica Sergio Mattarella terrà il suo tradizionale discorso di fine anno, il primo dell'era Giorgia Meloni. Il Capo dello Stato parlerà…

'Today, the President of the Republic, Sergio Mattarella, will deliver his traditional end-of-year speech, the first of the Giorgia Meloni era. The Head of State will speak…'.

In Italian, such occurrences are common in journalistic genres and narrative texts, where synonymic substitution is employed to avoid repetition. This raises the question of whether similar strategies are found in Chinese. According to the native speakers consulted in the present study, this strategy does not appear to be available: once a referent is introduced, it is typically recalled using lighter noun phrases, such as bare noun phrases with proper nouns or pronouns (thus falling into the category of coreferential anaphora, Section 4.2). This is their translation for (24):

(24') 今天，意大利共和国总统塞尔焦·马塔雷拉将发表他一年一度的年终讲话，这将是乔尔吉娅·梅洛尼执政以来的首次讲话。马塔雷拉总统将会谈到……

jīntiān	*Yìdàlì Gònghéguó*	*zǒngtǒng*	*Sài'ěrqiáo Mǎtǎléilā*	*jiāng*
today	Italy Republic	president	Sergio Mattarella	fut
fābiǎo	*tā*	*yì-nián-yí-dù*	*de*	
deliver	3SG	one-year-one-time	SP	
niánzhōng jiǎohuà	*zhè*	*jiāng*	*shì*	
year-end speech	this	fut	be	
Qiáo'ěrjíyà Méiluònì	*zhízhèng*	*yǐlái*	*de*	
Giorgia Meloni	govern	since	SP	
shǒucì jiǎohuà	*Mǎtǎléilā*	*zǒngtǒng*	*jiāng huì*	*tándào*
first speech	Mattarella	president	fut	talk.about

Sentence (25), from a news article, similarly displays a heavy noun phrase Zhōngguó guójiā zhǔxí Xí Jìnpíng 'President of the People's Republic of China Xi Jinping' which is later referenced by repeating only the proper noun, Xí Jìnpíng:

(25) (source: https://www.zaobao.com.sg/)

中国国家主席习近平星期三(12月18日)下午抵达澳门, […]。(one sentence follows) 这是习近平担任国家主席任内第三度访问澳门…

Zhōngguó	*guójiā*	*zhǔxí*	*Xíjìnpíng*			
China	state	chairman	Xi.Jinping			
xīngqīsān	*(12 yuè*	*18 rì)*	*xiàwǔ*	*dǐdá*	*Àomén*	
Wednesday	(Dec.	18th)	PM	arrive	Macao	
zhè	*shì*	*Xíjìnpíng*	*dānrèn*	*guójiā*	*zhǔxí-rèn*	*nèi*
this	be	Xi.Jinping	serve.as	state	chair-person	in
dìsān	*dù*	*fǎngwèn*	*Àomén*			
third	time	visit	Macao			

‘Chinese President Xi Jinping arrived in Macau on Wednesday afternoon (December 18) […]. This is Xi Jinping’s third visit to Macau during his presidency…’.

However, one may wonder whether such cases might emerge if a larger body of texts is considered. Is this phenomenon also attested in Chinese newspapers? To what extent do Chinese and Italian differ in coreference encoding? These questions have been addressed through a corpus study conducted on political news reports, covering diplomatic travels by the presidents of China and Italy, consisting in 60 texts, 30 for each language.

In 100% Italian texts, repetition avoidance was achieved, in addition to proforms and faithful anaphors – i.e. NPs with the same head nouns (cf. Section 5.2.1), through a number of synonyms, such as *il Presidente Mattarella - il Capo dello Stato* (cf. Section 5.2.2), as in (26):

(26)

Il giorno successivo, sabato 9 novembre, il Presidente Mattarella ha incontrato il Primo Ministro, Li Qiang, e il Presidente dell‘Assemblea Nazionale del Popolo, Zhao Leji. Al termine dei colloqui, il Capo dello Stato ha poi deposto una corona al Monumento degli Eroi del Popolo in Piazza Tiananmen…

‘The following day, Saturday, November 9, President Mattarella met with Prime Minister Li Qiang and the Chairman of the National People’s Congress, Zhao Leji. At the end of the talks, the Head of State laid a wreath at the Monument to the People’s Heroes in Tiananmen Square’.

On the contrary, no such instances were found in Chinese, which only employed proforms and faithful anaphors – i.e. NPs with the same head nouns, as e.g. *Yìdàlì Gònghéguó Zǒngtǒng Sài’ěrjiō · Mǎtǎléilā (Sergio Mattarella)* ‘The President of the Italian Republic, Sergio Mattarella’ and *Mǎtǎléilā Zǒngtǒng* ‘President Mattarella’ in (27), or *Guójiā Zhǔxí Xí Jìnpíng* and *Xí Jìnpíng* in (28). Full NP repetitions were also observed.

(27) 意大利共和国总统塞尔焦·马塔雷拉(Sergio Mattarella)于今天即11月12日结束了对中华人民共和国的国事访问。在意大利与中华人民共和国建交55周年前夕, 此次访问证实了意大利进一步深化双边关系的意愿。

Yìdàlì Gònghéguó Zǒngtǒng Sài’ěrjiō · Mǎtǎléilā (Sergio Mattarella) yú jīntiān jí 11 yuè 12 rì jiéshùle duì Zhōnghuá Rénmín Gònghéguó de guóshì fǎngwèn. Zài Yìdàlì yǔ Zhōnghuá Rénmín Gònghéguó jiànjiāo 55 zhōunián qiányè, cǐ cì fǎngwèn zhèngshíle Yìdàlì jìnyībù shēnhuà shuāngbiān guānxì de yìyuàn.

11月8日星期五, 中华人民共和国主席习近平在人民大会堂为马塔雷拉总统举行了欢迎仪式, 并进行了双边会谈。

11 yuè 8 rì xīngqīwǔ, Zhōnghuá Rénmín Gònghéguó Zhǔxí Xí Jìnpíng zài Rénmín Dàhuìtáng wèi Mǎtǎléilā Zǒngtǒng jǔxíngle huānyíng yíshì, bìng jìnxíngle shuāngbiān huìtán.

‘The President of the Italian Republic, Sergio Mattarella, concluded his state visit to the People’s Republic of China on November 12. This visit, taking place on the eve of the 55th anniversary of diplomatic relations between Italy and the People’s Republic of China, reaffirmed Italy’s commitment to further deepening bilateral relations. On Friday, November 8, President Xi Jinping of the People’s Republic of China hosted a welcome ceremony for President Mattarella at the Great Hall of the People and held bilateral talks’.

(28) 2024年11月8日下午, 国家主席习近平在北京人民大会堂同来华进行国事访问的意大利总统马塔雷拉举行会谈。

2024 nián 11 yuè 8 rì xiàwǔ, Guójiā Zhǔxí Xí Jìnpíng zài Běijīng Rénmín Dàhuìtáng tóng lái Huá jìnxíng guóshì fǎngwèn de Yìdàlì Zǒngtǒng Mǎtǎléilā jǔxíng huìtán.

习近平指出, 今年是中意建立全面战略伙伴关系20周年。

Xí Jìnpíng zhǐchū, jīnnián shì Zhōng-Yì jiànlì quánmiàn zhànlüè huǒbàn guānxì 20 zhōunián.

‘On the afternoon of Nov. 8, 2024, President Xi Jinping held talks at the Great Hall of the People in Beijing with Italian President Mattarella, who is in China for a state visit. Xi noted that this year marks the 20th anniversary of the establishment of a comprehensive strategic partnership between China and Italy’.

4.1.4 Categorical Identification and Generic Reference

In Korzen’s outline **[fig. 2]**, the second major type of exophoric reference is categorical identification, more commonly referred to in linguistic literature as generic reference. Korzen further refines this category by distinguishing between generic and prototype reference. In the case of generic reference, the noun phrase denotes an entire category rather than an individual referent, presenting a description or evaluation as universally applicable to all members or subsets of that category (29-30). Alternatively, in prototype reference, the noun phrase points to a representative member of the category, serving as a generalised example of that class (31):

(29) C. 坐汽车比火车贵八九倍 (source https://www.sohu.com/)

zuò	*qìchē*	*bǐ*	*huǒchē*	*guì*	*bā-jiǔ*	*bèi*
sit	car	compare	train	expensive	8-9	time

I. Usare l'auto costa da otto a nove volte di più che prendere il treno.
E. Driving a car is eight to nine times more expensive than taking a train.

(30) C. 老虎是猫科动物 (source https://www.sohu.com/)

lǎohǔ	*shì*	*māokē*	*dòngwù*
tiger	be	feline	animal

I. La tigre è un felino.
E. The tiger is a feline.

(31) C. 我已经买车了 (BCC)

wǒ	*yǐjīng*	*mǎi*	*chē*	*le*
1SG	already	buy	car	PFV/COS

E. I've already bought a car.
I. Ho già comprato la macchina

While the above examples involve bare NPs in Chinese and definite-article-marked NPs in Italian, as noted by Korzen, data show that indefinite NPs may also carry generic reference in all three languages:

(32) C. 一个年轻人应当有志气。 (Lu, Pan 2009, Morbiato 2018)

yí	*ge*	*niánqīng-rén*	*yīngdāng*	*yǒu*	*zhìqì*
one	CLF	young-man	should	have	ambition

I. Un ragazzo giovane dovrebbe avere un po' di ambizione.
E. A young man should be ambitious.

(33) C. 咱生个孩子吧。(CCL)

zán	*shēng*	*ge*	*háizǐ*	*ba*
1PL	be.born	CLF	child	SFP

I. Facciamo un bambino.
E. Let's have a baby.

4.2 Anaphora

With regard to anaphora, Korzen (2023) proposes a scheme with a threefold division: coreferential, associative, and resumptive anaphors **[fig. 3]**. Coreferential anaphora (identity) involves an anaphor that refers back to a noun phrase (NP) in the preceding context, with both the anaphor and the antecedent pointing to the same entity. For example, in the sentence *John was tired. He went to bed*, *he* is a coreferential anaphor referring to *John*. While coreferential anaphora is the most straightforward and widely studied form of reference, associative and resumptive anaphora are often more complex, as they involve more nuanced relationships and therefore merit closer attention. Associative anaphora, also known as partial coreference or bridging, refers to something that is contextually or logically connected to the antecedent, as in *We checked the picnic supplies. The beer was warm*. This type of anaphora relies on shared knowledge, derived either from the immediate discourse or from the broader context. Resumptive anaphora, finally, involves the use of an anaphor to refer back to an entire clause, sentence, or a portion of it – often to emphasise or elaborate on previously mentioned information. For instance, in *The car broke down. It happened in the middle of the highway*, the word *it* is a resumptive anaphor referring not just to *the car*, but to the entire event of the car breaking down.

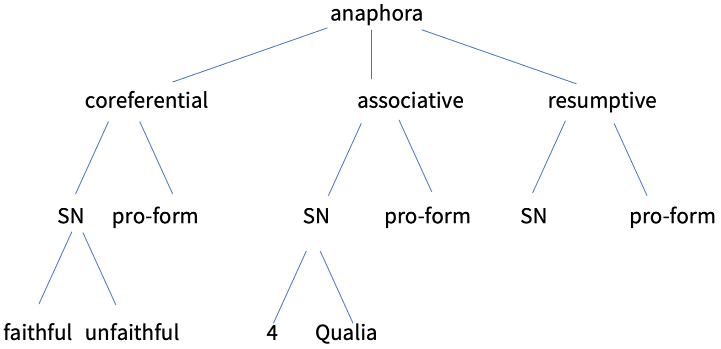

Figure 3 Anaphoric relationships (slightly adapted from Korzen 2023, 110)

4.2.1 Coreferential Anaphora

Coreferential anaphors are expressions that refer back to a specific antecedent previously introduced in the discourse. As in deictic reference, both definite noun phrases (34a-b) and pro-forms (34c) can be used. When a full noun phrase is used, Korzen (2023) distinguishes between faithful anaphora, where the head of the NP is lexically identical to the antecedent (34a), and unfaithful anaphora, where the head differs lexically, often being a hypernym, as e.g. *car* and *vehicle* in (34b).

(34) (Korzen 2023, 110)

Ho visto un'automobile. non conosco.
(a) *L'automobile* era di una marca che
(b) *Il veicolo* era di una marca che non conosco.
(c) Non *l*'avevo mai vista prima.

'I saw a car.
(a) *The car* was of a brand I don't know.
(b) […] *The vehicle* was of a brand I don't know
(c) I had not seen it before'.

A few observations on Korzen's example are, however, in order. As he states, it is a well-known fact that the anaphoric strength of a noun phrase exceeds that of a pronoun. In other words, strong anaphoric devices, such as full noun phrases, serve to either introduce entirely new referents into the discourse or to reintroduce previously mentioned, cognitively active referents after a significant absence or when other intervening referents have been introduced. Conversely, weaker means (e.g. pronouns or zero anaphors) are typically used to encode informationally given or highly accessible items.[6] This, however, implies that (34a) and (b) above would be appropriate, for instance, only in a context where another utterance intervenes between the two, such as (34a') below. Furthermore, sentence (b) sounds quite unnatural in a conversation context as that imagined by Korzen. A term like il veicolo would be more appropriate in contexts such as news reports or official documents (e.g. police reports on car accidents).

(34a') Ho visto *un'automobile* nel cortile ieri sera. C'erano anche un motorino e una bici. *L'automobile* era di una marca che non conosco.
'I saw a car in the courtyard last night. There was also a scooter and a bicycle. The car was of a brand I don't know'.

Proforms may also include zero anaphora (dropped pronouns). However, a noteworthy point concerns the licensing conditions for zero anaphora in these three languages. With rare exceptions, English allows coreferential subject omission only in adjacent clauses and never across sentence boundaries and does not permit zero anaphora in object position (exceptions include, for example, recipe texts and certain other genre-specific text types). Italian, as said earlier, allows the omission of all pronominal subjects (pro-drop), as the subject can be inferred from verb inflection agreement. However, pronominal objects must be overt, at the very least encoded by a clitic. Finally, Chinese permits the omission of both verbal arguments, even across different clauses (see Morbiato 2024, 127).

---

[6] Ariel 2009; Chafe 1994; Givón 1983; Huang 2000b.

(35) E. **He** dressed up, **Ø**$_j$ went down the stairs$_k$, **Ø**$_j$ walked out **the front door**$_i$. Then **he** closed **it**$_j$ behind him.

I.	**Ø**$_j$	Si	vest-ì,	**Ø**$_j$	scese	le	scale$_k$
		REFL	dress.up-PST.3SG		go.down-PST.3SG	the	stairs

e	**Ø**$_j$	usc-ì	da-lla	**porta**	**d’**	**ingresso**$_\iota$.
and		exit-PST.3SG	from-the	door	of	front

Poi **Ø**$_j$	**la**$_\iota$	chiu-se	a-lle	sue	spalle.
Then	3SG.F	close-PST.3SG	at-the	his	shoulder

C. 他穿好衣服, 走下楼梯, 走出前门。然后就顺手把门关上了。

tā$_j$	*chuān-hǎo*	*yīfú,*	**Ø**$_j$	*zǒu-xià*	*lóutī*$_k$,
3SG	dress-RES	cloth		walk-down	stairs

Ø$_j$	*zǒu-chū*	***qián-mén***$_i$	*ránhòu*
	walk-exit	front-door	afterwards

Ø$_j$	*jiù*	*shùnshǒu*	*bǎ*	***mén***$_i$	*guān-shàng-le*
	then	smoothly	BA	door	close-up-PFV/COS

4.2.2 Unfaithful Anaphora

According to Korzen, unfaithful coreferential anaphors are noun phrases whose heads differ lexically from their antecedents, often taking the form of hypernyms or synonyms (34b). These anaphors serve two main functions. The first is stylistic, aimed at avoiding repetition – a common preference in certain genres such as journalistic writing, where Italian in particular tends to discourage reiteration. This use contributes to lexical variation within the text, a phenomenon especially prominent in Italian. Lexical variation, however, generally reduces anaphoric strength compared to lexical identity. As a result, demonstratives – e.g. *quel* in (36), which explicitly anchor the anaphor to its antecedent within the same co-text, occur more frequently in unfaithful than in faithful anaphora (Korzen 2023, 114):

(36) Non si vuole con ciò pretendere che gli Stati Uniti cambino il proprio sistema, anche se sono molte e autorevoli le voci che si sono levate in quel paese [→ gli Stati Uniti] a chiedere l’adozione di una legislazione in questa materia.

‘This does not intend to claim that the United States should change its system, even though many authoritative voices have risen in that country [→ the United States] to call for the adoption of legislation on this matter’.

The second function of unfaithful anaphora relates to information structure: such anaphors introduce new information about the antecedent, adding descriptions, evaluations, or other details. For instance, in journalistic style, an unfaithful anaphor may broaden or reframe the antecedent to convey new content. This practice is found in both English and Italian, where such expressions enrich the discourse by layering additional meaning onto the initial referent.

(37) Tra i due litiganti, prova ad inserirsi il terzo incomodo: Cameron. *Il leader britannico* ha incontrato sia Putin che Obama, ed ora vuole giocarsi le sue carte.
'The US-Russia duel. […] Between the two quarrellers, Cameron, a third nuisance, tries to join in. *The British leader* met both with Putin and Obama and now wants to play his cards'.

Korzen's comparison of Italian and Danish reveals a significantly higher frequency of variation across different linguistic levels in Italian, including in coreferential anaphora. This variation is shaped not only by language-specific and cultural norms but also by text type: it is more frequent in "creative" genres (e.g. narratives, descriptions, arguments) than in technical, legal, or regulatory texts, where semantic precision is paramount.

Is unfaithful anaphora also available in Chinese? According to native informants, who were asked to translate the examples discussed so far, it is not. Below is a translation provided for (34):

(34'') 昨晚我在院子里看到了一辆汽车，
(a) 是我不认识的品牌。
(b) 那辆汽车是我不认识的品牌。

zuó-wǎn		*wǒ*	*zài*	*yuànzǐ-lǐ*	*kàn-dào-le*	*yī*	*liàng*	*qìchē*	
yesterday-night		1SG	be.at	yard-in	see-RES-PFV	one	CLF	car	
(a)	*shì*	*wǒ*	*bù*	*rènshí*	*de*	*pǐnpái*			
	be	1SG	NEG	know	SP	brand			
(b)	*nà*	*liàng*	*qìchē*	*shì*	*wǒ*	*bù*	*rènshí*	*de*	*pǐnpái*
	that	CLF	car	be	1SG	NEG	know	SP	brand

'I saw a car in the yard last night. That car/it was a brand I didn't know'.

With regard to (24), a faithful anaphora, i.e. the repetition of the NP head *Mattarella*, is felt to be the best option. As for (34), native speakers also reject the unfaithful anaphora option, claiming that only types (a) and (c) are acceptable – i.e. either faithful coreferential anaphora or proform – while the use of synonyms does not seem justified (34''). Nevertheless, a dedicated case study based on corpora will examine this type of anaphoric form in more detail (see Section 4.2.3).

Crucially, Chinese exhibits a particularly interesting pattern resembling unfaithful anaphors, but not found in the other two languages. It involves a proform, typically a pronoun, preceded by modifiers, usually adjectives or relative clauses. Just like unfaithful anaphors, this structure serves to provide additional information (either objective or subjective, such as the speaker's opinion or stance) about a referent already introduced in the discourse, which is now referred to via a pronominal proform.

(38) 为什么没有人帮帮可怜的他们? (BCC)

wèishénme	*méi*	*yǒu*	*rén*	*bāng-bāng*	*kělián*	*de*	*tāmen*
why	NEG	have	person	help-help	pitiable	SP	they

'Why doesn't anyone help 'the pitiable them'?'

(39) 李衍秀今年才开始"改行"种植棉花。37岁的他一直兼任永安村实业公司总经理。(BCC)

Lǐ	*Yǎnxiù*	*jīnnián*	*cái*	*"gǎiháng"*
Li	Yanxiu	this.year	only	"switch.profession"

zhòng	*miánhuā*	***37 suì***	***de***	***tā***	*yīzhí*
plant	cotton	37 year.old	SP	1SG	always

jiānrèn	*Yǒng'ān*	*cūn*	*shíyè*	*gōngsī*	*de*	*zǒng*	*jīnglǐ*
hold	Yong'an	village	industry	company	SP	general	*jīnglǐ*

'Li Yanxiu only this year, 'changed his profession' to cotton farming. 'The 37 yo he'/ Since he was 37 years old, he has been the part-time general manager of the Yongan Village Industrial Company'.

Translation into English and Italian of these sentences is not straightforward, as a similar pattern is entirely absent. This pattern does not fully fall into the category of unfaithful referential anaphors as defined by Korzen, since the head is not a different noun but rather a pronoun. Nonetheless, as said, it exhibits the functions of unfaithful anaphors, namely referring back to a highly accessible referent and adding information about that previously introduced referent. It certainly deserves further attention in corpus-based studies to explore its communicative functions and possible structural tendencies.

4.2.3 Case Study 2: Unfaithful Anaphors in Web News

The issue of unfaithful anaphora in Chinese calls for further investigation. To this end, a case study was conducted on a sample corpus comprising texts on two distinct topics: web news on political trips and general news reports, collecting 60 texts, 30 for each language. The quantitative analysis is compelling and shows that 100% of Italian texts involved repetition avoidance, whereas Chinese texts used either proforms or faithful anaphora, with very few exceptions. Below are some examples where anaphoric forms are highlighted:

(40) Un ordigno ha colpito ieri in tarda serata un'automobile in via Manzoni, a Latina. Il veicolo appartiene a un insegnante.
'An explosive device struck a car late last night on Via Manzoni in Latina. The vehicle belongs to a teacher'.

(41) Scontro alla rotonda tra un camion e un'automobile: grave il passeggero del veicolo.
'Collision at the roundabout between a truck and a car: the passenger of the vehicle is in serious condition'.

(42) In condizioni critiche, il venditore di un'auto su cui lo stesso stava viaggiando lungo via Slovenia. Alla guida, il potenziale acquirente del veicolo, sul sedile di fianco, il proprietario del mezzo. La trattativa, con test incluso, ha preso la piega peggiore alle 17.30, quando l'automobile ha svoltato per immettersi nella rotonda attigua.
'In critical condition, the seller of a car in which he was traveling along Via Slovenia. Driving was the potential buyer of the vehicle, while the owner sat in the passenger seat. The negotiation, including a test drive, took a turn for the worse at 5:30 PM when the car turned to enter the adjacent roundabout'.

In Italian, within this genre, avoiding repetition feels almost obligatory. Notably, in example (42), four different nouns – *auto, veicolo, mezzo,* and *automobile* – are used to refer to the same entity within a very brief passage consisting of just three sentences. In Chinese, on the contrary, all instances are NPs headed by the word *chē* 'car', with few exceptions where the noun is *chēliàng* 'vehicle', as can be seen below in (43) and (44).

(43) 最近，温州龙港的黄先生向“浙里帮”反映，8月31日一早，他停放在家楼下的轿车发生了一起“离奇”的车祸，而车辆后续的赔偿问题迟迟得不到解决。

zuìjìn	*Wēnzhōu*	*Lónggǎng*	*de*	*Huáng*	*xiānshēng*
recently	Wenzhou	Longgang	SP	Huang	Mr.
xiàng	*"Zhèlǐ Bāng"*	*fǎnyìng*	*8*	*yuè*	*31 rì*
towards	"Zheli Bang"	report	8	month	31 day
yī-zǎo	*tā*	*tíngfàng*	*zài*	*jiā*	*lóuxià*
one-early	3SG	park	at	home	downstairs
de	*jiàochē*	*fāshēng-le*	*yī*	*qǐ*	*"líqí"*
SP	car	happen-PFV	one	CLF	"bizarre"
de	*chēhuò*	*ér*	*chēliàng*	*hòuxù*	*de*
SP	accident	but	vehicle	subsequent	SP
péicháng	*wèntí*	*chíchí*	*dé*	*bùdào*	*jiějué.*
compensation	issue	long.time	get	NEG	resolve

'Recently, Mr. Huang from Longgang, Wenzhou, reported to *Zheli Bang* that on the morning of August 31, his sedan, which was parked downstairs at his home, was involved in a 'bizarre' accident. However, the issue of subsequent compensation for the vehicle has yet to be resolved'.

(44) 通过现场的公共视频可以看到，撞向黄先生车的是一辆“蔚来”电动汽车，车辆在拐弯后，突然加速向着路边冲去。

tōngguò	*xiànchǎng*	*de*	*gōnggòng*	*shìpín*
through	scene	SP	public	video
kěyǐ	*kàndào*	*zhuàng*	*xiàng*	*Huáng*
can	see	crash	towards	Huang
xiānshēng	*chē*	*de*	*shì*	*yī*
Mr.	car	SP	be	one
liàng	*“Wèilái”*	*diàndòng*	*qìchē*	*chēliàng*
CLF	“NIO”	electric	car	vehicle
zài	*guǎiwān*	*hòu*	*tūrán*	*jiāsù*
at	turn	after	suddenly	accelerate
xiàngzhe	*lùbiān*	*chōng*	*qù*	
towards	*roadside*	*rush*	*go*	

‘Footage from a public surveillance camera at the scene shows that the vehicle that crashed into Mr. Huang’s car was a NIO electric car. After making a turn, the vehicle suddenly accelerated and veered toward the roadside’.

Based on the corpus examined, Chinese and Italian exhibit contrasting tendencies with respect to faithful versus unfaithful anaphora. In sum, Italian tends to favor unfaithful anaphoric expressions, frequently employing a wide range of synonyms, hypernyms, or semantically related expressions. This flexibility allows for variation in reference while maintaining discourse cohesion. Chinese, by contrast, shows a marked preference for faithful anaphora, where the anaphoric element closely mirrors its antecedent – often through repetition of the lexical head alone – thereby reinforcing cohesion through lexical continuity. These cross-linguistic tendencies reflect distinct discourse strategies and underscore the influence of language-specific constraints on referential coherence.

4.2.4 Associative Anaphora

Associative anaphora (Korzen 2023; Hawkins 1978; Löbner 1998) refers to entities that are associated with, but not identical to, their antecedents. This phenomenon – also known as bridging[7] plays a crucial role in discourse comprehension. Clark (1977) argues that bridging relies on background assumptions or inference, often involving implicatures. An entity introduced into a discourse – the anaphor – stands in a particular relation to some previously mentioned discourse entity, as in *I looked into the room. The ceiling was very high*. This relation is not explicitly stated by linguistic means. Yet it is an essential part of the discourse content because necessary for successfully interpreting a discourse (Irmer 2011, 223). The relationship can take various semantic forms, as illustrated below with examples from the cooking domain. Bridging also presents a significant challenge in natural language processing (NLP), as parsers and computational tools often struggle to establish these associative links (Fang, Baldwin, Verspoor 2022). Bridging cross-reference anaphora is most commonly encoded using definite NPs (Huang 2000a). However, other forms are also possible, including numerals, quantifiers, measure words, and indefinite adjectives or pronouns (highlighted in bold below):

(45) Lisa opened **the picnic basket**. **The sandwiches** were perfectly crafted. (Setting-element)

(46) Cut **the lemon** and take **the seeds** out. (Whole-part/necessary parts)

(47) Slice **the bread** into **pieces**. (Partitive)

(48) Add all **dry ingredients**, except **yeast**. (Set-membership)

(49) Add all **fruits**, except **strawberries**. (Hypernym-hyponym)

(50) Bake for 15 minutes, then turn **the oven** off and let it cook. (Necessary role/tool)

(51) Remove sinew and membrane from the chicken breast, chop, or mince very finely to almost a paste. Mix with **½ a tablespoon of the corn flour**, milk […] With one chopstick pick up **a little of the chicken mixture** and drop that in the stock, go on until **all the chicken** is used up and you have **lots of chicken bits/balls** pea size. (Fang et al. 2022)

One way to formally account for this "associability" is through the four qualia roles introduced in Pustejovsky's Generative Lexicon (1995, 67, 85). Pustejovsky conceptualises qualia structure as an inherent semantic property of lexemes, consisting of four roles: the formal role (e.g. attributes like shape, size, position, and colour), the constitutive role (e.g. elements such as material, weight, components, and content), the agentive role (e.g. the entities or factors involved in the creation of the object), and the telic role (e.g. the object's purpose or function). These semantic roles offer a framework for understanding associative anaphora (Korzen 2023). For example, after introducing a car, one can easily refer to associated entities such as its parts (wheels, doors), properties (colour, size), individuals involved with it (e.g. a mechanic), or its purpose (e.g. a trip).

---

[7] Asher, Lascarides 1998; Bos, Buitelaar, Mineur 1995; Clark 1977; Huang 2000a.

4.2.5 Case Study 3: Partial Coreference in Culinary Texts

Morbiato (2024) conducted a corpus analysis of evolving anaphors in culinary texts to compare the use of different anaphoric forms (e.g. full NPs over pronouns or zero forms) in Chinese, English, and Italian, based on the corpus also analysed in Morbiato, Cani 2024. The data consists of recipe texts drawn from user-generated blogs, describing similar recipes, i.e. the preparation of pumpkin bread rolls, in English, Chinese, and Italian. Partial coreference can be observed in both languages, although with different means. Here are some text excerpts:

(52) (I.14a)

Mett-ete	in	una	ciotola	il	
put-IMP.2PL	in	a	bowl	the	
quantitativo	di	polpa	di	zucca	(120gr),
quantity	of	pulp	of	pumpkin	
(b)	la	Pasta Madre	(a	pezzetti),	lo
	the	sourdough	in	pieces	the
zucchero	e	parte	d-ell'	acqua	
sugar	and	part	of-the	water	

'Place 120 grams of pumpkin puree, the sourdough cut into pieces, sugar, and a portion of the water (e.g. 100 grams) in the bowl. Mix everything thoroughly to achieve a creamy texture'.

(I.15a)

Aggiung-ete	la	farina	e	gradualmente	
add-IMP.2PL	the	flour	and	gradually	
inseri-te	la	restante	acqua	(b)	può
insert-IMP.2PL	the	remaining	water		might
darsi	che	non	serva	tutt-a	
be	that	NEG	need	all-F.SG	

'Add the flour and gradually incorporate the remaining water; you might not need all of it'.

(I.17a)

Pirl-ate	un	po'	l'	impasto
spin-IMP.2PL	a	bit	the	dough

a	formare	un-a	bell-a	palla	lisci-a
to	form	a-F	nice-F	ball.F	smooth-F
(b)	ole-ate	Ø	leggermente	e	
	oil-IMP.2PL	Ø	slightly	and	
mett-ete-la	in	una	ciotola		
put-IMP.2PL-F.3SG	in	a	bowl		

‘Give the dough a gentle spin to form a nice, smooth ball. Lightly coat with oil, place in a bowl’.

(I.18b)

stend-ete	con	il	mattarello	l’
spread-IMP.2PL	with	the	rolling.pin	the
impasto	sgonfiando-lo	bene	ma	delicatamente,
dough	deflating-3M.SG	well	but	gently

‘Take the dough out of the fridge and allow it to come to room temperature. Roll out the dough with a rolling pin, gently deflating it, and stretch it until it reaches around 5 mm in height’.

In Italian, both measure words and expressions (e.g. *un dito d’acqua* ‘a finger of water’, *parte dell’acqua* ‘part of the water’) and clitic partitive pronouns occur, as e.g. ‘ne’, followed by the quantity *120gr.* ‘use 120 g (of it)’. Hypernyms and collective nouns/expressions are used as well, such as *tutto* ‘everything in (40) (14c) to denote all the aforementioned ingredients.

(53) (C.9) (a) 南瓜泥$_g$, 用南瓜$_j$去皮切块Ø蒸熟, (b) 然后Ø入榨汁机Ø榨成汁,

nánguā	*ní$_g$*	*yòng*	*nánguā*	*qù*	*pí*
pumpkin	puree$_g$ TOP	use	pumpkin	remove	skin
qiē	*kuài*	*Ø*	*zhēng-shú*	*ránhòu*	*Ø*
cut	pieces	Ø	steam-cook	then	Ø
rù	*zhàzhījī*	*Ø*	*zhà-chéng*	*zhī*	
insert	juicer	Ø	press-become	juice	

‘For the pumpkin puree, peel the pumpkin, cut into pieces and steam, then put in the juicer and juice’.

(C.10) 然后除黄油外所有材料放入厨师机内。

ránhòu	*chú*	*huángyóu*	*wài*	*suǒyǒu*	*cáiliào$_m$*
then	except	butter	out	all	ingredient
fàng-rù	*chúshījī*	*nèi*			
put-insert	mixer	in			

‘Place all the ingredients in the mixer, excluding the butter’.

(C.15) 然后Ø整理成面团, 入不锈钢碗里。

ránhòu	*Ø*	*zhěnglǐ-chéng*	*miàntuán*	*rù*	*bùxiù*
then	Ø	arrange-become	dough	insert	stainless
gāng	*wǎn*	*lǐ*			

steel bowl in
‘Next, knead the mixture until it forms a dough and place it in a stainless-steel bowl’.

In Chinese, similarly, whole-part relationships are present, as e.g. *nánguā* ‘pumpkin’ and *pí* ‘skin’, as well as quantifiers *suǒyǒu cáiliào* ‘all ingredients’. Similar considerations hold for English (see Morbiato 2024).

However, in this type of texts a different kind of reference is striking observable, and namely expressions that refer to an entity while also indicating a change in its properties or attributes over time. These are, in the literature, referred to as evolving anaphors (Asher 2000; Korzen 2009; Oguz et al. 2022) and are common in procedural texts, as e.g. recipes, where they typically refer to an ingredient or component of a recipe that has undergone some transformation.

Evolving anaphors are observable both the Italian and Chinese corpus, with similar referential mechanisms. For Italian, one instance is (I.15)-(18), where the referent undergoes several changes of states, while being referred to in an array of different ways, e.g. with the same NP (e.g. *l’impasto*, ‘the dough’ in (17a) and (18.b)), clitic pronouns, zero anaphors (I.17b), or different NPs (e.g. *un composto cremoso* in (I.14) *una bella palla liscia* ‘a nice, smooth ball’ in (I.17). In Chinese, evolving anaphors abound and are mainly encoded through zeroes, and can be observed in the many passages in example (53) above. Morbiato (2024), along with Korzen (2009) extensively examined this phenomenon, which will therefore not be further discussed here. However, it seems necessary to incorporate this type of anaphoric relationship into the overall framework outlined above, distinguishing it from bridging or associative anaphora. This distinction is further justified by the fact that such a relationship is not easily accounted for within the Qualia Structure theory discussed above, as it involves the evolution of an entity over time, potentially reaching ontological thresholds. For instance, this can include transformations in physical state, such as a solid turning into a liquid, as exemplified in (42), in Ø *rù zhàzhījī* Ø *zhà-chéng zhī* ‘Put in the juicer and juice (lit. to become juice)’.

4.2.6 Resumptive Anaphors

Resumptive anaphors encapsulate not just a noun phrase but an entire preceding clause, sentence, or a substantial portion thereof, thereby foregrounding the topical content of the discourse. In other words, this type of anaphora does not establish reference to a discrete nominal expression but instead targets a broader propositional or discursive segment – typically to emphasise, reiterate, or evaluate a previously stated idea. Consequently, the antecedent is interpreted as a third-order entity, which demands a particularly “strong” anaphoric form (Korzen 2023). According to Korzen, in Italian (and similarly in English), when the anaphor is pronominal, it is most often a demonstrative pronoun; when it takes the form of a noun phrase, it is typically a demonstrative-modified NP.

(54) (Korzen 2019, 80)

Nel cortile del nostro Parlamento è in corso – forse ancora – un’indegna manifestazione […]. Penso, quindi, che questo problema vada segnalato […].

‘In the forecourt of our Parliament a shameful demonstration is – probably still – in progress […]. I therefore think that this problem should be reported […]’.

(55) (Korzen 2014, 258)

Evidentemente, la questione dell’immigrazione è il fatto principale dietro il quale si nascondono poi tutte le cause di razzismo e di xenofobia, ed è su questo che deve rivolgersi tutta l’attenzione dei governi europei.

‘Clearly, the issue of immigration is the main thing behind which all the causes of racism and xenophobia are hidden, and it is to this that all the attention of European governments must be directed’.

The Chinese forms encoding this type of anaphoric relationship are not dissimilar. Like Italian and English, Chinese also uses demonstratives. Native speakers suggest that the possible translations of (54) and (55) above can be as follows, demonstrating a series of resumptive means that are similar, such as *zhè.ge wèntí* ‘this problem’ in (56) and *zhè zhèng-shì… zhòngdiǎn* ‘this is the key point’ in (57).

(56) 在议会的前院，可能仍在进行一场令人羞耻的示威活动[…]。因此，我认为这个问题应该被报道[…]。

zài	*yìhuì*	*de*	*qiányuàn*	*kěnéng*	*réng*
at	forecourt	sp	before	maybe	still
zài	*jìnxíng*	*yī*	*chǎng*	*lìng*	*rén*
be.at	hold	one	CLF	lead	people
xiūchǐ	*de*	*shìwēi huódòng*	*yīncǐ*	*wǒ*	*rènwéi*
shame	SP	*demonstra-tion*	therefore	1SG	think
zhè	*ge*	*wèntí*	*yīnggāi*	*bèi*	*bàodào*
this	CLF	problem	should	BEI	report

(57) 显然，移民问题是隐藏种族主义和仇外心理根源的核心问题，这正是欧洲各国政府必须全力关注的重点。

xiǎnrán	*yímín*	*wèntí*	*shì*	*yǐncáng*
obviously	immigration	problem	be	conceal
zhǒngzú.zhǔyì	*hé*	*chóuwài*	*xīnlǐ*	*gēnyuán*

racism	and	xenophobia	psychol.	root
de	*héxīn*	*wèntí*	*zhè*	*zhèng-shì*
SP	core	problem	this	right-be
Ōuzhōu	*gè-guó*	*zhèngfǔ*	*bìxū*	*quánlì*
Europe	each-country	government	must	full-effort
guānzhù	*de*	*zhòngdiǎn*		
focus	SP	key.point		

5 Conclusions

This paper offered an overview of the different types of referential relationships, with a particular focus on exophora and anaphora, and detailed the linguistic forms used to encode these relationships in Chinese as compared to Italian and English. Korzen's (2023) work on referential relationships in Italian served as a valuable baseline, allowing the identification of a range of interesting similarities and differences with Chinese and English, although with some remarks and additions as detailed below.

The corpus data showed encoding strategies in the three languages, both with similarities and with some nuanced differences as outlined in the various sections. With regard to exophoric forms, all languages exhibited full noun phrases with demonstratives or possessives, headless demonstrative NPs, pronouns, and zero forms. Despite these similarities, some differences emerged, including the cooccurrence of articles and possessives in Italian, absent in English and Chinese; argument omission is allowed in Italian and Chinese, though for distinct reasons and with different licensing conditions; and different means to express partitivity (with proforms in Italian/English vs. zeroes in Chinese). Finally, Italian distinguishes between weak and strong pronouns – a contrast absent in the other two languages. With regard to generic reference, data show that in addition to bare NPs in Chinese and definite-article-marked NPs in Italian/English, indefinite NPs may also carry generic reference in all three languages.

With regard to anaphora, similar encoding strategies are found, although differences exist in the distribution and licensing conditions of pronouns versus zero forms, with substantial variation across the three languages. Substantial differences, crucially, emerged in the use of synonyms/hypernyms and hyponyms in coreference, both exophoric and anaphoric. Notably, exophoric coreference and unfaithful anaphora, both involving synonyms, are attested in Italian as strategies to avoid repetition and provide additional information. However, these phenomena were not (or only extremely marginally) observed in Chinese, which tends to favour NP repetition, faithful anaphora, or the use of proforms. This holds across the various text types and genres examined in the case studies' corpus analysis. In particular, unfaithful anaphora (i.e. an NP with a lexically different head, as in "The US president" / "the leader") is not used in Chinese and is only marginally observable. Crucially, however, a distinct pattern emerged that fulfills a similar function, involving a pronoun modified by adjectives or relative clauses. Though the head is pronominal rather than nominal, its communicative function parallels that of unfaithful anaphora: it serves to provide additional (objective or subjective) information about a previously mentioned referent and has a similar information structure. This pattern, while structurally distinct yet functionally akin to unfaithful anaphora, warrants further corpus-based investigation – both to determine its frequency and distribution across genres and registers, and to gain deeper insight into its communicative and discourse-functional roles in comparison to unfaithful anaphora in other languages.

(58) 为什么没有人帮帮可怜的他们? (BCC)

wèishénme	*méi*	*yǒu*	*rén*	*bāng-bāng*	*kělián*	*de*	*tāmen*
why	neg	have	person	help-help	pitiable	SP	they

'Why doesn't anyone help 'the pitiable them'?'

Regarding Korzen's framework of referential relationships, it was found to be generally useful, but a few important cases deserve further elaboration. First, specific reference – being identifiable to the speaker but not to the interlocutor – was not included, despite its key role in introducing new discourse referents through indefinite NPs. Moreover, it seems necessary to incorporate evolving anaphora – i.e. anaphoric expressions referring to entities whose properties change over time – into the overall mapping of referential relationships, distinguishing it from bridging or associative anaphora. This distinction is further justified by the difficulty of accounting for such relationships within existing frameworks like the Qualia Structure theory **[fig. 4]**.

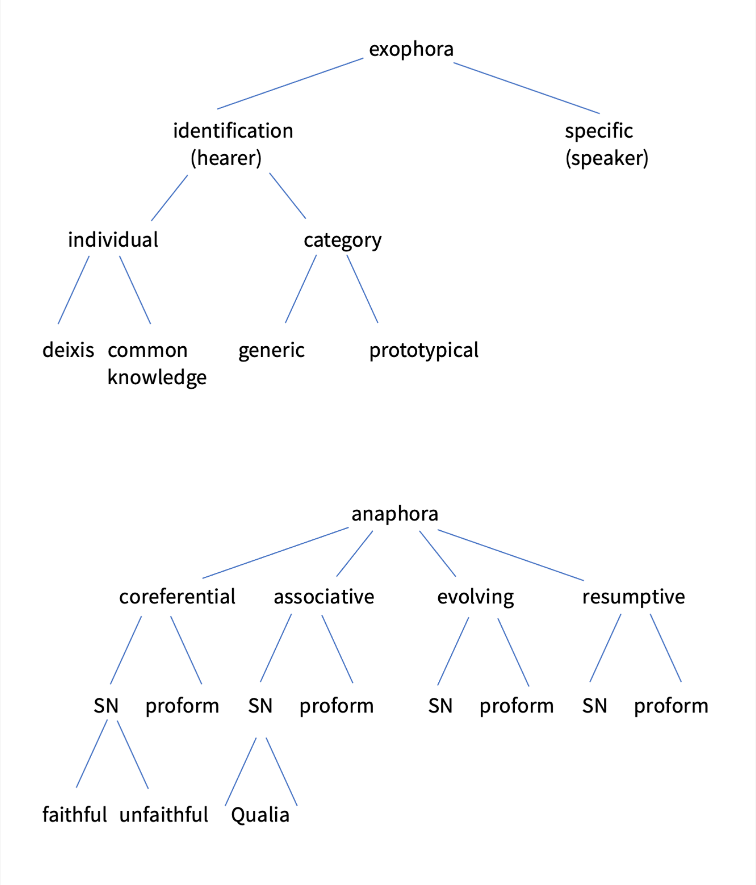
exophora
identification (hearer)
specific (speaker)
individual
category
deixis
common knowledge
generic
prototypical
anaphora
coreferential
associative
evolving
resumptive
SN
proform
SN
proform
SN
proform
SN
proform
faithful
unfaithful
Qualia

Figure 4 Mapping of exophoric and anaphoric relationships

Overall, the corpus-based comparison proved highly informative, particularly through the construction of comparable datasets within specific genres, which enabled more controlled and meaningful cross-linguistic contrasts. Based on preliminary qualitative and quantitative observations from relatively small corpora, the study does not aim for full representativeness. Nonetheless, it provides valuable initial insights that highlight the potential of corpus-driven research. These findings emphasise the central role of such approaches in revealing both universal tendencies and language-specific strategies in referential practices, and they may serve as a springboard for more extensive future research – especially through larger, genre-specific corpus investigations. It is hoped that this work will help pave the way for such developments.

Bibliography

Alcón, E.M.P.; Jordà, S. (2008). “Pragmatic Awareness in Second Language Acquisition”. Hornberger, N.H. (ed.), *Encyclopedia of Language and Education*. Boston (MA): Springer US, 1948-59.

Ariel, M. (2009). “Accessibility Theory”. Mey, J.L. (ed.), *A Concise Encyclopaedia of Pragmatics*. Oxford: Elsevier, 1-4.

Asher, N. (2000). “Events, Facts, Propositions, and Evolutive Anaphora”. Varzi, A.; Higginbotham, J.; Pianesi, F. (eds), *Speaking of Events*. Oxford: Oxford University Press, 123-50.

Asher, N.; Lascarides, A. (1998). “Bridging”. *Journal of Semantics*, 15(1), 83-113.

Bos, J.; Buitelaar, P.; Mineur, A. (1995). *Bridging as Coercive Accommodation*. `https://arxiv.org/abs/cmp-lg/9508001`.

Chafe, W.L. (1976). “Givenness, Contrastiveness, Definiteness, Subjects, Topics, and Points of View”. Li, C. (ed.), *Subject and Topic*. New York: Academic Press, 25-55.

Chafe, W.L. (1994). *Discourse, Consciousness, and Time: The Flow and Displacement of Conscious Experience in Speaking and Writing*. Chicago; London: The University of Chicago Press.

Clark, K. (1977). “Bridging”. Johnson-Laird, P.N.; Cathcart Wason, P. (eds), *Thinking: Readings in Cognitive Science*. Cambridge: Cambridge University Press, 411-20.

Comrie, B. (1989). “Some General Properties of Reference-tracking Systems”. Doug, A.; Atkinson, M.; Durand, J.; Grover, C.; Sadler, L. (eds), *Essays on Grammatical Theory and Universal Grammar*. Oxford: Clarendon Press, 37-51.

Cornish, F. (2010). “Anaphora: Text-Based or Discourse-Dependent? Functionalist vs. Formalist Accounts”. *Functions of Language*, 17(2), 207-41.

Cornish, F. (2014). “Discourse Anaphora”. Brown, K. (ed.), *Encyclopedia of Language and Linguistics*. Oxford (RU): Elsevier, 631-8.

Fang, B.; Druckenbrodt, C.; Akhondi, S.A.; Thorne, C.; Baldwin, T.; Verspoor, K. (2022). *Annotation Guideline for RecipeRef: A Corpus for Modeling Anaphora Resolution of the Procedural Text of Recipes*. `https://biaoyanf.github.io/files/papers/reciperef-dataset.pdf`.

Fang, B.; Baldwin, T.; Verspoor, K. (2022). “What Does It Take to Bake a Cake? The RecipeRef Corpus and Anaphora Resolution in Procedural Text”. *Proceedings of the Annual Meeting of the Association for Computational Linguistics*, 3481-95.

Foley, W.A; Van Valin, R.D. (1984). *Functional Syntax and Universal Grammar*. Cambridge: Cambridge University Press.

Givón, T. (1983). *Topic Continuity in Discourse: A Quantitative Cross-Language Study*. John Benjamins Publishing.

Green, K. (2009). “Deixis and Anaphora: Pragmatic Approaches”. Mey, J.L. (ed.), *A Concise Encyclopaedia of Pragmatics*. Oxford: Elsevier, 178-81.

Horn, L.R.; Ward, G. (eds) (2004). *The Handbook of Pragmatics*. Oxford: Blackwell Publishing.

House, J.; Kádár, D.Z. (2021). *Cross-Cultural Pragmatics*. Cambridge: Cambridge University Press.

Huang, C.-T.J. (1984). “On the Distribution and Reference of Empty Pronouns”. *Linguistic inquiry*, 15(4), 531-74.

Huang, Y. (1994). *The Syntax and Pragmatics of Anaphora: A Study with Special Reference to Chinese*. New York: Cambridge University Press.

Huang, Y. (2000a). *Anaphora: A Cross-Linguistic Study*. New York: Oxford University Press.

Huang, Y. (2000b). “Discourse Anaphora: Four Theoretical Models”. *Journal of Pragmatics*, 32(2), 151-76.

Huang, Y. (2015). “Anaphora, Modern”. *Encyclopedia of Chinese Language and Linguistics Online*.

Huang, Y. (2016). “Aspects of Anaphora in Chinese and in Some Germanic, Romance, and Slavic Languages, the ‘Syntactic’ Versus ‘Pragmatic’ Language Typology, and Neo-Gricean Pragmatics BT - Pragmemes and Theories of Language Use”. Allan, K.; Capone, A.; Kecskes, I. (eds), *Pragmemes and Theories of Language Use. Perspectives in Pragmatics, Philosophy & Psychology*, vol 9. Cham: Springer International Publishing, 21-43. `https://doi.org/10.1007/978-3-319-43491-9_2`.

Irmer, M. (2011). *Bridging Inferences: Constraining and Resolving Underspecification in Discourse Interpretation*. Berlin; Boston: De Gruyter.

Korzen, I. (2009). “Tipologia Anaforica: Il Caso Della Cosiddetta ‘Anafora Evolutiva’”. *Studi di grammatica italiana*, XXV, 323-57.

Korzen, I. (2014). “Frequency, Forms and Functions of Cleft Constructions in Romance and Germanic: Contrastive, Corpus-Based Studies”. De Cesare, A.-M. (ed.), *Cleft Sentences. Italian-Danish in Contrast*. Berlin: De Gruyter, 217-75.

Korzen, I. (2019). “Anaphors and Text Structure in Romance and Germanic Languages: Typologies in Comparison”. *Langages*, 214(2), 75-90.

Korzen, I. (2023). “Di Che Cosa Parliamo Quando Siamo Infedeli? Alcune Puntualizzazioni Sulle Valenze Di Rinvio: Anafora, Catafora Ed Esofora”. *Globe: A Journal of Language, Culture and Communication*, 17(December), 106-18.

Lambrecht, K. (1994). *Information Structure and Sentence Form: Topic, Focus, and the Mental Representations of Discourse Referents*. Cambridge: Cambridge University Press.

Levinson, S.C. (1983). *Cambridge Textbooks in Linguistics Pragmatics*. Cambridge: Cambridge University Press.

Levinson, S.C. (2006). “Deixis”. Horn, L.R.; Ward, G. (eds), *The Handbook of Pragmatics*. Wiley, 97-121. `https://doi.org/10.1002/9780470756959.ch5`.

Li, C.N.; Thompson, S.A. (1979). “Third-Person Pronouns and Zero-Anaphora in Chinese Discourse”. Givón, T. (ed.), *Syntax and Semantics: Discourse and Syntax*. New York: New York Academic Press, 311-35.

Mey, J. (ed.) (2009). *A Concise Encyclopaedia of Pragmatics*. Oxford: Elsevier Ltd.

Mitkov, R. (2022). “Anaphora Resolution”. Mitkov, R., (ed.) *The Oxford Handbook of Computational Linguistics*. Oxford; New York: Oxford University Press, 707-29.

Morbiato, A. (2020). “Chinese Sentence-Initial Indefinites: What Corpora Reveal”. Basciano, B.; Gatto, F.; Morbiato, A. (a cura di), *Corpus-Based Research on Chinese Language and Linguistics*. Venice: Edizioni Ca’ Foscari, 59-80. `http://doi.org/10.30687/978-88-6969-406-6/002`.

Morbiato, A. (2024). “Pragmatics as the Secret Ingredient: A Multilingual Study of Reference Tracking and Evolving Anaphors in Culinary Narratives”. Steels, L.; Porzel, R. (eds), *Narrative-Based Understanding of Everyday Activities: An AI Cookbook*. Venice, 121-41. `https://zenodo.org/records/14199174`.

Morbiato, A.; Cani, P. (2024). “The Linguistic Ingredients of Recipes: Aspect, Causation, and Argument Structure”. Steels, L.; Porzel, R. (eds), *Narrative-Based Understanding of Everyday Activities: An AI Cookbook*. Venice, 99-120. `https://zenodo.org/records/14199174`.

Oguz, C.; Kruijff-Korbayova, I.; Vincent, E.; Denis, P.; Van Genabith, J. (2022). “Chop and Change: Anaphora Resolution in Instructional Cooking Videos”. *2nd Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics and the 12th International Joint Conference on Natural Language Processing - Findings of* *the Association for Computational Linguistics: AACL-IJCNLP 2022*, 364-74. `https://aclanthology.org/2022.findings-aacl.34.pdf`.

Prince, E.F. (1981). “Toward a Taxonomy of Given-New Information”. Cole, P. (ed.), *Radical Pragmatics*. New York: Academic Press, 223-55.

Pu, M.; Pu, Q. (2014). “The Topic Chains in Chinese: A Discourse Analysis and Application in Language Teaching”. *International Journal of Linguistics and Communication*, 2(1), 27-44.

Pustejovsky, J. (1995). “The Generative Lexicon”. *Computational Linguistics*, 17(4), 409-41.

Sidnell, J.; Enfield, N. (2017). “Deixis and the Interactional Foundations of Reference”. *The Oxford Handbook of Pragmatics*. Oxford: Oxford University Press, 217-30.

Sun, K. (2020). *The Complexity of Zero Anaphora in Chinese Discourse: An Event-Related Perspective*.

Tao, L.; Healy, F. (1996). “Cognitive Strategies in Discourse Processing: A Comparison of Chinese and English Speakers”. *Journal of Psycholinguistic Research*, 25(6), 597-616.

Tao, L.; Healy, F. (2005). “Zero Anaphora: Transfer of Reference Tracking Strategies from Chinese to English”. *Journal of Psycholinguistic Research*, 34(2), 99-131.

Tu, X. (2010). *A Discourse-Oriented Approach to Automatic Chinese Zero Anaphora Resolution* [PhD dissertation] Boston: Boston University.

Verschueren, J. (2009). “Introduction”. Verschueren, J.; Ostman, J.-O. (eds), *Key Notions for Pragmatics*. Amsterdam; Philadelphia: John Benjamins Publishing Co., 1-27.

Xu, Y. (2015). “Reference Processing and Discourse”. *Encyclopedia of Chinese Language and Linguistics Online*. Leiden: Brill. `https://referenceworks.brill.com/view/entries/ECLO/COM-000015.xml`.