Nominal classification: does it play a role in referent disambiguation?\footnote{Acknowledgements: The support of the AHRC (grants: Combining gender and classifiers in natural language, grant AH/K003194/1 and Lexical splits: a novel perspective on the structure of words, grant AH/N006887/1) is gratefully acknowledged. I wish to thank Matthew Baerman, Oliver Bond, Dunstan Brown, Greville Corbett, Sebastian Fedden, Nathan Hill, Marcin Kilarski, Gunter Senft, Helen Sims-Williams, Ruth Singer and Marc Tang, as well as the two anonymous reviewers, for helpful comments and suggestions. A preliminary version of this paper was presented at the workshop Gender and classifiers: Diachronic and synchronic variation, University of Surrey, Guildford on 28 January 2016 and the 46th Poznań Linguistic Meeting, Adam Mickiewicz University, Poznań on 16 September 2016. I would like to thank both audiences for their helpful comments and the subsequent discussion.}

Abstract

It is often claimed that reference tracking is a key function of nominal classification systems, not least because of the role such systems can play in referent disambiguation. This paper reports on the results of a comparative study of reference in texts from four languages, focussing specifically on the disambiguating function of nominal classification. The results strongly suggest that disambiguation is not a primary function of nominal classification systems. While gender and/or classifiers sometimes contribute to the avoidance of referential conflict, the reality is that the conditions have to be just right – all competing references must be of opposing genders, and those genders must be formally distinct – and this happens with surprisingly low frequency. We are better off viewing disambiguation as a convenient by-product of nominal classification systems that a language can exploit when conditions allow.

Keywords: nominal classification, gender, classifiers, referent, disambiguation, reference tracking, referential conflict.

1. Introduction

Many languages have systems of nominal classification and yet the purpose of these systems is often poorly understood and the subject of some debate. In an attempt to explain the existence of such systems, it is often claimed that they serve a reference tracking function. Foley & Van Valin (1984), for instance, claim that “gender functions as the dominant source of discourse cohesion” and others, including Corbett (1991), Heath (1983: 141), Huang (2000: 8) and Singer (2016), have also claimed that gender can function as a reference tracking device. Similar claims have also been made with regard to the function of nominal classifiers by Aikhenvald (2003: 329), Contini-Morava & Kilarski (2013) and Senft (2008).

Perhaps one of the most salient reasons that nominal classification systems have been claimed to have a reference tracking function is due to their power to disambiguate between
competing referents. For instance, in a language that classifies objects by their shape, if a verb is marked with a ROUND subject classifier an interlocutor knows that the subject of the clause must be an entity that is able to occur with the ROUND classifier, ruling out all other referents that do not fit. The gender-specific pronouns of English also illustrate how gender can serve to disambiguate between referents which would otherwise be ambiguous (compare 1a and 1b). But isolated examples such as these tell us very little about how systems of nominal classification function in discourse more generally, and the degree to which they may be employed in a disambiguating function in reference tracking.

(1) a. Mary and John bumped into each other. He smiled at her.
    b. Mary and Sarah bumped into each other. She smiled at her.

An alternative view is that nominal classification systems serve no reference tracking function, if indeed any function at all. This view is expressed by Trudgill (1999: 140) who states that “...it is not at all clear what gender is for. [...] The function of noun classes and grammatical gender in particular in human languages is actually largely obscure.” Especially in the case of languages which have a small gender system (e.g. the three-gender system of German), there are sound reasons for making this claim; for gender to be of any use in disambiguating between referents, for instance, competing references must belong to opposing genders and the likelihood of this holding is severely limited in such systems (see discussion in Trudgill 1999).

Yet regardless of whether reference tracking (either generally speaking, or at a more specific level, such as referent disambiguation) is a core function of nominal classification systems or simply a complementary (or opportunistic) one, there are countless languages that deal with reference tracking in the absence of any sort of nominal classification.

With this in mind, one point of departure in investigating the role of nominal classification is to compare the referent disambiguation strategies of languages which do not have nominal classification to ones which do possess such a system. This approach can be taken one step further by also looking at languages with different nominal classification systems separately, i.e. investigating how languages with gender, on the one hand, and classifiers, on the other, handle referent disambiguation. (See Dixon 1986, Fedden & Corbett 2017 and Senft 2008 for discussions on the different types of nominal classification systems.)

This paper reports on the findings of such a study: narrative texts from four languages, which differ typologically from each other with regard to nominal classification, were studied in order to ascertain what role, if any, gender and classifiers play in disambiguating references. The four languages were selected in an attempt to cover the typological space. They comprise: Skolt Saami (Finno-Ugric, Finland), a language with no nominal classification; Spanish (Romance), a language with gender; Kilivila (Austronesian, Papua New Guinea), a language with classifiers; and Mian (Trans New Guinea, Papua New Guinea), one of the rare instances of a language displaying both gender and classifiers.

The results of this study strongly suggest that nominal classification plays a very restricted role in disambiguating referents, at least in the languages under investigation. Overwhelmingly, the same factors that serve to correctly disambiguate referents in Skolt Saami (including the use of full NPs, the level of activation of a referent, and semantic and syntactic restrictions) were found to be in play in the other three languages in the study to a similar degree. And in the majority of cases where gender, or a classifier, were identified as having a potential disambiguating function, this co-occurred with other factors which would
almost certainly have successfully disambiguated the referent in the absence of nominal classification.

This paper is structured as follows. Section 2 lays the groundwork for the rest of the paper by defining a number of key terms that will be used, as well as delineating the scope of the study by being clear about what it does not cover. Section 3 provides a detailed account of the methodology used in the study, but begins with a brief summary of some existing methodologies employed in reference tracking studies and a look at a number of issues that were encountered with them. In addressing these issues, the rationale is provided for adopting an innovative approach to the task at hand. Section 4 takes a cursory look at each language of the study in turn, covering Skolt Saami, Spanish, Kilivila, and Mian, respectively. Section 5 presents the results and Section 6 concludes the paper.

2. Terminology and scope

2.1. Terminology

A **referent** is an entity in the physical world, and includes both animates such as *a dog*, *the next-door neighbour*, and *the US president*, and inanimates, such as *my shopping list*, *Mt. Everest*, and *a table*. However, a referent can also be an image in a person’s mind (e.g. *the Loch Ness Monster*) or more abstract concepts such as *the weather*. Huang (2012) defines a referent as “what is referred to by the use of a referring expression”.

A referential device (or referring expression) is “*any linguistic expression that can be used in an utterance to refer to a particular entity or set of entities in the external world*” (Huang 2012) (note that here, ‘external world’ refers to that which is outside the linguistic world and thus includes referents in a person’s mind). This includes proper names and noun phrases which may spell out the identity of a referent, but also other forms of linguistic expression, such as free pronouns, bound pronouns, demonstratives, and zero reference, all of which are known as reduced referential devices (Kibrik 2011). The decision to use one referential device over another, say a pronoun in place of a noun, is referred to as referential choice.

The link between a referential device and a referent is known as **reference**, which Huang defines as “*the relationship between a linguistic expression and an entity, activity, property, relationship, etc. in the external world, to which it is used to refer*”. In many ways, then, reference tracking, which Huang (2012) defines as “*keeping track of the entities referred to in an ongoing discourse*”, might be better thought of as referent tracking, since it is the identity of the referent that matters.

These definitions capture a crucial fact about referents: that is, while referents are talked about in discourse, they themselves do not occur in discourse. It is important to recognise this separation between a referent in the external world on the one hand, and references in the linguistic world on the other, and to understand the one-to-many relationship between a referent and references to it. For example, Elizabeth Alexandra Mary Windsor as a person in the physical world is a single referent, but a plethora of references may be used in talking about her including *she*, Ø [zero reference], *Her Majesty*, *The Queen*, *Elizabeth II*, or, in addressing her, *Ma’am*, *mum*, *grandma*, and so on, depending on who is talking to her.
One of the potential problems with reduced referential devices is that, by their very nature, they do not spell out the identity of the referent they are referring to. This can result in a referential conflict or referential ambiguity arising, where “there is more than one possible candidate competing for the referent of a referring or an anaphoric expression” (Huang 2012). In reality, however, such conflicts are surprisingly rare. This is because there are many factors in play which serve to eliminate (or reduce) the potential for a referential conflict, such as semantics, person/number, animacy, switch-reference, and context. These factors are collectively referred to as referential aids since they can help to correctly identify a referent.

Inasmuch as gender helps to eliminate referential conflict, then, it too can be thought of as a referential aid. This is because gender marking itself does not refer to an entity (and hence cannot be considered a referential device), but it may assist in disambiguating between referents of opposing genders. In this paper, classifiers are likewise considered to be a further type of referential aid, although it should be noted that in some languages classifiers can also be used to refer (see, for example, Seifart (2005) on Miraña) and so the referential device vs. referential aid distinction is less clear-cut when it comes to classifiers.

2.2. Scope of the study

This paper is concerned only with the issue of referent disambiguation in the context of discourse (i.e. across clauses). By way of a simple example comprising only two clauses, consider the sentence “after the leopard had been chasing the gazelle for five kilometres, it began to tire”. In the adverbial clause, two referents – the leopard and the gazelle – are introduced. In the main clause, reference is made to one of these referents with the pronoun it (a reduced referential device). In the case of this sentence, however, we are faced with a referential ambiguity, and it is unclear which animal began to tire. Our real-world knowledge of how a leopard catches its prey may lead us to conclude that it was the gazelle that tired (in which case real-world knowledge would be serving as a referential aid), but in reality we cannot be sure of the correct interpretation here. In much longer stretches of discourse, with many more participants on the scene, and much more going on, the task of keeping track of referents becomes even more crucial.

This paper does not purport to address other claimed functions of nominal classification, whether they be other discourse functions, such as ‘reference management’ (where the presence/absence or position of (predominantly) classifiers is used to indicate definiteness, specificity or referentiality) and ‘re-presentation of referents’, or semantic functions, such as expansion of the lexicon or individuation (see Contini-Morava & Kilarski 2013 for a discussion of these and other functions). Nor does it deal with the syntactic functions of gender of (i) encoding grammatical relations, or (ii) assigning the correct interpretation to an otherwise ambiguous constituent. These last two functions are important to bear in mind, since they are functions of gender which are observed in Mian and Spanish, respectively.

In Mian, subjects (and objects, too, in the case of seven verbs) are encoded by gender-marked verbal affixes. In a transitive clause, then, provided the two participants differ with regard to gender, the grammatical role of each argument is easily recoverable, regardless of word order. This is exemplified in (2) (Fedden 2011: 363) where, regardless of the order the two arguments appear in, the meaning remains unchanged since the verbal affixes in both cases indicate that the subject is masculine and the object is feminine. (Correctly determining the grammatical role of an argument is a far more involved process than this simplistic
example might suggest, however, entailing such factors as word order and information structure; of course, Mian must rely on another strategy when both arguments of a transitive clause are of the same gender.)

(2) a. \( \text{naka}=\ e \quad \text{unáng}=\ o \quad \text{wa-têm}'-\Ø-e=be \)
    
\[
\begin{array}{llllll}
\text{man}=&\text{SG.M} & \text{woman}=&\text{SG.F} & \text{3SG.F.OBJ}=\text{see}.\text{PFV}-\text{3SG.M.SBJ}=\text{DECL}
\end{array}
\]

‘The man has seen the woman.’

b. \( \text{unáng}=\ o \quad \text{naka}=\ e \quad \text{wa-têm}'-\Ø-e=be \)
    
\[
\begin{array}{llllll}
\text{woman}=&\text{SG.F} & \text{man}=&\text{SG.M} & \text{3SG.F.OBJ}=\text{see}.\text{PFV}-\text{3SG.M.SBJ}=\text{DECL}
\end{array}
\]

‘The man has seen the woman.’

More importantly, however, this function of gender is restricted to the level of the clause and does not relate directly to the topic of reference tracking. This paper is not concerned with the role gender might play in correctly identifying the respective grammatical roles of two or more references in a given clause, but rather the role gender might play in disambiguating the identity of a referent at any given point in discourse. This notion is illustrated in (3). As John is the only male referent, there is no doubt that John is the subject in (3b), since English has strict SVO word order and there is a masculine pronoun in subject position. But the key question here is not ‘who is the subject?’ but rather ‘who does he refer to?’. In this example, there is only one male participant on the scene and so he must refer to John. The same cannot be said about the feminine pronoun her in (3b), as there are two female participants on the scene and hence the identity of the referent is unclear.

(3) a. John loved Sally, but he couldn’t get Mary out of his head.
   b. He decided he had no choice but to tell her.

Turning now to Spanish, consider example (4). In (4), the adjective asustada ‘scared’ is marked for feminine gender and, as a result, it is clear that it is an attributive modifier of the object (la mujer ‘the woman’). In (5), the same adjective signals masculine gender, and so here the adjective is interpreted as making a predication about the subject. In (6), however, both the subject and the object are of the same gender and thus the interpretation is ambiguous.

(4) \( \text{el} \quad \text{hombre} \quad \text{miró} \quad \text{a} \quad \text{[la} \quad \text{mujer} \quad \text{asustada]} \)
    
\[
\begin{array}{llllll}
\text{DEF.M} & \text{man(M)} & \text{look.PST.3SG} & \text{ANIM} & \text{[DEF.F} & \text{woman(F)} & \text{scared.F} \text{]} \\
\end{array}
\]

‘The man looked at the scared woman.’

(5) \( \text{el} \quad \text{hombre} \quad \text{miró} \quad \text{a} \quad \text{[la} \quad \text{mujer} \quad \text{]} \quad \text{asustado} \)
    
\[
\begin{array}{llllll}
\text{DEF.M} & \text{man(M)} & \text{look.PST.3SG} & \text{ANIM} & \text{[DEF.F} & \text{woman(F)} & \text{scared.M} \\
\end{array}
\]

‘The man looked at the woman, scared.’

(6) \( \text{el} \quad \text{hombre} \quad \text{miró} \quad \text{al} \quad \text{niño} \quad \text{asustado} \)
    
\[
\begin{array}{llllll}
\text{DEF.M} & \text{man(M)} & \text{look.PST.3SG} & \text{ANIM.DEF.M} & \text{boy(M)} & \text{scared.M} \\
\end{array}
\]

‘The man looked at the scared boy / the man looked at the boy, scared.’
The purpose of showing these examples is to highlight the fact that, even though gender does serve to disambiguate in such circumstances, the study reported on in this paper is concerned with keeping track of, say, *el hombre* ‘the man’, regardless of whether he is referred to with a NP, a pronoun or even a zero; it is not concerned with determining which of the participants is being modified.

3. Methods of investigating reference tracking

A detailed account of the methodology used in this study is presented in section 3.3. Prior to this, however, in section 3.1, is a brief summary of some of the early methods that were developed to investigate reference tracking. These methods were found to have a number of limitations and be problematic in several ways, as explained in section 3.2.

3.1. Existing methods of investigating reference tracking

One of the earliest methods for studying reference tracking was proposed by Givón (1983) in his volume on topic continuity in discourse. This volume is a quantitative cross-linguistic study of reference tracking and the contributing authors all employ Givón’s methodology. His method attempts to quantify participant reference by looking at three ‘discourse measurements’, namely referential distance, potential interference and persistence.

Referential distance (or ‘look-back’) refers to the gap between a reference in discourse and the immediately preceding reference to the same referent. This measurement is expressed in terms of number of clauses; for example, a given reference is assigned a score of ‘4’ if the referent is last mentioned four clauses back in the discourse. Givón proposes an arbitrary maximum score of 20, which is assigned to a reference when no previous mention of the same referent is found in the preceding 20 clauses. It should be noted that ‘last mentioned’ does not necessarily mean there must be an overt reference in the discourse, but it can also include zero reference to a referent which is an argument of the predicate.

Potential interference (or ‘ambiguity’) assesses whether or not there are references in the ‘immediately preceding register’ which relate to other referents and which could therefore have a disruptive effect on topic identification in a clause. Givón defines ‘immediately preceding register’ as being between 1–5 clauses back, with most contributions in the volume settling on three clauses. The rationale for this measurement is that if the identification of a topic is unambiguous for a given number of preceding clauses, then the presence of other potentially interfering references further back in the discourse should not have a significant negative impact on a hearer’s ability to correctly identify a referent. In most of the studies in Givón (1983) the score assigned for potential interference is a binary one: if no potentially interfering references are found in the specified preceding register a score of ‘0’ is assigned; if one or more potentially interfering references are found, then a score of ‘1’ is given.

A reference is treated as instantiating potential interference only if it is semantically compatible with the clause in question. Thus, the referential device *it* in our earlier example ‘after the leopard had been chasing the gazelle for five kilometres, it began to tire’ would be assigned a score of 1, since one cannot be entirely sure whether it was the leopard or the gazelle which began to tire. However, the same reference *it* in ‘after the leopard had eaten
the gazelle, it ran to the watering hole’ would be assigned a score of 0 since on semantic grounds it is not possible for it to refer to the gazelle, since we know it is dead.

The third discourse measurement, persistence (or ‘decay’) looks at how long a referent continues an uninterrupted presence as a semantic argument of the clause in the subsequent discourse, and so differs from the other two measures in that it does not look back at the preceding discourse. Persistence is expressed in terms of the number of clauses ahead in which reference to the same referent is made. As soon as a clause which does not mention the referent is encountered the counting stops, thus a minimal score of ‘0’ is assigned if the following clause does not mention the referent; there is no maximum score.

While Givon’s method is a useful start in the quantification of references in discourse, it only considers participants and references to them on a timeline and does not consider other factors of discourse organisation. Tomlin (1987) identifies this as a potential shortcoming and proposes that the overall structure of discourse should also be taken into account by considering ‘episodes’ and ‘paragraphs’ of a discourse (i.e. points in the discourse where there is a shift in topic or focus). Tomlin hypothesises that a full noun phrase is used to reinstate referents across an ‘episode boundary’, even in cases where the same referent is mentioned in the preceding clause, while a pronoun (or other reduced referential device) is used to maintain reference within an episode.

Dooley & Levinsohn (2001) also propose a method that moves away from a simple linear view of reference, focussing instead on the various discourse contexts in which references appear. Each reference is assigned to one of eight contexts (S1, S2, S3, S4, N1, N2, N3, N4), which indicates (i) whether the reference in question is a subject or non-subject (S = subject, N = non-subject), and (ii) what role the referent played when it was last mentioned. Discourse context ‘S1’ indicates that the subject is the same as in the previous clause or sentence, ‘S2’ indicates that the subject was the addressee of a speech reported in the previous sentence, ‘S3’ that the subject was involved in the previous sentence in a non-subject role, and ‘S4’ other changes of subject. Discourse context ‘N1’ indicates that the referent occupies the same non-subject role as in the previous sentence, ‘N2’ that the addressee of a reported speech was the subject (speaker) of a speech reported in the previous sentence, ‘N3’ indicates that the referent was involved in the previous sentence in a different role than that covered by N2, and ‘N4’ is assigned to other non-subject references than those covered by N1–N3.

Once each reference has been assigned to a discourse context, ‘default encodings’ (i.e. the referential device that is used by default in a given discourse context) can be identified. For instance, one might find that eighty per cent of references in discourse context N1 are encoded with a pronoun – thus the default encoding associated with N1 is a pronoun by virtue of the high percentage of references encoded this way. References which do not align with the default encoding are therefore considered to be marked and an explanation for this is sought for each reference it applies to.

Each of these three methods has its own strengths and the methods proposed by Givón and Dooley & Levinsohn, in particular, have been employed in a number of reference tracking studies (e.g. Clark 2000, Córdova 2009, Dieterman 2002, Edwards 2011, Forker 2009, Isaac 2007, MacLeod 2012, Runge 2006, Stirling 2008). However, despite some strengths, these methods were not found to be particularly suited to investigating the potential disambiguating role of nominal classification systems. This was in part due to the fact that they were developed for a distinct purpose, but also because of a number of shortcomings that they were found to have. These limitations will be outlined in the following section.
3.2. Limitations of existing methods

The three methods discussed in section 3.1 are fundamentally concerned with understanding how references are encoded in discourse (i.e. what referential device is used) and what factors drive a speaker’s choice of one referential device over another. Givón’s method, for example, presupposes a correlation between the choice of referential device and the accessibility of a referent, which in turn is determined by its referential distance and potential interference scores. Tomlin’s method, likewise, focuses on understanding when and why different referential devices are used, although it goes one step further in correlating this with discourse structure. Dooley & Levinsohn’s method is built around the concept of default encodings and predicts the referential device that is expected for a given discourse context.

While referential choice is an important factor in any study of reference tracking, it is not the primary concern of this paper. Instead, in this paper the focus shifts to looking at whether there is any evidence to support the notion that systems of nominal classification serve a disambiguating function. Rather than asking the question ‘What are the factors that influence the choice of referential device in this language?’ the focus of this paper shifts to asking the question ‘Regardless of the referential device in use, to what extent do gender/classifiers aid in disambiguating referents?’. Another question we might ask is ‘Do languages with gender, classifiers, or both, suffer less from referential conflicts than languages without nominal classification?’ But even though the emphasis is on the role that nominal classification may or may not play in disambiguating referents, this does not imply ignoring referential choice completely; instead, we might reframe the question about referential devices to ask ‘Do languages with gender, classifiers, or both, allow for a lower proportion of NPs as a result?’ (for more on referential density, see Bickel 2003).

This shift in focus means that the methods discussed in section 3.1 do not lend themselves particularly well to the study reported on in this paper; they suffer from a number of shortcomings, which made it necessary to devise a new methodology. These shortcomings revolve around the fact that all of the methods discussed involve counting clauses (e.g. referential distance = number of clauses to the antecedent of a reference; persistence = number of clauses a referent continues an uninterrupted presence) and problems quickly arise with complex clauses and possessive NPs.

Firstly, it is unclear how previous methods would handle adverbial clauses or other subordinate clauses. Should they be treated in the same way as main clauses and included in any strategies which involve counting, or should they be ignored? Clark (2012: 7) observes that many of the studies in Givón (1983) had to adapt their strategies in this regard and notes that “complement clauses, as well as clauses that represented direct speech quotations, were not considered to be ‘gaps’ when the referent did not appear in the clause, but they were counted as occurrences when the referent did appear in the clause” [underlining in the original]. For the authors of the various studies in that volume to have adopted such an ad hoc approach to counting clauses is a strong indicator that the requirement to chunk discourse into clauses should either be abandoned or else fine-tuned. It should go without saying, though, that ignoring such clauses is not a viable option, since they often contain crucial information about referents. This is clearly exemplified in (7), where the identity of the referents mentioned in the main clause by means of the pronouns he and them, are only recoverable from the adverbial clause.
After the man had washed the dishes, he placed them in the drying rack.

Secondly, there is no clear indication of how previous methods deal with nested clauses, in particular relative clauses. In a similar fashion to example (7), references in relative clauses can hold the necessary information to correctly identify a referent mentioned in a main clause, so they cannot be disregarded. In (8), for instance, the participant dog is mentioned in a relative clause modifying the participant the boy, but in the following clause the dog is referred to with only a pronoun. At the same time, how should one count clauses where two or more arguments of a single clause are split by an intervening relative clause?

The boy, whose dog had run away, started crying. He’d only bought it that day.

And thirdly, the methods discussed say nothing about how to deal with referents which modify other referents, i.e. in the context of possession and part–whole relations. The two examples in (9) demonstrate that a possessor and its possessee can, once established, both be referred to by reduced referential devices in the discourse which follows. In a language with gender, such as Spanish, both a possessor and a possessee can control agreement in subsequent discourse (10).

a. The [children’s] mother asked [them] to stop pestering [her]

b. A bee landed on the [girl’s] head, so [she] gave [it] a quick shake

La madre del niño lo llevó al colegio

‘The [boy’s] mother took [him] to school’

La madre del niño pidió que la llamara

‘The [boy’s] mother asked that (he) call [her]’

While these methods may suffer from other shortcomings, those described above are the ones of greatest significance for the purpose of this study. The methodology designed for this study, which attempts to overcome these issues, is outlined in section 3.3.

3.3. Methodology

In order to address the issues outlined in section 3.2, the approach of chunking discourse into clauses was disregarded altogether. Instead, each and every reference in a discourse was identified and analysed in the same manner, regardless of whether it was, for example, a subject in a main clause, an oblique argument in an adverbial clause, a direct object in a nested relative clause, and so on. Under this approach, possessors and possessees are treated as two individual references, even where they function as a single argument of a clause.

The result of this method is a string of references in the linear order in which they occur in the discourse. This shifts the method from one of counting the number of clauses (between a reference and its antecedent, say), to one where the object of the counting is the number of references. Consider the excerpt from Lewis Carroll’s (1865) classic Alice’s
Adventures in Wonderland in (11); if this excerpt is split into clauses, then under Givón’s method one could calculate a referential distance of ‘2’ between it in clause (iv), and its antecedent, Dodo, in clause (ii).

(11)

i. Then they all crowded round her once more,
ii. while the Dodo solemnly presented the thimble, saying
iii. ‘We beg your acceptance of this elegant thimble’; and,
iv. when it had finished this short speech,
v. they all cheered.

On the other hand, if each and every reference is extracted from the passage, as in (12), where each reference is presented between braces and then listed below the excerpt, then one can instead refer to the fact that the reference it is the fifth reference after its antecedent, the Dodo.

(12)

Then {they} all crowded round {her} once more, while {the Dodo} solemnly presented {the thimble}, saying '{We} beg {your} acceptance of {this elegant thimble}'; and, when {it} had finished {this short speech}, {they} all cheered.³

1. they
2. her
3. the Dodo
4. the thimble
5. we
6. your
7. this elegant thimble
8. it
9. this short speech
10. they

Since this approach treats all references equally (regardless of position in a clause, grammatical role, semantic role, and so on), it resolves the issues outlined in section 3.2 relating to: (i) adverbial/complement clauses; (ii) (nested) relative clauses; and (iii) possessive NPs.

It should be noted here that when a verb occurs without an overt argument (i.e. zero reference), rather than trying to determine at what point in the text the reference would have occurred if it were overt, the corresponding verb is taken as a proxy for the reference. For example, in the sentence ‘the lion ate the gazelle and Ø ran to the waterhole’, the verb ran stands in for the zero reference; in this example it is easy to determine the position of the zero reference, given the strict word order of English, but this is not true of all the languages in this study. In the case of zero reference in Kilivila (see section 4.3), which has serial verb

² Although Alice’s Adventures in Wonderland is a highly stylised text, which does not reflect the nature of the more natural narratives used in this study (with the exception of Spanish), it is used here purely as a means of illustrating the methodology, since it demonstrates multiple distinct references in a short passage of text.
³ The referent ‘acceptance’ has been omitted on purpose, due to it being an abstract noun.
constructions, only one reference is posited per serial verb construction. The Kilivila example in (13) involves a serial verb construction consisting of three finite verbs, with no overt argument, but is nevertheless treated as a single instance of zero reference.

(13) [ e-sakaula e-va i-toli ] okepapala la kwava
    [ 3-run 3-go.to 3-stand ] at.the.side.of POSS.3SG wife

‘(He) goes running to stand beside his wife’

Once all references have been extracted from a text, they are entered into a database together with key information pertaining to the reference, such as the referential device used (e.g. NP, pronoun, agreement affix, zero reference, etc.), part of speech (e.g. noun, verb, clitic) and any grammatical information they encode (person, number, gender). Taking the list of references in (12), the data would then look like that presented in (14).

(14) ID Reference Device PoS Number Person Gender
    1 they Pro N pl 3 –
    2 her Pro N sg 3 Fem.
    3 the Dodo NP N sg 3 –
    4 the thimble NP N sg 3 –
    5 we Pro N pl 1 –
    6 your Pro N sg 2 –
    7 this thimble NP N sg 3 –
    8 it Pro N sg 3 Neuter
    9 this speech NP N sg 3 –
    10 they Pro N pl 3 –

Following this step, the referent of each reference is identified and other relevant information is then added, including information relating to the animacy of each participant and to whether it is a protagonist of the story. This is exemplified in (15).

(15) ID Reference Referent Animate? Protagonist?
    1 they party of animals Y –
    2 her Alice Y Y
    3 the Dodo Dodo Y Y
    4 the thimble thimble – –
    5 we party of animals Y –
    6 your Alice Y Y
    7 this thimble thimble – –
    8 it Dodo Y Y
    9 this speech speech – –
    10 they party of animals Y –

Once this procedure has been carried out for all the references in a discourse, the references are scored according to three criteria, using scoring strategies which are explained in sections 3.3.1–3.3.3. The first criterion assesses the activation level of the referent on the
basis of its presence and role in the immediately preceding discourse. This is an important step, since activation plays a crucial role in reference tracking (see, for instance, Chafe 1996, Kibrik 2011) and it is thus not possible to look solely at the role gender and classifiers might play to the exclusion of other factors. An activation score is calculated for all references, regardless of type. The second and third criteria assess the contribution that gender (together with number) and classifiers, respectively, make to eliminating other (potentially) conflicting references. References which spell out the identity of a referent (i.e. most NPs, except where the NP is a pronoun, demonstrative, quantifier, etc.) are not assessed according to these final two criteria. Also, it goes without saying that only references in Spanish and Mian are assessed for the contribution of gender, and only references in Mian and Kilivila are assessed for the contribution of classifiers.

3.3.1. Activation score

The activation score is made up of four separate components which are aggregated to give an overall activation score. These four components are loosely based on some of the “activation factors” that Kibrik (2011: chapter 11) uses in his discussion of a cognitive multifactorial approach to referential choice. The values used for the scores are also loosely based on Kibrik, where the higher the value the more activated a referent is. (It is worth noting here that, while this scoring system may not appear particularly scientific in nature, the absolute scores are not really what matters, but rather it is the relative scores which count. If a referent is awarded a score of 1.5, it can be seen as 50% more activated than a referent with a score of 1.0.)

The four factors can be broadly defined as (i) frequency, (ii) grammatical role, (iii) animacy and (iv) protagonisthood, and can be described as follows:

(i) **FREQUENCY** – the total number of times the referent is mentioned in the preceding 10 references. The raw count is then converted into a value from 0 to 1, in accordance with Table 1.

(ii) **GRAMMATICAL ROLE** – this element of the score is based on the intersection of two variables: (i) the last mention of the referent in question, in terms of number of references; and (ii) the grammatical role of the referent the last time it was mentioned, on a four-way scale from subject > object > indirect or secondary object > any other grammatical role. The scores awarded, again between 0 and 1, are laid out in Table 2.

(iii) **ANIMACY** – if a reference refers to an animate referent it is awarded a score of 0.2; inanimates receive a score of 0. This element of the activation score is included in recognition of the fact that animate participants are much more likely to be active discourse participants than inanimate or abstract referents. However, the score awarded for the animacy element is relatively low, to give priority to the other factors contributing to the overall activation score.

(iv) **PROTAGONISTHOOD** – referents which are identified as being one of the principle characters of a story are awarded an additional score of 0.5. Protagonists are typically animate referents that are present throughout most of the narrative and identifying them is usually a straightforward and uncontroversial task. In the example taken from *Alice’s Adventures in Wonderland*, Alice and the Dodo would be considered as protagonists, while the party of animals, the
speech and the thimble would not be. While it could be argued that
protagonists, by virtue of their predominance in a narrative, will already
score highly in terms of frequency, this would miss the fact that it is likely
much easier for a speaker to re activate a protagonist with a reduced
referential device after a period of absence from the discourse than it would
be for a non-protagonist.

<table>
<thead>
<tr>
<th>Frequency</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>≥ 5</td>
<td>1.0</td>
</tr>
<tr>
<td>4</td>
<td>0.8</td>
</tr>
<tr>
<td>3</td>
<td>0.6</td>
</tr>
<tr>
<td>2</td>
<td>0.4</td>
</tr>
<tr>
<td>1</td>
<td>0.2</td>
</tr>
<tr>
<td>0</td>
<td>0.0</td>
</tr>
</tbody>
</table>

Table 1. Element of activation score based on frequency of mentions

Table 1 shows the scores assigned to references based on the frequency of the
referent in the immediately preceding ten references. The reason for looking back ten
references was to build up a more accurate picture of the referent’s frequency. In calibrating
the scoring system it was found that, when coding every reference in a passage of discourse,
a referent which is otherwise very frequent can quite easily not be mentioned in the
preceding five or so references. Looking back five references or less, therefore, resulted in
some relatively frequent references being scored zero. On the other hand, looking back more
than ten references seemed unnecessary; if a referent had not been mentioned a single time
in the preceding ten references it was treated as not being frequent enough to increase its
activation level. Ten references was therefore chosen as the optimal ‘look-back’ distance for
this element of the activation score.

<table>
<thead>
<tr>
<th>Last mention</th>
<th>Subject</th>
<th>Object</th>
<th>Indirect object</th>
<th>Other</th>
</tr>
</thead>
<tbody>
<tr>
<td>-1</td>
<td>1.0</td>
<td>0.9</td>
<td>0.8</td>
<td>0.7</td>
</tr>
<tr>
<td>-2</td>
<td>0.8</td>
<td>0.7</td>
<td>0.6</td>
<td>0.5</td>
</tr>
<tr>
<td>-3</td>
<td>0.6</td>
<td>0.5</td>
<td>0.4</td>
<td>0.3</td>
</tr>
<tr>
<td>-4</td>
<td>0.4</td>
<td>0.3</td>
<td>0.2</td>
<td>0.1</td>
</tr>
<tr>
<td>-5</td>
<td>0.3</td>
<td>0.1</td>
<td>—</td>
<td>—</td>
</tr>
</tbody>
</table>

Table 2. Element of activation score based on last mention and grammatical role

Table 2 shows the scores assigned to references based on the grammatical role of the
referent when it was last mentioned. To assign this score, it is first necessary to count how
many references back in the text the referent was last mentioned. If the referent was last
mentioned six or more references ago, no score is awarded. If, however the referent was last
mentioned five or fewer references ago, a score is awarded based on that reference’s
grammatical role. Consider, for instance, the final occurrence of it (in boldface) in (16) (which
includes the excerpt in 12): this is a reference to the Dodo, which was last mentioned five
references back in the discourse, at which point it was the subject of the clause (‘the Dodo
solemnly presented the thimble’). This element of the score, for the reference \textit{it}, is therefore 0.3 (the intersection of the ‘Subject’ column with the ‘-5’ row in Table 2).

(16) ‘Of course,’ \textit{(the Dodo)} replied very gravely. ‘What else have \{you\} got in \{your\} \{pocket\}?’ \textit{(he)} went on, turning to \{Alice\}. ‘Only \{a thimble\},’ said \{Alice\} sadly. ‘Hand \{it\} over here,’ said \textit{(the Dodo)}. Then \{they\} all crowded round \{her\} once more, while \textit{(the Dodo)} solemnly presented \{the thimble\}, saying ‘\{We\} beg \{your\} acceptance of \{this elegant thimble\}’; and, when \textit{(it)} had finished \{this short speech\}, \{they\} all cheered.

The reason for looking back at the preceding ten references for the frequency element of the score, but only the preceding five references for the grammatical role element of the score, is because these two elements are fundamentally different in nature: the frequency element is of a quantitative nature, and so it is necessary to look back over a larger number of references to gain a better idea of a referent’s frequency, whereas the grammatical role element is of a qualitative nature, where the effect of the grammatical role is considered less relevant to a referent’s activation level the further back it is.

Now let us consider how the total activation score is calculated from the four elements outlined above for this same reference, \textit{it}, in (16). (It is interesting to note that the author of this story alternates between the masculine pronoun \textit{he} and the neuter pronoun \textit{it} when referring to this character, no doubt a clash between personification of \textit{the Dodo}, on the one hand, and the general rule of English grammar which refers to lower animates with \textit{it}, on the other.) The list of references from this excerpt are presented in (17). References to \textit{the Dodo} are presented in boldface, while the reference which is the focus of the scoring is also shaded.


Thus in assessing the level of activation of \textit{the Dodo} at the point in which it is referred to with the pronoun \textit{it}, a score of 1.4 would be given, which is made up as follows:

(i) \textbf{FREQUENCY} – in the ten references which precede reference #18 (i.e. references 8–17), the Dodo is mentioned twice; as a result, it gets a frequency score of 0.4 (see Table 1).

(ii) \textbf{GRAMMATICAL ROLE} – the last time the Dodo was mentioned before reference #18 was reference #13 (i.e. 5 references back in the discourse) where it was the subject of the adverbial clause. The score is therefore 0.3 (see Table 2).

(iii) \textbf{ANIMACY} – the Dodo is an animate entity and is thus awarded an animacy score of 0.2.

(iv) \textbf{PROTAGONISTHOOD} – the Dodo is one of the main characters in this story and so is awarded a score of 0.5.

(v) Summing up the scores for the individual components gives a total score of 1.4.
3.3.2. Contribution of gender score

As explained in section 2.2, this paper is concerned only with the issue of referent disambiguation in the context of discourse (i.e. across clauses) and does not purport to address other claimed functions of nominal classification, such as those outlined in Contini-Morava & Kilarski (2013). In this regard, then, when assessing the contribution that gender makes to referent disambiguation, the key question becomes: to what extent does gender aid in reducing competing referents? For the purpose of quantifying this contribution, a simple formula is used, as outlined below.

When developing the methodology for this study and calibrating the scoring method for the contribution of gender, it became clear that looking back ten references (as with the frequency score, see section 3.3.1) was unnecessary, as it is extremely rare to have ten competing references on the scene at any one time, whereas looking back only five references (as with the grammatical role score, see section 3.3.1) was too restrictive and uninformative. Therefore, it was decided that the contribution of gender score should look back at the preceding seven references.

Consider a passage of text in which there are five characters: A, B, C, D, E. (Note that, in discussing the contribution of gender score it is necessary to turn our attention away from the Alice in Wonderland excerpt used in the previous sections, and begin with an abstract representation of a passage of text.) Now, let us assume that they occur in a passage of text in the order presented in (18).

(18)   ID   Referent
       1   A
       2   B
       3   A
       4   C
       5   D
       6   D
       7   E
       8   B

First, the number of unique referents which occur in the seven references that precede the reference in question (here, reference number 8) is determined. The referent itself is excluded from this count, so if all seven preceding references relate to the referent, the number of unique referents would be 0. If the referent is not mentioned at all, then the maximum number of unique referents would be 7, but a referent may occur multiple times across the seven references thereby reducing this figure. In example (18), there are four unique referents represented (A, C, D, E) in the seven references which precede Reference 8. Reference 2 is not treated as a fifth unique referent, since the identity of the reference under examination is also Referent B.

Once the number of unique referents has been established, the next step is to determine how many of these unique referents are marked for a gender that is formally distinct from the gender-marking of the reference in question (and can therefore be
eliminated as potential antecedents).\(^4\) In (19) – a repetition of (18) with gender values added, which for the purpose of this illustration should be understood to be formally distinct from each other – three of the four competing referents (A, D and E) are marked for gender-number values which are distinct from the reference in question; referent C, however, is marked for the same gender-number value as Referent B. (Reference 2 is a reference to Referent B, so this is not counted.) Again, it is important to note that multiple references to the same referent should only be counted once. In example (19), although the F.PL marking of Referent D occurs twice, it should only be counted once, and the same goes for Referent A.

<table>
<thead>
<tr>
<th>ID</th>
<th>Gender (and number) marking</th>
<th>Referent</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>M.SG</td>
<td>A</td>
</tr>
<tr>
<td>2</td>
<td>M.PL</td>
<td>B</td>
</tr>
<tr>
<td>3</td>
<td>M.SG</td>
<td>A</td>
</tr>
<tr>
<td>4</td>
<td>M.PL</td>
<td>C</td>
</tr>
<tr>
<td>5</td>
<td>F.PL</td>
<td>D</td>
</tr>
<tr>
<td>6</td>
<td>F.PL</td>
<td>D</td>
</tr>
<tr>
<td>7</td>
<td>F.PL</td>
<td>E</td>
</tr>
<tr>
<td>8</td>
<td>M.PL</td>
<td>B</td>
</tr>
</tbody>
</table>

The final step in calculating the ‘contribution of gender’ score is to divide the number of referents whose gender marking is formally distinct from that of the referent in question by the number of unique referents. This is illustrated in (20) for the example given in (19).

\[
\frac{\text{referents with distinct gender marking}}{\text{unique referents}} = \frac{3}{4} = 0.75
\]

In this case, there are four unique referents, three of which have gender-number marking which is formally distinct from that of the referent in question. This means that, thanks to gender-number marking, the potential for referential conflict with regard to reference 8 has been reduced by 75% (indicated by the score of 0.75).

If Referent C had not been the same gender as Referent B in this example, the resulting ‘contribution of gender score’ would have been \(4/4 = 1.0\), indicating a 100% reduction (or rather elimination) of referential conflict.

The converse of this would be a situation where all of the unique referents display gender-number marking which is the same as that of the referent in question. This would result in a contribution of gender score of \(0/x\) (where ‘x’ is the number of unique referents), which in all cases would result in 0.0, or 0% reduction in referential conflict.

A further example is given in (21). For the sake of drawing a comparison, there are again four unique referents (A, C, D and E). However, this time two of them (A and E) share the gender-number marking of the referent in question, leaving only two referents (C and D) which can be eliminated on the basis of gender-number marking. The contribution of gender

\(^4\) The importance of only counting referents whose gender marking is formally distinct is a pertinent issue in Mian, where syncretisms in the agreement markers of the four genders result in only a two-way formal distinction in the singular and a two-way formal distinction in the plural, despite the existence of four controller genders. See section 4.4 for more details.
score in this instance is 0.5 (i.e. 2/4), which reflects the fact that fifty percent of the competing referents have been eliminated.

(21) ID Gender marking Referent
1 M.PL A
2 M.PL B
3 M.PL A
4 M.SG C
5 F.PL D
6 F.PL D
7 M.PL E
8 M.PL B

To demonstrate this scoring system, consider the Spanish passage taken from the well-known fable The Pied Piper in (22). In particular, consider the reference éstos 'these' (the masculine plural form of the proximal demonstrative pronoun) on the final line, whose referent is the leaders (of the Council). Note that, in cases of zero reference, the corresponding verb has been tagged; in many case the person and number of the subject can be recovered from the form of the verb in Spanish.

(22) A la mañana siguiente, {el flautista} se presentó ante {el Consejo} y {reclamó} a {los prohombres} de {la ciudad} {las cien monedas de oro} prometidas como {recompensa}. Pero éstos, [...] [le] contestaron [...] ‘The next morning, the flute played presented himself before the council and demanded from the leaders of the city the hundred gold coins promised as a reward. But they (lit. these) answered him …’

The list of references from this excerpt are presented in (23) together with their respective gender-number values. The one case of zero reference, which corresponds to the tagged verb in (22), has been entered as Ø. References to the leaders are presented in boldface, while the reference which is the focus of the scoring is also shaded. In the seven references that immediately precede Reference 8, there are six unique referents (el flautista is referred to twice). One of these, los prohombres ‘the leaders’, is co-referential with the referent in question, leaving a total of five other referents.

(23) Reference Gender-number
(i) el flautista (the pied piper) M.SG
(ii) el Consejo (the council) M.SG
The contribution of classifier score

The contribution that classifiers play in referent disambiguation is handled in exactly the same way as has been outlined for gender in section 3.3.2. That is, the number of referents which occur with a classificatory particle (Kilivila) or verbal classifier (Mian) which is (formally) distinct from that of the referent in question is divided by the number of unique referents mentioned in the preceding seven references.

A crucial difference here, however, concerns the relative frequency of classifiers when compared with obligatory gender marking. This is especially pronounced in Mian, where classifiers are restricted to occurring with only around 40 verbs.

4. Data

The data used in this study comprise narrative texts from four languages, which differ typologically from each other with regard to nominal classification. The four languages were selected in an attempt to cover the typological space. They comprise: Skolt Saami (Finno-Ugric, Finland), a language with no nominal classification; Spanish (Romance), a language with gender; Kilivila (Austronesian, Papua New Guinea), a language with classifiers; and Mian (Papuan, Papua New Guinea), one of the rare instances of a language displaying both gender and classifiers. This four-way typology is illustrated in (25).

(25)

<table>
<thead>
<tr>
<th>Classifiers</th>
<th>Gender</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>No</td>
<td></td>
<td>Skolt Saami</td>
</tr>
<tr>
<td>Yes</td>
<td></td>
<td>Kilivila</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Spanish</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Mian</td>
</tr>
</tbody>
</table>

The texts selected for inclusion in the study were all narrative texts, since the presence of an appropriate number of third person animate/human references was an important consideration; procedural texts, for example, tend to involve mostly first/second person references and references to inanimate entities and thus they do not lend themselves to a study of reference tracking. All the texts were fictional texts, typically involving a number of human characters, other animates (which may or may not be personified) and inanimates.
A total of five texts were studied per language, and the total number of references encoded across the four languages was 2775. The total number of referents represented was 575. Before presenting the results, a brief overview of the four languages is given, paying particular attention to their systems of nominal classification.

4.1. Skolt Saami

Skolt Saami is a Finno-Ugric language spoken in the far northeast of Finland by between 150 and 300 speakers (Feist 2015). The language exhibits neither gender nor classifiers and was therefore selected for this study as a control language in order to compare how it deals with referent disambiguation in the absence of nominal classification and to contrast this with the other languages in the study.

Verbs in Skolt Saami mark the person and number of the subject, meaning the person of the subject is clear in the absence of a pronoun, and instances of zero reference abound. In addition, the demonstrative pronouns also serve as third person references.

4.2. Spanish

Spanish, which differs drastically from the other three languages in this study by virtue of the fact it has over 400 million speakers, is one of the best-known examples of a language with gender. Gender in Spanish is essentially restricted to a masculine–feminine contrast — a small number of pronouns also have a neuter form (lo, ello, eso, esto, aquello) used to refer to ideas and concepts, but essentially Spanish can be thought of as a two-gender system.

As is typical of gender systems, nouns in Spanish are inherently either masculine or feminine. The gender of many Spanish nouns is evident from their form, with masculine nouns typically ending in -o, and feminine nouns typically ending in -a, but there are numerous exceptions, e.g. clima ‘weather (masc.)’, mano ‘hand (fem.)’. There are general tendencies for nouns which do not end in -o or -a (nouns in -dad, -ión and -ez are predominantly feminine, nouns in -e are predominantly masculine, etc.), but even so there remain a number of nouns whose gender cannot be determined from their form.

Gender agreement targets within the noun phrase are articles, adjectives, demonstratives, a subset of numerals and some quantifiers (see example 26), and at the level of the clause gender is marked on the direct object when it is expressed pronominally or as a clitic (27), as well as on modifiers external to the NP as already seen in examples (4) – (6).

(26) a. tiene **un-a** casa blanc-a
   have.PRS.3SG ART.INDEF-F.SG house(F).SG white-F.SG
   ‘s/he has a white house’

   b. viste **es-as** tres-cient-as **libras?**
   see.PST.2SG DIST-F.PL three-hundred-F.PL pounds(F).PL
   ‘did you see those three hundred pounds?’

(27) a. quiero **ver-la**
   want.PRS.1SG see=OBJ(F).SG
   ‘I want to see her’
b. no lo recibirás hasta mañana
   no OBJ(M).SG receive.FUT.2SG until tomorrow
   ‘you won’t receive it (masc.) until tomorrow’

There is no doubt that gender in Spanish can serve to disambiguate in all sorts of situations – a clear example of this was given in example (4) – but this function relies on the two referents in question being of opposing genders, a condition that cannot be relied upon in a language with such a restricted gender system. Even so, this gender distinction is a feature of Spanish that speakers are able to exploit when the conditions are right, and so the question remains as to whether or not they do so in the context of referent disambiguation.

4.3. Kilivila

Kilivila is an Austronesian language spoken by around 25,000 people on the Trobriand Islands off the western tip of Papua New Guinea (Senft 1986). It is particularly well known for its extensive system of classificatory particles\(^5\), first described by Malinowski (1920). A total of 177 classifiers have been documented for Kilivila (Senft 1996).

Classifiers surface as prefixes or infixes on demonstrative pronouns, adjectives and numerals, which require concord with the class of noun they refer to. This is neatly illustrated in (28) (Senft, personal communication), where the classifier \textit{ke}, used when the noun refers to a wooden object (and thus glossed as CLF.WOODEN), appears as a prefix on both the numeral \textit{yu} ‘two’ and the adjective \textit{manabweta} ‘beautiful’, and as an infix in the demonstrative pronoun \textit{ma-}-\textit{na} ‘this’.

\begin{align*}
\text{(28)} & & \text{ke-} & \text{yu} & \text{waga} & \text{ma-ke-si-na} \\
& & \text{CLF.WOODEN-two} & \text{canoe} & \text{DEM-CLF.WOODEN-PL-DEM} \\
& & \text{ke-manabweta} & \text{le-kota-si} \\
& & \text{CLF.WOODEN-beautiful} & \text{3.PST-arrive-PL} \\
& & \text{‘these two beautiful canoes arrived’}
\end{align*}

While the nature of Kilivila classifiers might be considered more ‘gender-like’ with respect to agreement with multiple constituents (see Fedden and Corbett 2017: 22), they also display the more ‘classifier-like’ behaviour of a one-to-many relationship between nouns and classifiers. This allows speakers to temporarily classify nouns to emphasise certain characteristics, whereas in a gender system a noun’s gender is typically fixed (although this is not always the case). This is illustrated in (29) where the noun \textit{yena} ‘fish’ appears in four different numeral phrases, but each time with a distinct classifier to emphasise a different attribute of the fish. In (29a) the default classifier for animals \textit{na} is used, but in (29b) the classifier \textit{kevala}, used to refer to items being batch dried, appears with the numeral \textit{lima} ‘five’ – the resulting meaning is thus not ‘five fish’ but ‘five batches of fish’. A similar situation holds in (29c), where the classifier \textit{oyla} refers to a bunch of fish on a string.

\begin{align*}
\text{(29)} & & \text{ke-} & \text{yena} & \text{na} \\
& & \text{CLF.WOODEN-two} & \text{DEM} & \text{DEM} \\
& & \text{ke-} & \text{kevala} & \text{lima} & \text{si} \\
& & \text{CLF.WOODEN-beautiful} & \text{3.PST-arrive-PL} & \text{DEM} \\
& & \text{‘five batches of fish arrived’}
\end{align*}

\(^5\) Although the term ‘classificatory particle’ captures the fact that classifiers in Kilivila consist of a morpheme affixed to another word (primarily demonstrative pronouns), the term \textit{classifier} will nevertheless be used throughout this paper.
(29)  a.  na-tala       yena  
     CLF.ANIMAL-one  fish  
     ‘one fish’

b.  kevala-lima    yena  
    CLF.BATCH.DRYING-five fish  
    ‘five batches of smoked fish’

c.  oyla-lima       yena  
    CLF.STRING-five  fish  
    ‘five strings with stringed on fish’

d.  pwasa-lima   pwasa-tala  yena  
    CLF.ROTTEN-five  CLF.ROTTEN-one  fish  
    ‘six rotten fish’

In all the examples in (28) and (29) a head noun is present, and so the presence of a classifier in these contexts cannot be seen as serving a disambiguating role. However, nouns may be elided from an NP, in which case the classifiers on the modifying elements can result in the identity of the referent being easily recoverable. A simple example of this is presented in (30), where the numeral _tala_ ‘one’ is repeated three times. Since each occurrence of the numeral occurs with a classifier, the intended referent is unequivocal.

(30)  bi-bodi  te-tala  na-tala  gudi-tala  
     3.FUT-benefit  CLF.MALE-one  CLF.FEMALE-one  CLF.CHILD-one  
     ‘it will benefit each man, woman, and child’

Although reference tracking has been claimed to be one of the functions of classifiers in Kilivila, almost no instances of elided nouns where the classifier was retained on a modifying element were found in the texts used for the present study, and so the results of this study (see Section 5) do not provide compelling evidence to support this hypothesis, at least not with regards to referent disambiguation.

4.4. Mian

Mian, a Papuan language of the Ok family (Trans New Guinea), is spoken in the central highlands of Papua New Guinea by fewer than 2000 speakers (Fedden 2011: 2). The language is one of the few languages which are known to have two concurrent systems of nominal classification, with both a gender system and a system of verbal classifiers (but see Fedden & Corbett 2017 and Corbett, Fedden & Finkel 2017 for in-depth discussions as to the nature of these systems).

Mian has four genders: masculine, feminine, neuter 1 and neuter 2. The masculine and feminine genders are assigned to male and female humans, respectively, while neuter 1 is assigned to most inanimates, and neuter 2 to locations, weather phenomena, abstract nouns, and some tools, among other entities. Agreement targets are articles (which appear as clitics on the noun), pronouns, and verbs. The clitic articles are presented in Table 3; the
agreement forms are different on pronouns and verbs, but the same pattern of syncretisms is observed.

<table>
<thead>
<tr>
<th></th>
<th>SG</th>
<th>PL</th>
</tr>
</thead>
<tbody>
<tr>
<td>MASCULINE</td>
<td>=e</td>
<td>=i</td>
</tr>
<tr>
<td>FEMININE</td>
<td>=o</td>
<td>=i</td>
</tr>
<tr>
<td>NEUTER 1</td>
<td>=e</td>
<td>=o</td>
</tr>
<tr>
<td>NEUTER 2</td>
<td>=o</td>
<td>=o</td>
</tr>
</tbody>
</table>

Table 3 – Mian clitic articles

The examples in (31) show gender marked on all three targets: the pronoun, the article and the verb. In addition to the subject indexing on the verb, found on all finite verbs, the verb têm’ ‘see’ is one of seven verbs in Mian which also displays object agreement (Fedden 2011: 265–267), so in fact here we see gender marking occurring four times in each clause.

(31) a. ē unáng=o
3SG.M woman(f)=ARTICLE.SG.F ‘he sees the woman’
wa-têm’-Ø-e=be 3SG.F.OBJ-see.PFV-REALIS-3SG.M.SBJ=DECL
b. ō naka=e
3SG.F man(m)=ARTICLE.SG.M ‘she sees the man’
a-têm’-Ø-o=be 3SG.M.OBJ-see.PFV-REALIS-3SG.F.SBJ=DECL

While there are four controller genders in Mian (that is, genders defined in terms of what agreements the controller requires), the syncretisms mean there are only two target genders (that is, the number of gender distinctions observed on the target): =e and =o in the singular, and =i and =e in the plural. This is, of course, of particular relevance in a discussion about referent disambiguation; it is precisely in the absence of the controller when gender may serve to disambiguate between referents, yet with only targets to go off only a two-way gender distinction can be observed.

Aside from gender, Mian has a system of six classifiers, which appear as verbal prefixes, and which classify the transitive object of about forty verbs of object handling or movement (e.g. take, give, throw), as illustrated in (32), as well as the subject of the intransitive verb ‘fall’.

(32) mën=e
gol-â-Ø-i-o=be
string_bag(n1)=ARTICLE.SG.N1 CLF.BUNDLE-put.PFV-REALIS-1SG.SBJ-EP=DECL ‘I have put down the string bag.’

The classifiers, presented in Table 4, categorise objects largely, but not exclusively, on semantic grounds. The M-classifier, for instance, incorporates all nouns of masculine gender, and a subset of neuter 1 nouns. The F-classifier incorporates all but three feminine nouns, and all neuter 2 nouns. The semantic basis for the remaining four classifiers, however, is much more transparent.

<table>
<thead>
<tr>
<th></th>
<th>SG</th>
<th>PL</th>
</tr>
</thead>
<tbody>
<tr>
<td>M-CLASSIFIER</td>
<td>dob-</td>
<td>dol-</td>
</tr>
</tbody>
</table>
Table 4 – Mian classifiers (NB. only one allomorph presented)

Crucially, the forty-or-so verbs which take a classifier do not overlap with the seven verbs which index the direct object in the gender system, so there is never a case where a classifier is in competition with an object agreement marker on the verb. Compare the two examples in (33) – both clauses involve a masculine and a feminine participant, yet in (33a) the verb ‘see’ is one of the seven verbs which require the direct object to be indexed in accordance with the gender system, while in (33b) the verb ‘take’ belongs to the set of forty verbs which require the object to be indexed with a verbal classifier.

(33)  
a. unáŋɡ̥=o  
woman(ː)=ARTICLE.SG.F  
‘he saw a woman’

b. unáŋɡ̥=o  
woman(ː)=ARTICLE.SG.F  
‘he took a wife’

5. Results

In this section, the findings of this study are presented from three perspectives in line with the three scores assigned to each reference, as discussed in Section 3, namely:

1. The correlation between referential choice and the activation score of the referent;
2. The contribution of gender in reducing or eliminating referential conflict; and
3. The contribution of classifiers in reducing or eliminating referential conflict.

5.1 Correlations between referential choice and activation scores

It is no great surprise, of course, that references will generally surface as a full NP when the referent’s activation level is low and as a reduced device when the activation level is high. This has been demonstrated by Kibrik (2011: 380), who also shows how, at the intersection of high and low activation levels, both a full NP and reduced device are often equally likely, all else being equal.

Independently of a referent’s high activation level, however, speakers must decide if the use of a reduced referential device may result in referential conflict. Kibrik (2011: 394) refers to this step in referential choice as a ‘filter’ that checks the projected choice. If referential conflict is likely, a full NP may be used even where an activation score might be indicative of a reduced device. The question here, then, is whether gender and/or classifiers – acting as referential aids – reduce the likelihood of referential conflict to such an extent as to affect the overall correlation between referential choice and activation scores. For
instance, we might hypothesise that a language without gender or classifiers uses more full NPs (if we also assume it has a poorer repository of referential aids at its disposal).

In analysing the data, references were divided into three groups: (i) those whose activation score is less than 1; (ii) those where it is equal to or greater than 1, but less than 2; and (iii) those where it is equal to or greater than 2. Within each group, references were categorised into three groups according to the referential device employed: (i) full NPs, (ii) pronouns, and (iii) other devices. The category ‘other devices’ consisted predominantly of instances of zero reference, but also served as a catch-all category for references that did not fit neatly into the other two categories, including NPs that did not clearly identify the referent (e.g. demonstratives, quantifiers, modifiers serving as NP heads). The data are presented in Table 5.

<table>
<thead>
<tr>
<th>Activation level</th>
<th>Referential choice</th>
<th>Skolt Saami</th>
<th>Kilivila</th>
<th>Spanish</th>
<th>Mian</th>
</tr>
</thead>
<tbody>
<tr>
<td>≥2</td>
<td>Full NP</td>
<td>20</td>
<td>32</td>
<td>9</td>
<td>12</td>
</tr>
<tr>
<td></td>
<td>Pronominal</td>
<td>37</td>
<td>45</td>
<td>24</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>Other</td>
<td>66</td>
<td>96</td>
<td>57</td>
<td>180</td>
</tr>
<tr>
<td></td>
<td></td>
<td>123</td>
<td>173</td>
<td>90</td>
<td>196</td>
</tr>
<tr>
<td>≥1 &lt;2</td>
<td>Full NP</td>
<td>59</td>
<td>93</td>
<td>61</td>
<td>32</td>
</tr>
<tr>
<td></td>
<td>Pronominal</td>
<td>32</td>
<td>53</td>
<td>42</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>Other</td>
<td>79</td>
<td>125</td>
<td>141</td>
<td>178</td>
</tr>
<tr>
<td></td>
<td></td>
<td>170</td>
<td>271</td>
<td>244</td>
<td>213</td>
</tr>
<tr>
<td>&lt;1</td>
<td>Full NP</td>
<td>226</td>
<td>288</td>
<td>316</td>
<td>211</td>
</tr>
<tr>
<td></td>
<td>Pronominal</td>
<td>11</td>
<td>10</td>
<td>19</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>Other</td>
<td>25</td>
<td>43</td>
<td>58</td>
<td>83</td>
</tr>
<tr>
<td></td>
<td></td>
<td>262</td>
<td>341</td>
<td>393</td>
<td>299</td>
</tr>
</tbody>
</table>

Table 5 – correlation between activation scores and referential choice

Figure 1 is a box and whisker plot showing the distribution of activation scores by referential device category for all 2775 references in this study. An independent-samples t-test on these data indicated that activation scores for NPs ($M = 0.57$, $SD = 0.68$) were significantly lower than for all other referential devices combined ($M = 1.67$, $SD = 0.70$), $t(2773) = 42.02$, $p < .001$, $d = 1.60$. 
Starting with references which have activation scores lower than 1, Figure 2 clearly illustrates how, as expected, all four languages favour a full NP in these environments. In all four languages, pronouns barely occur when the reference has a low activation score, while non-pronominal reduced referential devices are more prevalent than pronouns. Skolt Saami has the highest occurrence of full NPs (86% of references with a low activation score were full NPs), while Mian has the lowest (71%). Kilivila and Spanish fall midway between the two, at 84% and 80%, respectively.
Turning to references with a mid activation score ($\geq 1 < 2$), a number of observations can be made from Figure 3. Firstly, as to be expected, the prevalence of full NPs drops across all languages: in the case of Skolt Saami from 86% down to 35%; in Kilivila an almost identical pattern holds, going from 84% to 34%; in Spanish the drop is from 80% to 25%; while Mian witnesses the biggest fall of all, from 71% to 15%.

The more interesting observation, however, concerns the differences between the individual languages, with Mian in particular witnessing a marked drop in the use of full NPs. Between the Skolt Saami and Mian values, there is a 20% difference in the use of full NPs and an 18% difference in the use of pronominal elements. The Kilivila values are both within a percentage point of the Skolt Saami ones, with the Spanish value for full NPs falling midway between Skolt Saami/Kilivila and Mian (while the use of pronominal elements is only slightly lower).
Figure 4 illustrates the pattern that emerges when a reference has an activation score that is equal to or greater than 2. Here the relative frequencies of full NPs and pronouns have switched places, exactly as we would anticipate, with the exception of Mian where the figures for both are, in any case, approaching zero.

![Bar chart showing percentage of references with high activation scores (≥2) in three referential device categories]

It is interesting to note that, in all three data sets, the language with classifiers only (Kilivila) patterns with the language with neither gender nor classifiers (Skolt Saami). The language with gender only (Spanish) shows a slight drop in the use of full NPs and pronominals in mid- and high-activation environments, and the language with both gender and classifiers displays the more pronounced drop. While these may be interesting observations, in order to see if gender or classifiers are potentially playing a role in referent disambiguation, it is necessary to look at the contribution they each make to either reducing or eliminating referential conflict.

5.2 Contribution of gender in reducing or eliminating referential conflict

While full NPs in Mian and Spanish have an inherent gender, the very fact they are full NPs usually (although not always) means that the identity of the referent is unambiguous. As such, gender can only really be seen to be playing a role in the case of reduced referential devices. For Mian, this reduces the number of references in question from 690 to 435 (63% of the total number of references in the Mian data); in the case of Spanish, the number drops from 722 to 336 (47%). Of the remaining references, a significant number (280 for Spanish, 112 for Mian) are unmarked for gender (for example, in cases of zero reference) and so gender clearly cannot play a role in reducing or eliminating referential conflict. These figures are presented in Table 8, together with the results for references where gender may play a role: in Mian, we are left with 323 references (47% of the total number of references encoded) and in Spanish a mere 56 references remain (less than 8% of the total references encoded).

In Table 8, references are divided into one of five categories: (i) references where gender is unable to play a role (due to the reference being a full NP); (ii) references where gender is unable to play a role, since the reference is unmarked for gender; (iii) references...
where gender eliminates referential conflict (percentages are shaded); (iv) references where gender reduces referential conflict (percentages are shaded); and (v) references where gender could assist in reducing referential conflict, but fails to do so.

<table>
<thead>
<tr>
<th>Gender unable to play a role (full NP)</th>
<th>Spanish</th>
<th>Mian</th>
<th>Spanish</th>
<th>Mian</th>
</tr>
</thead>
<tbody>
<tr>
<td>Referential conflict eliminated</td>
<td>386</td>
<td>255</td>
<td>53.4%</td>
<td>36.9%</td>
</tr>
<tr>
<td>Referential conflict reduced</td>
<td>280</td>
<td>112</td>
<td>38.8%</td>
<td>16.2%</td>
</tr>
<tr>
<td>Referential conflict unchanged</td>
<td>7</td>
<td>80</td>
<td>1.0%</td>
<td>11.6%</td>
</tr>
<tr>
<td></td>
<td>49</td>
<td>204</td>
<td>6.8%</td>
<td>29.6%</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>39</td>
<td>0.0%</td>
<td>5.7%</td>
</tr>
</tbody>
</table>

722  690

Table 8 – Reduction/elimination of referential conflict ascribed to gender

Perhaps the most striking thing about these results is that fact that, in Spanish, gender is unable to play a disambiguating role in 92% of references. In 53% of references gender is unnecessary, as the identity of the referent is spelt out with a full NP, but this still leaves 39% of references that could potentially be ambiguous, but which are unmarked for gender. Of the 8% of references which remain, gender only serves to eliminate referential conflict in 1% of them, with the remaining 7% benefitting from a reduction in conflict.

In Mian, gender serves to eliminate referential conflict in 12% of cases, and reduce it in 30% of cases, so it certainly plays a more significant role than it does in Spanish. Even so, 16% of references are unmarked for gender, and in a further 6% of references gender is marked but does not assist in reducing referential conflict.

In both Spanish and Mian, the data show how a reduction in referential conflict is more common that an elimination of referential conflict. Indeed, for 39 (6%) of the Mian references, even a reduction in referential conflict is not achieved (i.e the contribution of gender score is ‘0’). The breakdown of the values given for the reduction of referential conflict are presented in Figure 5, taking into account only the 56 (Spanish) and 323 (Mian) references where gender is able to play a role.
These figures clearly illustrate how a reduction in referential conflict can still leave a reference exposed to potential ambiguity. In Spanish, of the 56 references where gender is able to play a role – already a very small proportion of the total number of references – in the case of 29 (52%) of them the referential conflict is reduced by 55% or less. In Mian, of the 323 references where gender has the potential to play a role, in the case of 152 (47%) of them the referential conflict is reduced by 55% or less, or not at all.

Taking such stark figures into account, it is difficult to even contemplate that disambiguation might be the primary function of gender systems. At the same time, however, it is clear that gender can contribute to reducing referential conflict to varying degrees given the right conditions. It is probably better, therefore, to see gender as just one of the many factors at play in discourse, such as context, semantics, and so on (cf. Kibrik (2011: 390) on the ‘multiplicity of factors’ in discourse).

5.3 Contribution of classifiers in reducing or eliminating referential conflict

Finally we turn to the question of whether classifiers in Mian and Kilivila can be shown to contribute to reducing or eliminating referential conflict. In light of the significant differences between the two systems – that is, Kilivila has 177 classifiers, while Mian has a mere 6, and even those are restricted to occurring with only forty or so verbs – we might expect classifiers in Kilivila to play a far greater role in disambiguating references.

As with the analysis of gender in Section 5.2, it is first necessary to recognise that in the case of full NPs the reference is typically unambiguous. In addition to this, in both Mian and Kilivila, referential devices often occur without any classifier. The remaining references are the only ones where classifiers might be playing a disambiguating role. The figures are presented in Table 10.

<table>
<thead>
<tr>
<th>Kilivila</th>
<th>Mian</th>
<th>Kilivila</th>
<th>Mian</th>
</tr>
</thead>
<tbody>
<tr>
<td>Classifier unable to play a role (full NP)</td>
<td>386</td>
<td>255</td>
<td>49%</td>
</tr>
<tr>
<td>Reduced device with no classifier</td>
<td>372</td>
<td>416</td>
<td>47%</td>
</tr>
<tr>
<td>Reduced device with classifier</td>
<td>27</td>
<td>37</td>
<td>3%</td>
</tr>
<tr>
<td>Total</td>
<td>785</td>
<td>708</td>
<td>100%</td>
</tr>
</tbody>
</table>

Table 10 – Reduction/elimination of referential conflict ascribed to classifiers

These figures speak for themselves. Given that the number of reduced referential devices co-occurring with a classifier in Kilivila and Mian is 3% and 5%, respectively, this is fairly compelling evidence that classifiers are not contributing anything of significance to the avoidance of referential conflict.

6. Conclusion

Reference tracking is often claimed to be a function of nominal classification systems. By way of example, Huang (2000: 8) treats gender systems as one of “four major types of reference-
tracking system operating in discourse”, alongside switch-reference systems, switch-function systems and inference systems. This is not entirely surprising: firstly, there is no doubt that gender/classifiers can, and do, serve to disambiguate between referents when the conditions are right, perhaps leading us to conclude this is their raison d’être, and secondly, when faced with a linguistic phenomenon that might at first glance appear superfluous, we tend to want to ascribe a function to it.

What is conspicuously lacking in the literature, however, is any sort of systematic investigation which attempts to assess the validity of this often-cited function of nominal classification. This is not to suggest that the topic of reference tracking has been neglected – indeed, there are many studies that address this topic – but rather that reference tracking studies tend to be concerned with topics such as referential choice or referential density; these are very different questions to that of the role that gender/classifiers might play in referent disambiguation.

Of course, a study of an individual language, especially one with an elaborate gender or classifier system, may clearly demonstrate the disambiguating potential of that particular language’s nominal classification system, but such studies do not consider the bigger picture; that speakers of a given language use a resource which is at hand to disambiguate between referents does not answer the question of whether nominal classification, cross-linguistically, exists for this purpose.

The aim of this paper, then, was to bridge this gap by investigating the role, if any, that gender and classifiers play in referent disambiguation. The study presented in this paper differs from previous studies of reference tracking in a number of key ways:

(i) it is not concerned with topics such as referential density, but shifts the focus instead to nominal classification systems and the role they might play in reducing or eliminating referential conflicts (i.e. referent disambiguation);
(ii) an entirely new methodology was devised for the purpose of this study, one which not only allows for the results to be quantitative in nature, but also facilitates comparisons across languages;
(iii) it is typological in nature and so rather than focussing on languages with either gender or classifier systems – or indeed concurrent systems, as in this study – it views these in the context of a language which has neither, thereby allowing us to cover the typological space.

First, this paper looked at referential choice and how it correlated to how activated a referent was in the discourse. Although this was not the main focus of the study, it was a useful starting point. Here the results were much as anticipated, with full NPs preferred when the activation level was below 1, and reduced referential devices preferred when the activation level was 2 or above. Of the four languages in the study, however, Mian stood out in this regard, displaying the lowest number of full NPs across all levels of activation. One possible interpretation of this fact is that Mian, with its concurrent systems of nominal classification, allows for a higher level of reduced referential devices than the other three languages in the study. However, an alternative (and perhaps more plausible) explanation could simply be that Mian places greater cultural value on events than it does on identities and referents; in their study comparing Belhare (Sino-Tibetan, Nepal) with Russian, Stoll and Bickel (2009: 554) hypothesise that such cultural differences in discourse may explain the significantly lower referential density observed in Belhare. In a different study, involving
Belhare, Maithili and Nepali, Bickel proposes that speakers will be more concerned with the referents involved in an event if the syntax of their language “requires higher processing attention to NP information” (2003: 733) and when “there is no linguistic pressure to focus on referents [...] speakers will focus more on events”. So, while Mian stands out in the present study as the language with the lowest referential density, this fact must be treated with caution, as it may be unrelated to the presence of nominal classification.

In any case, lower referential density in itself does not imply that references in Mian discourse are less susceptible to referential conflict. To answer that question, we turn to the results of the investigation into the contribution of gender and classifiers. The results here were rather striking.

Considering first the role of gender, the results showed that the gender system of Spanish eliminated referential conflict for a mere 1% of references, and helped to reduce referential conflict for a further 7% of references. Notably, however, there were no instances where gender had no effect at all on reducing the referential conflict, rather the reason that only 8% of references could be shown to benefit from gender was down to the fact that 53% of references were realised as full NPs and 39% of references were simply not marked for gender (primarily in cases of zero reference).

In the case of Mian, gender served to eliminate referential conflict for 12% of references and reduced referential conflict for 30% of references. However, unlike in Spanish, there were 6% of references where gender was unable to aid in reducing the potential conflict, due to the gender marking being formally identical to all other competing referents. By way of comparison with Spanish, in the Mian data 37% of references were full NPs and 16% were not marked for gender.

Evidently, then, gender has a greater potential to disambiguate between conflicting referents in Mian than it does in Spanish, but when faced with a figure as low as 12% (= Mian references for which referential conflict has been eliminated) it is hard to claim that gender is doing much work at all. It is worth remembering here that the figures for ‘reduction in referential conflict’ cover a rather broad category that includes references whose referential conflict is only reduced by 11%. The figure of 30% (= Mian references for which referential conflict has been reduced) would be as low as 13% if we were to only consider those references whose referential conflict was reduced by 56% or more.

Turning to the role of classifiers, the results were even more striking. Of the 785 Kilivila references encoded, a mere 27 references (3% of all references) consisted of a reduced referential device involving a classifier. All other references were either full NPs (49% of all references), where even if a classifier were present it would not have served a disambiguating role, or reduced referential devices not involving a classifier (47%). A similar picture held for Mian, where only 37 (5%) of 708 references consisted of a reduced referential device involving a classifier. In light of these figures, it was deemed futile to calculate ‘contribution of classifier’ scores for these references.

The results of this study, then, cast light on a particularly contentious issue, namely, whether one of the purposes of nominal classification systems is to serve a reference tracking function. By deiving a novel methodology for investigating the role of gender and classifiers in discourse, and adopting a typological perspective, this paper has been able to demonstrate the surprisingly limited role that gender and classifiers appear to play in one area of reference tracking, namely referent disambiguation. And of course, one must not forget that, even in those cases where gender/classifiers can be shown to either reduce or eliminate referential conflict, this does not necessarily imply that speakers use these as referential aids. There are
many other factors at play in discourse which speakers have recourse to in order to identify referents (perhaps most importantly that of context) and there are many languages that manage just fine without any system of nominal classification.

This, of course, leaves wide open the question of what nominal classification systems exist for, but recent research into redundancy in cognition points to the possibility of an entirely different function of gender, that of easing language processing effort. Gibson et al. (2019: 399) explain that nouns which are more predictable in context, for example as a result of gender, are easier to produce and understand, since language processing effort is associated with predictability. Thus, the redundant marking of a feature like gender may serve to make the linguistic signal more robust to noise.

In conclusion then, while there are of course times when gender or classifiers may be shown to be serving a disambiguating role in discourse, it is perhaps better to view such cases as a convenient by-product of nominal classification systems, rather than one of their primary functions.
References


