

Acquisition of a transparent gender system:

A comparison of Italian and Croatian

3

4 Abstract

5 This study compares the acquisition of the gender system and gender values in two
6 languages with transparent gender: Italian and Croatian. The study focuses on the
7 different degrees of transparency between the two languages by taking into account the
8 extended nominal paradigm.

9 We have conducted an adjective elicitation task on a total of 60 monolingual Italian
10 and Croatian children divided in two age groups (Italian=3;0 and 3;10, Croatian= 2;10
11 and 4;2).

12 The results have revealed that the Italian gender system is mastered already by the
13 youngest child (age=2;6) and that the two gender values are acquired simultaneously;
14 whereas Croatian children show a significant difference in the error ratio between the
15 two age groups which indicates that the gender system is not yet acquired in the younger
16 group (average age=3;0). Additionally, the results suggest that feminine is the first gender
17 to be mastered in Croatian due to the regularity of its paradigm, and that neuter is the
18 most problematic gender for the children, likely due to its lower frequency and
19 syncretism with masculine in the case paradigm.

20

21 Keywords: gender acquisition, gender agreement, transparency, Italian, Croatian

22

23 1. Introduction

24 The aim of this article is to compare how Italian and Croatian monolingual children
25 acquire the gender system and the individual gender values therein. Here we are testing
26 for acquisition through adjectival agreement and will consider gender agreement of a
27 value to be acquired when the error rate is below 5%, i.e. at 95% of correctness.

28 In the current study, gender is considered an inherent property of the noun reflected
29 in agreement with other elements (i.e. determiners and adjectives) (Corbett, 1991). Both
30 Croatian and Italian have transparent gender systems and a rich morphology
31 (Gudmundson, 2010; Kovačević, Palmović, & Hržica, 2009). Having a transparent
32 gender system entails that the gender of the noun is evident from its phonological form,

1 as nouns have a gender assigned as it is a part of their lexical entry (Kupisch, Müller, &
2 Cantone, 2002). However, as stated by Audring (2014), all languages show some
3 divergence from transparency which means that we cannot consider transparency a
4 categorical property (binary distinction between a transparent and opaque system), but
5 we have to place it on a continuum. The different degree of transparency between the
6 two languages will be defined by taking into consideration the declension classes of the
7 nouns in each language as well as the full nominal paradigm (i.e. number for Italian,
8 number and case for Croatian), as gender is rarely marked on the noun itself but it is
9 present overwhelmingly elsewhere which makes it a complex feature in language
10 (Audring, 2014).

11 Italian has obligatory gender-marked articles that occur with the noun and thus add
12 to the transparency of the language, since articles are defined as the first and most
13 frequent syntactic cue for gender (Chini, 1995). Croatian does not have articles and
14 consequently also no obligatory determiners: the noun may be accompanied by gender-
15 marked demonstratives, possessives, or adjectives, but most of the time it is bare.
16 Additionally, Croatian has three gender values while Italian has two, which according to
17 Audring (2014) is one of the dimensions of the complexity of a gender system. Depending
18 on the degree of transparency of the gender values, this may affect the acquisition
19 process; but more importantly, the three Croatian gender values are not equally well
20 represented in child directed speech (CDS) with the neuter amounting to only 6% of
21 nouns (Kovačević et al., 2009). Another factor taken into consideration from Audring
22 (2014) is syncretism as we will also be taking into consideration the extended nominal
23 paradigms of nouns and adjectives in the two languages which include gender and
24 number for Italian, and gender, number, and case for Croatian.

25 Clear formal cues contribute to the transparency of a gender value and are thus a
26 very relevant factor for the acquisition of a gender system (Karmiloff-Smith, 1981; Levy,
27 1983). The gender-marked article adds to the transparency (in favor of Italian); the
28 infrequent occurrence of neuter in Croatian is also a factor that needs to be taken into
29 consideration. Overall, in the next sections we will compare the transparency ration of
30 the two languages and based on that make predictions regarding the timing and mastery
31 of gender in the respective languages.

1 The methodology consisted of a picture-based elicitation task. The children saw
2 images depicting referents of different grammatical genders and they were prompted to
3 describe them by using adjectives. The use of the correct agreement between the
4 adjective and the noun was considered an indication for the acquisition of gender. We
5 have decided to test the acquisition of gender through agreement, more precisely the
6 agreement of adjectives, as these are optional and agree in gender with the noun in both
7 languages, thus providing a comparable testing ground for the gender systems of the two
8 languages.

9 The results have shown considerable differences in timing of the mastery of gender
10 in the two language groups, and we discuss this based on the different degrees of
11 transparency.

12 The paper is structured as follows: In the next section we provide a description of
13 the gender system of the two target languages, and in section 3 we provide an overview
14 of the acquisition of gender followed by an overview of the acquisition of gender in the
15 two target languages. In section 4, we describe the aims of the current study and lay out
16 the research questions, following that the methodology is described (section 5). Section
17 6 focuses on the results, followed by a discussion (section 7) and conclusions (section 8).

18

19 2. The gender systems of Croatian and Italian

20 In the following sections we will describe the gender system of Italian and Croatian in
21 order to outline the level of transparency of each language. We will focus on the
22 following factors: number of genders and their distribution in the vocabulary and
23 corpora, the agreeing elements, the nominal paradigm and its syncretisms. According to
24 Audring (2014) syncretism adds to the complexity of a gender system because it reduced
25 the number of gender markers that provide unambiguous gender information.

26

27 2.1 The Italian gender system

28 The Italian gender system has two values, M and F, and the gender is expressed through
29 morphophonological properties of the noun ending (Chini, 1995). The distribution of the
30 two genders in the language is 60% of M nouns and 40% of F nouns (Costa, Kovačić,

1 Franck, & Caramazza, 2003). Italian nouns can be divided in various declension classes
2 based on the noun endings in the singular and plural.

3 The Italian gender system is considered transparent because the majority of Italian nouns
4 ends in *-o/-i* (sg./pl.) which signals M, or in *-a/-e* which signals F. According to Nocetti
5 (2002), M is the unmarked gender, and deriving F nouns is possible via suffixations when
6 the nouns are animate. Gudmundson (2010) conducted a corpus study on the LIP
7 corpus¹ (Voghera et al., 2014) and found that this (transparent) group of nouns constitutes
8 71,2% of tokens used. The third most frequent noun class amounting to 21% of the
9 corpus consisted of the *-e/-i* ending and is ambiguous with respect to gender as it
10 contains M nouns (*cane/i-dog/s*), F nouns (*volpe/i-fox/es*), or both depending on the
11 natural gender of the referent (*insegnate/i-teacher/s*). Chini (1995, p. 81) specifies that
12 only the first two classes are unambiguous when it comes to gender.

13 The full categorization of Italian nouns based on their endings along with their frequency
14 in LIP is summarized in table 1. Here, we are reporting the results as described by
15 Gudmundson (2010).

16 As we can see in table 1, the aforementioned transparent classes are denoted as A and B
17 for M and F gender respectively, and the others are opaque if the semantic or
18 morphological information is not present. Thus, she claims that the system is not
19 unambiguous and that it is only partially transparent. However, the frequencies in the
20 table show that classes A and B are the more numerous ones, and together with class C,
21 they comprise 92,68% of the corpus.

22

23

24

25

26

27

¹ Corpus of spoken Italian with diaphasic, diatopic and diamesic varieties.

Noun class	Gender	Ending sg/pl	Example	Translation	Frequency in LIP (%)
A	M	-o/-i	Libro/i	Book/s	39,80
B	F	-a/-e	Padella/e	Pot/s	31,40
C1	F	-e/-i	Volpe/i	Fox/s	21,49
C2	M	-e/-i	Cane/i	Dog/s	
C3	M&F	-e/-i	Insegnate/i	Teacher/s	
D	F	-á/-á	Abilitá	Ability	2,96
E	M	-a/-i	Problema/i	Problems/s	1,46
F	M	-cons/-cons	Camion	Truck	0,70
G1	M	-a/-i	Artista/i	Artist/s	0,44
G2	F	-a/-e	Artista/e	Artist/s	
H1	F	-i/-i	Analisi	Analysis	0,40
H2	M	-i/-i	Domani	Tomorrow	
I	F	-o/-o	Foto	Photograph	0,24
J	M	-í/-í	Lunedí	Monday	0,29
K	M.sg -F.pl	-o/-a	Uovo/a	Egg/s	0,25
L	M	-è/-è	Caffé	Coffe/s	0,09
M	M	-a/-a	Cinema	Cinema	0,07
N	F	-ù/-ù	Virtú	Virtue	0,02
Other					0,41

1 Table 1: The Italian declension classes

2

3 Additionally, when class C is broken down into groups based on derivational
4 morphemes, it is revealed that there are morphemes in this class that unambiguously
5 signal the gender of a noun, such as *-ione* (F) (*stagione*-season), *-tore* (M) and *-trice* (F)
6 respectively (*attore/attrice*-actor/actress), *-iere* (M) (*giocolliere*-juggler), *-ame* (M)
7 (*fogliame*-foliage), and *-udine* (F) (*solitudine*-loneliness); which according to
8 Gudmundson (2010) account for 52% of noun class C. Thus, even if the majority of noun
9 classes in Italian can be classified as opaque, when the size of the groups and the nouns
10 that appear there in is considered together with frequency, it is evident that the majority
11 of the nouns are transparent (Gudmundson, 2010).

12 The gender of the opaque nouns becomes explicit through agreement, thus the rich
13 inflectional system of the nominal domain in Italian should be a valuable resource for

1 the acquiring child. Elements that have gender agreement with the noun are: articles,
 2 determiners, adjectives, quantifiers, possessives, wh-words, relative clauses and the past
 3 participle (Chini, 1995; Gudmundson, 2010). Here, we will focus on the description of
 4 the article and adjectival systems: the paradigm of the former is crucial for our hypotheses
 5 while the latter is relevant because we used adjectives as test items in our task. We start
 6 by describing the article system because it is an obligatory marking of the noun.

7 Italian articles are marked for definiteness, number, and gender (Caselli, Leonard,
 8 Volterra, & Campagnoli, 1993). All the Italian articles are presented in table 2.

9

Gender	Singular	Example	Plural	Example
Definite				
M	il	Il libro	i	I libri
	lo	Lo zaino	gli	Gli zaini
	L'	L'albero		Gli alberi
F	la	La padella	le	Le padelle
	l'	L'ape		Le api
Indefinite				
M	un	Un libro, un albero	NA	
	uno	Uno zaino		
F	una	Una padella	NA	
	un'	Un'ape		

10

11 Table 2: The Italian article system

12

13 Italian articles are sensitive to the phonology of the initial syllable of the noun: *il/un* (M)
 14 and *la, una* (F) are the unmarked forms; *lo, gli,* and *uno* are used when the M noun starts
 15 with a s+consonant (e.g. *lo stallone/gli stalloni* -stallion/s) or with a z, x, j, pn, ps, gn, and
 16 sc (*lo zaino/gli zaini*-backpack/s) whereas *l'* (both for M and F nouns) and *un'* (for F) are
 17 used with nouns beginning with a vowel (e.g. *l'albero*-M 'tree', *l'anatra*-F 'duck')
 18 (Gudmundson, 2010; Pizzuto & Caselli, 1992). Chini (1995) defines the paradigm of M
 19 article as more complex than that of the F article, because it contains more allomorphs,
 20 and also because the unmarked form for 'il' does not end in -o as the typical M nouns
 21 do. The gender cue is however neutralized with 'l'' especially if the noun also belongs
 22 to the C-class (*l'ape*, 'bee').

The presence of the article is a crucial factor in our study because it is the first and most frequent syntactic gender cue. Moreover, it is the only one when the noun does not have transparent assignment (Chini, 1995, p. 93). Despite the presence of opaque declension classes in Italian, this frequent cue contributes to the transparency of the gender and, as we will see from some previous studies, contributes to facilitating gender acquisition.

Adjectives also show gender agreement and their ending morphemes are equivalent to what we have seen for the noun classes in table 1, taking *-o/-i* for M and *-a/-e* for F. However, not all adjectives show gender agreement: some adjectives agree only in number and have the *-e/-i*, while some do not have variation at all. According to Chini (1995) the group of adjectives displaying gender and number features is crucial for acquiring the declension rule for adjectives. The types of adjectival agreement are presented in table 3.

Adjective	Gender	Ending sg./pl.	Example	Translation
Class 1	M	-o/-i	Piccolo libro/ piccoli libri	Small books/s
	F	-a/-e	Piccola padella/ piccole padelle	Small pot/s
Class 2	/	-e/-i	Grande libro/ grandi libri	Big book/s
			Grande padella/ grandi padelle	Big pot/s
Class 3	/	/	Libro blu ⁴ / libri blu	Blue book/s
			Padella rosa/ padelle rosa	Pink pot/s

Table 3: The Italian adjectival classes

The presence of an adjective does not exclude the article, thus both the adjective and the article are valuable gender indicators in (1).

(1) a. *Il libro bianco*⁵

b. *La padella bianca*

The-M book white-M

the-F pot white-F

“The white book”

“The white pot”

⁴ Some adjectives can be placed pre- and post- nominally while others cannot, but this is beyond the scope of this paper so this is not a differentiating criterion for the current study. Adjective also agree with the noun both in attributive (as above) and predicative position, but again, gender is marked in both cases so we will not differentiate between the two.

⁵ Note that the Italian adjective can be placed in pre- and postnominal position and some variation to meaning may apply. However, this is not the focus of the current study as the adjective is gender marked in either position.

1 Summarizing, the majority of Italian nouns has a transparent gender assignment, But
2 because of obligatory article inflected for gender even this nouns are disambiguated.

3

4 2.2 The gender system of Croatian

5 Croatian has a rich inflection system and nouns are marked for gender, number and case.
6 The gender of the noun is an inherent property, while number and case are considered
7 morphological properties (Mihajlović, 2014). Croatian has three gender values: M, F, and
8 N. An inquiry of the 4000 most spoken (noun) lemmas in adult spoken language revealed
9 that 43.3% are M, 42.9% are F, and N takes up the remaining 13,7% (Vuletić, 1991)in
10 (Kovačević et al., 2009, p. 157).

11 The morpho-phonological properties of the noun act as a reliable cue for gender:
12 M nouns end in a consonant, e.g., *jelen* (deer), *prozor* (window); F nouns end in *-a*, e.g.,
13 *kuća* (house); and N nouns end in *-o* or *-e*, e.g., *nebo* (sky), *more* (sea). However, some
14 exceptions apply for M and F as there are M nouns that end in an *-o*, such as *pepeo*
15 (“ash”) or *ugao* (“corner”), but also ending in *-a* such as *gazda* (“boss”); and F nouns
16 ending in a consonant, such as *kost* (“bone”) (Barić et al., 2005).

17 Croatian has three declension classes, but these do not have a 1:1 distribution with
18 the three gender values. The criteria for grouping the nouns in declension classes is based
19 on the ending that the noun takes in the genitive singular: class *-a*, class *-e*, and class *-i*.
20 The first class includes M and N nouns, the second class includes mostly F nouns with
21 some M exceptions such as *gazda* (boss/owner), while the *-i* class contains only F nouns,
22 the ones ending in a consonant (Barić et al., 2005, p. 103). Here we can already see that
23 M and N nouns are in closer relation than the F nouns are. Regarding the *-i* declension
24 class, these nouns are opaque and they were excluded from the study.

25 In table 4 we present the full paradigm of Croatian declension. This includes the
26 three declension classes divided into their gender values, and a separation between M
27 animate and inanimate nouns since these differ in one point of the paradigm (accusative
28 case)⁷.

⁷ This sub-division of the M gender value is not expected to interfere with the current study because (i) the target of the elicitation task is in Nominative, (ii) the declension of adjectives does not differ based on the animacy criteria

Singular	Class -a			Class -e	Class -i
	M animate	M inanimate	N	F	F
NOM	Jelen (deer)	prozor (window)	nebo (sky)	Kuća (house)	Kost (bone)
ACC	jelen-a	prozor	nebo	kuć-u	kost
GEN	jelen-a	prozor-a	neb-a	kuć-e	kost-i
DAT	jelen-u	prozor-u	neb-u	kuć-i	kost-i
VOC	jelen-e	prozor-e	nebo	kuć-o	kost-i
LOC	jelen-u	prozor-u	neb-u	kuć-i	kost-i
INS	jelen-om	prozor-om	neb-om	kuć-om	kost-i
Plural					
NOM	jelen-i	prozor-i	neb-a	kuć-e	kost-i
ACC	jelen-e	prozor-e	neb-a	kuć-e	kost-i
GEN	jelen-a	prozor-a	neb-a	kuć-a	kost-iju
DAT	jelen-ima	prozor-ima	neb-ima	kuć-ama	kost-ima
VOC	jelen-i	prozor-i	neb-a	kuć-e	kost-i
LOC	jelen-ima	prozor-ima	neb-ima	kuć-ama	kost-ima
INS	jelen-ima	prozor-ima	neb-ima	kuć-ama	kost-ima

1 Table 4: The full Croatian declension system

2

4 In table 4 we can see the overwhelming syncretism between M and N (shaded in gray).
5 Syncretism is defined as the relation between two or more words in a paradigm that have
6 different morphosyntactic features but are identical in form (Matthews, 2007). The
7 syncretism of the *-ima* (pl) ending also extends to the F *-i* stem declension class in DAT,
8 LOC, and INS case. The fact that M and N nouns are syncretic in most of the paradigm
9 makes the gender of a noun opaque when expressed in these cases, and since there are
10 no obligatory determiners, the noun will often be bare. The F nouns from class *-e*,
11 included in this study, are only syncretic (same form and function of the suffix) with the
12 other two gender values in INS.sg and GEN.pl. We can see from table 4 that the F nouns
13 in class *-e* have also some other shared suffixes as the suffix *-a* is shared by NOM F and
14 ACC M, the suffix *-u* is shared between DAT M and ACC F, *-e* is shared between VOC M

1 and GEN F. But since words are rarely produced in isolation, it will be obvious from the
2 context, for example, whether a word is a DAT or an ACC. Thus, when case is
3 disambiguated, the gender should not be ambiguous either. Thus, even if there are
4 syncretism across the paradigm including all three genders, it is crucial to note that M
5 and N are part of the same declension class and that are syncretic within the same case,
6 whereas the syncretism between M and F is across case. Taking this into account, the F
7 (-e class) is the most transparent of the three genders in Croatian as it has a clearly
8 distinguishable gender marker across the paradigm. Thus, the Croatian system is
9 transparent when the base form (i.e. Nominative) is considered, but when we look at the
10 full paradigm ambiguities between M and N arise. Considering that the N is also
11 considerably less frequent than M, it is plausible that the children will initially perceive
12 N nouns as M. On the contrary, the acquisition of F should proceed quite straightforward.

13 When it comes to agreement, demonstratives, possessives, adjectives and some
14 verbs show gender agreement. However, since these are optional elements, the Croatian
15 noun is often bare, with only the morphology of the noun itself serving as a gender cue.

16 (2) a. *Djevojčica traži torbu.*

17 girl-F.NOM search-PRS bag-F.ACC.

18 "The/a⁹ girl is searching for a bag."

19 b. *Dječak gleda psa.*

20 boy-M.NOM look-PRS dog-M.ACC

21 "The/a boy is looking at a dog."

22 c. *Dijete voli more.*

23 child-N.NOM love-PRS sea-N.ACC

24 "The/a child loves the sea."

25 Of all of these elements we will focus only on the adjectival paradigm because adjectives
26 were elicited in our task. In table 5 the declension of the adjective *beautiful* is displayed
27 per gender. Adjectives agree with gender and are not dependent on the declension class

⁹ Both determiners are put in the translation because, due to the lack of an article system in Croatian, the definiteness value of the noun is ambiguous.

1 of the noun .

Singular	M		N	F
	short	long		
NOM	lijep-ø/	lijep-i	lijep-o	lijep-a
ACC	lijep-a/	lijep-og	lijep-o	lijep-u
GEN	lijep-a/	lijep-og	lijep-og	lijep-e
DAT	lijep-u	lijep-om	lijep-om	lijep-oj
VOC	/	lijep-i	lijep-o	lijep-a
LOC	lijep-u	lijep-om	lijep-om	lijep-oj
INS	lijep-im	lijep-im	lijep-im	lijep-om
Plural				
NOM	lijep-i		lijep-a	lijep-e
ACC	lijep-e		lijep-a	lijep-e
GEN	lijep-ih		lijep-ih	lijep-ih
DAT	lijep-im		lijep-im	lijep-im
VOC	lijep-i		lijep-a	lijep-e
LOC	lijep-im		lijep-im	lijep-im
INS	lijep-im		lijep-im	lijep-im

2 Table 5: The Croatian adjectival paradigm

3 As it is shown in table 5, Croatian has two adjectival forms for the singular M gender
4 value. The main distributional difference is that the short adjective can be in predicative
5 position (reserved only for the Nominative case) and the long one cannot, while both can
6 be in attributive position. In attributive position these mark a definiteness/specificity with
7 the long form being +DEF/SPEC and the short form -DEF/SPEC (Aljović, 2002; Trenkić,
8 2004). However, this sub-division of M adjectives should not affect the task at hand
9 because both forms signal the gender value and we are not interested in the correctness

1 of the pragmatic use. We can see once more how M (long) and N are neutralized in some
 2 places of the paradigm, but also how in the plural adjectives, gender distinction is barely
 3 made. However, we are currently no testing plurals.

4 Overall, the gender system of Croatian is transparent as the noun endings in the
 5 nominative singular quite unambiguously indicate the gender value of the noun; also the
 6 agreeing elements display a transparent gender marker. However, when the whole
 7 nominal paradigm is taken into consideration, the initially straightforward transparency
 8 becomes more limited: the M and N genders are revealed as much closer than initially
 9 shown by their declension in the nominative case, as syncretism of the declension is
 10 present in four out of the seven cases. F is the most transparent gender, as the syncretism
 11 it has with the other genders is within case only in INS in the nominal paradigm and
 12 never within case in the adjectival paradigm.

13

14 2.3 Main differences between Italian and Croatian

15 Here we will concisely summarize the crucial differences between Italian and Croatian
 16 which relevant for the predictions of our research questions (section 4).

17

	Number of gender values	Even distribution of gender values in the language	Obligatory determiner	Opaque nouns (assignment)	Syncretism in the paradigm
Italian	2	yes	yes	yes	limited
Croatian	3	no	no	no (yes) ¹¹	extensive

18 Table 6: Crucial differences between Italian and Croatian that can influence the degree of gender
 19 transparency

20 In the table above the facilitative factors of transparency are shaded in gray, and it is clear
 21 that Italian is the more transparent gender system: it has less gender values and these are
 22 more evenly distributed, it has the obligatory gender marked determiner (the article), and

¹¹ The *-i* class F nouns are opaque, but not part of the current study

1 the syncretism of the article is limited to the definite article /' used when the noun begins
2 in a vowel. Both languages have a set of nouns that are considered opaque: for Croatian
3 this is the -i declension class, but it is excluded from the task in this study giving Croatian
4 a less opaque gender assignment system, while Italian has what we have referred to as
5 C-class of nouns. The gender of the latter ones is however clearly disambiguated with
6 the obligatory gender-marked article. The Italian nominal paradigm is more limited than
7 the Croatian one because it does not include case. Thus, when comparing Croatian noun
8 to Italian ones only in the nominative, the two systems seem transparent to the same
9 degree. However, when all cases in Croatian are taken into account, it is revealed that
10 the distinction between M and N is neutralized across some cases, leaving F the least
11 ambiguous gender value. Additionally, the Italian M and F values are quite evenly
12 distributed in speech (Costa et al., 2003), as are in Croatian, however the N here is
13 significantly less frequent than the remaining two gender values (Kovačević et al., 2009)
14 which could affect the timing that it takes to be mastered.

15

16 3. The acquisition of gender: a general overview

17 From an acquisition perspective, there are two types of cues that contribute the
18 acquisition of gender: formal and semantic cues. In the previous sections we have
19 outlined the morpho-phonological cues for gender; these are considered formal cues.
20 Conversely, semantic cues correlate to the natural gender of the referent. Research on
21 various languages, such as Hebrew (Levy, 1983), French (Karmiloff-Smith, 1981), Russian
22 (Rodina, 2008), and Spanish (Pérez-Pereira, 1991) has shown that children rely more on
23 formal than semantic cues, so the latter will not be further discussed. Hence, semantic
24 cues will not be discussed further. Thus, if the formal cues are clear, i.e. transparent,
25 children should pay more attention to those. In this study we consider a gender
26 assignment system to be transparent when the phonological or morphological assignment
27 of gender allows for an accurate inference of the gender of the noun without having to
28 rely on agreement on other arguments; as a high complexity of the nature and number
29 of assignment rules may lead to a greater difficulty for the acquiring child (Audring,
30 2014). An example of a transparent system is Spanish with the final vowel of the noun
31 signaling its gender: -o for M and -a for F (*armario*-M 'closet', *mesa*-F 'table'), much like
32 Italian that was already described in section 2.1; whereas languages like Norwegian are

1 deemed to be an opaque system as the noun itself offers no, or very little, indication of
2 the gender (*stol*-M 'chair', *seng*-F 'bed', and *skap*-N 'closet'). It is however clear from the
3 description of the two systems described in the previous section, that Italian and Croatian
4 are transparent gender systems.

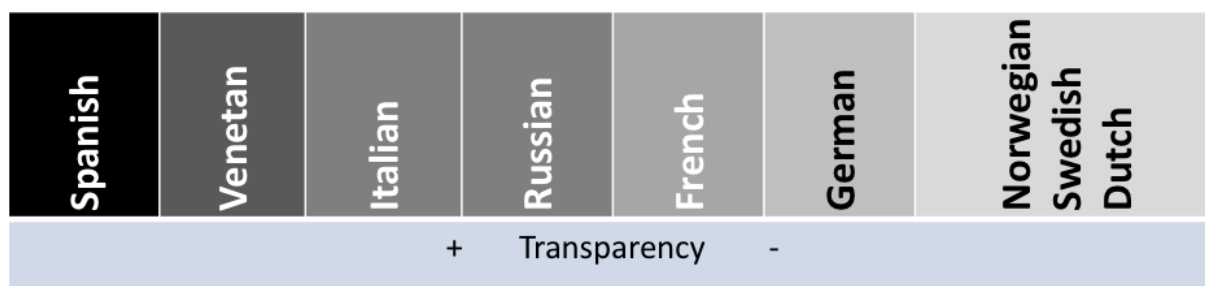
5 Other arguments receive gender by agreement. As Corbett (1991) states, in order to
6 establish the gender of a noun we have to use agreement as a test, since agreement is the
7 way in which gender is realized in language use. So, nouns have gender by assignment
8 (i.e. they are assigned a gender value in the lexicon), while the elements that agree with
9 the noun receive gender through agreement (Kupisch, Akpınar, & Stöhr, 2013). Unlike
10 assignment, a rich agreement system facilitates acquisition as languages in which a lot of
11 elements agree with the noun (e.g. Spanish, French) is mastered earlier than a system
12 with few agreement markers (e.g. Dutch) (Audring, 2014).

13 Previous studies have correlated the time course in the acquisition of the gender
14 system with its transparency: if the gender assignment system is transparent it will be
15 acquired more easily; as this property facilitates early use of correct agreement (Kupisch
16 et al., 2002; Rodina & Westergaard, 2015). On the other end of the spectrum, in
17 languages with opaque gender systems, this property is acquired late, such as in
18 Norwegian (Rodina & Westergaard, 2015).

19 Here the acquisition of a gender system is observed through agreement (of the
20 adjective) as has been done in previous studies (Karmiloff-Smith, 1981; Kupisch et al.,
21 2002; Rodina, 2008). Pérez-Pereira (1991) conducted a study on 160 Spanish children
22 of a wide age range (4-11-year-olds) that tested the semantic, morphological, syntactic
23 cues in conditions where two of the factors either indicated the same gender or were in
24 conflict. Converging cues made it easier to determine the correct gender, but in conflict
25 conditions, children payed more attention to the syntactic (agreement) than to the
26 morphological (assignment) cue. If this finding transpires cross-linguistically, Italian
27 children can have an advantage over Croatian children since the obligatory article in
28 Italian is an almost omnipresent syntactic cue for gender. Pérez-Pereira (1991) also found
29 that children tended to attribute M to F, which might be because of the unmarked status
30 of M (Pérez-Pereira 1991, p.584). If M has the same status in Italian and Croatian, we
31 can expect a similar outcome.

32 Transparency is not a binary feature and comes in degrees, therefore previous

1 studies have found pivotal differences in the acquisition of two languages with fairly
 2 similar gender systems. Smoczyńska (1985) compared Polish and Russian, and found
 3 that the Polish gender system is more easily acquired. The reason for this is the different
 4 degrees of transparency between the two languages. Rodina and Westergaard (2017)
 5 define two classes of opaque nouns in Russian: N nouns ending in an unstressed -o are
 6 indistinguishable from the typical -a ending of F nouns making the nouns in question
 7 ambiguous between F and N, and nouns ending in a palatalized consonant which may
 8 belong either to the M or the F gender. An effect of the degrees of transparency was also
 9 found between French and Italian as gender in Italian is acquired earlier than in French
 10 (Kupisch et al., 2002) because the formal regularities are more reliable in Italian. The
 11 French system is not considered opaque, as in both languages gender is unambiguously
 12 marked only on determiners but Italian has a more reliable morphology (Kupisch et al.,
 13 2002). Thus, also in the current study, we will look at degrees of transparency. Kupisch,
 14 Geiß, Mitrofanova, and Westergaard (2018) have already used this kind of approach to
 15 study the acquisition of gender in Russian children acquiring German and have
 16 developed a continuum of gender transparency that places these languages in the context
 17 of others (figure 1).



18
 19 Figure 1: Transparency scale of languages from Kupisch et al. (2018)

20 If Croatian were included in this scale, it would appear close to Russian. However, based
 21 on our description of the Russian opaqueness seen above and the general lack
 22 opaqueness in the nominative case in Croatian (2.2), we deem Croatian to be more
 23 transparent than Russian and thus placing it in between Russian and Italian, entailing that
 24 assignment in Italian is more transparent than assignment in Croatian. This entails that
 25 Italian and Croatian are not only both transparent languages when it comes to gender but
 26 are also very similarly transparent as are placed next to each other on the transparency
 27 scale.

1 3.1. The acquisition of the Italian gender system

2 For Italian, it has generally been reported that the Italian gender system is acquired in a
3 short time and with very few errors in the article system (Kupisch et al., 2002). When it
4 comes to the distribution of declension classes in child speech and child directed speech
5 (CDS), De Marco (2005) found that the two transparent classes are the most represented
6 ones, with nouns of what we referred to as the C-class being less frequent. No reports on
7 the other, much less frequent declension classes, are made in this study. Thus De Marco
8 (2005) found that A and B-class nouns (singular) amount to 63% in CDS (average across
9 the corpus), and to 66% in child speech; whereas the C-class nouns are produced at 11%
10 and 15%, respectively, by each group. Noccetti (2002, p. 62) also reports that the A and
11 B declension classes (to which she refers to as *productive microclasses*) are the most
12 frequent ones both in CDS, in the child's utterances, and also in adult to adult speech.
13 She also provides an analysis which shows how the proportion of noun types and tokens
14 divided per gender is fairly similar in the child's productions and in CDS.

15 When it comes to the opaque nouns, various studies (Belletti & Guasti, 2015;
16 Caselli et al., 1993; Kupisch et al., 2002; Noccetti, 2002) report that also these are
17 acquired easily. A corpus investigation conducted by Kupisch et al. (2002) revealed that
18 90% of all noun tokens belong to the two noun classes defined as transparent, and
19 additionally, in CDS, diminutive suffixes are used rather often which makes the gender
20 of the noun available also on C class nouns (i.e. *il cane/il canino*- "the dog/doggie", *la*
21 *volpe/la volpicina*- "the fox/foxy"). The Italian monolingual corpus investigated by
22 Kupisch et al. (2002) contained very few gender errors (1.9%) and they conclude that
23 gender in Italian is mastered early without significant deviations from the target gender.
24 According to Kupisch et al. (2002) articles have a relevant contribution for the early
25 acquisition of gender in Italian, because transparency reinforces the early acquisition of
26 articles, and once they are acquired it is an additional facilitation for the acquisition of
27 gender. This is also found in Caselli et al. (1993) as children produced articles in the
28 same proportion with all noun classes. However, Noccetti (2002) provides a more
29 detailed analysis of a corpus of a single child and discusses a more sequential acquisition
30 of Italian gender. In the *premorphology phase* (2;00-2;3) of gender acquisition the child
31 makes some assignment errors (i.e. gender mistakes in the plural) and the nouns rarely
32 have agreement, i.e. the article is omitted. But by the end of this phase, there are no

1 longer assignment errors in the number paradigms. In the *protomorphology phase* (2;3-
2 2;8) mistakes in agreement of C-class nouns are present as they are generalized to M;
3 however, the B-class nouns (F gender) are productive, for example the child pluralized a
4 proper name. In the final phase, *modularized morphology* (2;8-3;4), the child uses the
5 paradigms and agreement of all genders correctly. Thus, even though there are obviously
6 stages in the acquisition of Italian gender, we can see from Nocetti's (2002) data that
7 these phases are quite brief which probably why there were not observed by other,
8 perhaps less detailed, studies.

9 Thus, the presence of the article is thus crucial as it can be regarded as a kind of
10 gender morpheme (Chini, 1998). Accordingly, Pizzuto and Caselli (1992) found based
11 on corpus data of three children that the F article *la* is the first one to be attested (age=1;6).
12 The errors of commission observed in the study are related to the phonological choice of
13 M article (use of the default 'il' when 'lo' is required) and to the misinterpretation of
14 gender in C-class nouns (Pizzuto & Caselli, 1992). Pizzuto & Caselli's (1992) analysis
15 shows that even if the articles are attested early, only 'la' can be considered acquired by
16 all three children in the corpus. Moreover, what they conclude is that none of the major
17 inflectional paradigms that they investigated (articles, pronouns, clitics, and verbs) is fully
18 mastered by the age of 3;0. Thus, article as a gender cue certainly contributes to acquiring
19 the gender system, it does not entail that the system is immediately mastered. Therefore,
20 investigating whether gender is productive on agreeing elements other than articles can
21 give us insight into when the system is acquired. For example, Caselli et al. (1993)
22 conducted an elicitation task on three age groups and found that 'la' has a better accuracy
23 than 'il' in the youngest group, but agreement on adjectives is at ceiling in all three age
24 groups (Caselli et al., 1993, p. 384). According to Caselli et al. (1993) Italian grammatical
25 morphemes reach full mastery at ages 3;6-4;0.

26 Bottari, Cipriani, and Chilosi (1993) analyzed the production of monosyllabic place
27 holders that emerge before lexical items and that the correct morpheme 'la' form the F
28 article is realized before its M counterpart. They explain this in relation to 'la' having
29 vowel features analogous the noun, which the M article does not have. This is
30 strengthened by the fact that the plural article 'le' is also learned before 'il', in spite of
31 the lower frequency of the plural in the input. They report that the full version of 'il' and
32 'la' take over the place holders at 3;1 and 3;8 and thus claim that there are two stages

1 that lead to a correct production of articles: in the *linear agreement strategy* the
2 placeholder agrees with the noun ending, therefore -a is used for F, while an approximate
3 vowel is used for M. Thus, 'il' is acquired when *morphological insertion strategy* takes
4 place (Bottari et al., 1993)p.363 when the right morpheme has to be inserted.

5 Previous studies on Italian cited in this section have reported an early mastery of
6 the gender system, even for the classes that are deemed less transparent, because of a
7 frequent gender cue in the form of an article.

8

9 3.2. The acquisition of the Croatian gender system

10 The acquisition of Croatian is overall understudied, and thus reference to numerous
11 studies is not possible. Croatian is a language with rich derivational morphology which
12 enhances the transparency of lemmas, and this morphological richness could act as a
13 booster for the acquisition of the inflectional paradigms (Kovačević et al., 2009).
14 Kovačević et al. (2009) have analyzed corpus data from one of the children present in
15 the Kovačević (2004) corpus from ages 1;3-2;8. The study focuses mostly on the child's
16 incremental acquisition of the case system and only marginally looks at gender, focusing
17 mainly on the use of nouns in relation to the three genders rather than the gender
18 markings on agreeing elements. They noticed that the distribution of the three genders
19 reflects the distribution of the nouns in child directed speech of the same corpus: 35%
20 M, 59% F, and 6% N. Consequently, the child produced mostly F nouns (58%) followed
21 by M (36%) and N (6%). Other than that, there are no notes on the gender agreement of
22 these nouns, nor are there any indications of how accurate agreement is. However, in
23 relation to the acquisition of the case paradigm, the authors state that the first opposition
24 of inflected forms are found in regular F nouns between the Nominative and Accusative
25 case (Kovačević et al., 2009). They find that the child grasps the complex case system
26 easily and is using all seven cases, with different frequencies, at age 1;10. This implies
27 that all the children that participated in our task should be able to use the full
28 declensional paradigm. The corpus results from Kovačević et al. (2009, p. 165) report the
29 nominative and accusative reaching together 78% of the child's production at age 2;5.
30 This prevalence of the two aforementioned cases is also evident in the input. Recalling
31 table 4, NOM and ACC are not syncretic at any level, and considering the very low
32 frequency of N nouns (6%), Croatian children might still be at the stage in which the

1 syncretism between M and N is not evident. Since the corpus contains data only until
2 2;8, evidence of a more distributed use of the case paradigm or a more frequent use of
3 N nouns is not provided. We might assume that the neutralization of M and N becomes
4 more evident with increased exposure and usage of the full case paradigm, and we can
5 also speculate that this might be reflected in the acquisition of gender, namely that it will
6 take children longer to master M and N because it will take them longer to become aware
7 that these nouns belong to different genders.

8

9 4. The current study

10 The importance of this study lies in the fact that we are comparing two languages that
11 have transparent gender systems, which has been repeatedly found to be the key to an
12 early acquisition of gender (Levy, 1983). Based on this premise, both Italian and Croatian
13 children should have an early grasp of their respective gender systems. As Caselli et al.
14 (1993) put it 'it is widely accepted that inflectionally rich languages promote an early
15 acquisition of morphology', including Italian in this list of languages. Kovačević et al.
16 (2009) also supposed that the morphological richness of Croatian may promote the early
17 development of inflection.

18 However, the two target languages differ in other relevant features which may affect
19 the timing of acquisition of the system and the individual gender values, even if gender
20 is acquired early in both groups. We thus expect a high rate of correct agreement, but
21 the intra-language differences are expected to affect the timing of and perhaps the order
22 in which a particular gender is mastered. Thus, the relevance of this study lies in the fact
23 that we are analyzing how different factors in transparent languages can affect gender
24 acquisition. Additionally, the acquisition of the gender system in Croatian is relatively
25 new territory, with only one study taking it into consideration (Kovačević et al., 2009),
26 but not from the perspective of agreement. It will thus offer a starting point for further
27 investigation of the acquisition of gender in Croatian, according to which more specific
28 research questions could be tackled through more precisely designed tasks.

29 We have decided to elicit adjectives because the adjectival paradigms because of
30 their optional status in both languages and the fact that in both languages the declension
31 paradigm is regular and thus offers good grounds for comparison between the two
32 languages. The study aims to answer the following research questions:

- 1 1. Are Croatian children slower than Italian children to acquire the gender system?
- 2 2. Is the most regular gender (feminine) acquired first in both languages?
- 3 3. How do the acquisition paths differ in the two languages?

4
5 Based on the properties of the two gender systems and what has been found so far in
6 previous studies, we predict that Croatian will be acquired more slowly when compared
7 to Italian. This prediction is primarily based to the syncretic distribution of infections of
8 M and N in some parts of the case system (table 4) accompanied by the low frequency
9 of N nouns in CDS. Whereas in Italian the presence of the obligatory gender-marked
10 article is a relevant cue and children have been found to acquire gender
11 unproblematically.

12 If regularity and transparency is a relevant factor, F should be the gender that is
13 acquired first or with more accuracy in both languages. Recall from sections 2.1 and 2.2
14 that F is the more regular gender in both languages as in Italian the F article is more
15 regular than M due to its simpler realization, and in Croatian F nouns do not have within-
16 case syncretism with other gender values, while M and N show extended similarities.

17 We expect the three Croatian genders to be acquired in different stages, as we
18 expect the low frequency of N to result in a later acquisition of this gender value.

19 Overall, this study will compare the acquisition of two easily acquirable gender
20 system and it will thus reveal the nuances of how the gradience of gender transparency
21 affects accuracy.

22 23 5. Methodology

24 The task consisted of adjective elicitation using images of animals and objects denoting
25 referents of different genders. We have chosen to elicit adjectives since they are optional
26 and agree in gender with the noun in both languages.

27 28 5.1 Participants

29 The research was conducted on two samples of children, native speakers of Italian and
30 native speakers of Croatian (n=60, 27 male, 33 female). Each language group (n=30)
31 included two age groups of 15 children each. The children that were chosen for this
32 research were required to be native and monolingual speakers with both parents of either

1 Italian or Croatian nationality. The parents/caregivers were informed about the testing
2 and had to sign a consent form in order for their child to participate.

3 The age balance across the language groups was not entirely possible due to the
4 availability of children and parental consent. The language groups are still close in age
5 and we are able to make relevant statistical comparisons. The mean age of the younger
6 Croatian group (n=15, 6 male) was 2;10, whereas the older group (n=15, 10 male) had
7 a mean age of 4;2. The mean age of the younger Italian group (n=15, 3 male) was 3;0,
8 and the older group (n=15, 8 male) had a mean of 3;10. This means that the age range
9 of the Croatian children is wider than the one of the Italian participants, which is also
10 convenient as we expect this group to acquire gender at a slower pace than the Italian
11 children and with a wider age range, we will be able to observe this potential difference
12 more clearly.

13 Two participants were excluded from the younger groups due to their general lack
14 of adjective production; both children, one from the Italian and one from the Croatian
15 group, were 2;3. This does not indicate that gender is not or cannot be acquired before
16 age 2;6, but simply that children are not producing enough adjectives at this age.

18 5.2 Materials

19 Thirty images depicting animals and inanimate objects were used. The images were
20 downloaded from the Internet from open source websites. These images were selected
21 on the basis of the grammatical gender of the depicted noun in Italian and Croatian. This
22 means that in Italian fifteen images represent an M noun and the other fifteen represent
23 an F noun. In Croatian, the same images were distributed in a different manor across
24 genders: ten are M, ten are F, and ten are N. The test items were chosen for both
25 languages parallelly tracking the genders they were assigned to in each language. So for
26 example, when the item *tree* has been chose as stimuli, it was counted as an M noun for
27 Italian and N for Croatian, conversely the item *snail* took a F slot for Italian and M for
28 Croatian, the item *cat* took a M slot in Italian and a F slot in Croatian. As we went along
29 with the choice of the stimuli, these had to fill specific gender slots in both languages
30 depending on how many were left for each gender. Unfortunately, due to matching the
31 referents to the genders cross-linguistically, the Italian C-class nouns (n=6) are contained
32 only within M. However, as the previous literature has stated, the article is a very reliable

1 marker of gender (Chini, 1995) and Italian children have been found to master these
2 types of nouns easily (Belletti & Guasti, 2015).

3 All the items used are displayed in table 6 and are divided for language and gender.
4 Due to a considerable amount of N referents denoting abstract concepts, such as *sky* had
5 to be included. We did not explicitly check for frequency of these nouns in the two
6 languages, but during the piloting of the task all the participants were able to name all
7 the images included. The only somewhat problematic item was 'wing', but it could not
8 have been replaced due to a low availability of nouns that were N in Croatian and F in
9 Italian.

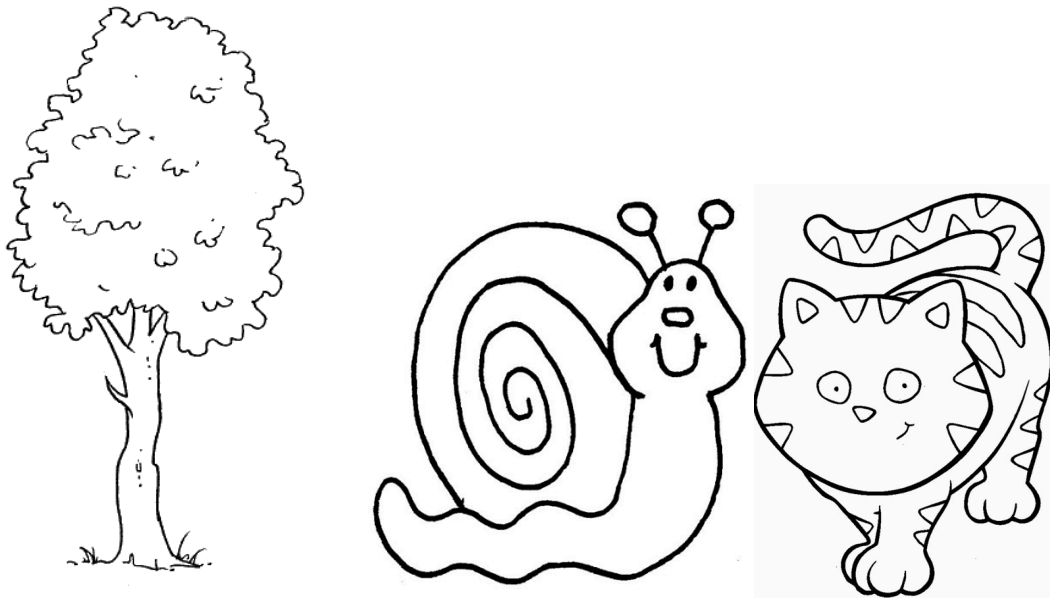
10
11

English translation	Italian noun	Gender in Italian	Croatian Noun	Gender in Croatian	
Lion	Leone	M	Lav	M	1
Frog	Rana	F	Žaba	F	2
Sun	Sole	M	Sunce	N	3
Grapes	Uva	F	Grožđe	N	4
Penguin	Pinguino	M	Pingvin	M	5
Snail	Lumaca	F	Puž	M	6
Tree	Albero	M	Stablo	N	7
Cat	Gatto	M	Mačka	F	8
Giraffe	Giraffa	F	Žirafa	F	9
Sheep	Pecora	F	Ovca	F	10
Sea	Mare	M	More	N	11
Dog	Cane	M	Pas	M	12
Butterfly	Farfalla	F	Leptir	M	13
Apple	Mela	F	Jabuka	F	14
Egg	Uovo	M	Jaje	N	15
House	Casa	F	Kuća	F	16
Mouse	Topo	M	Miš	M	17
Star	Stella	F	Zvijezda	F	18
Heart	Cuore	M	Srce	N	19
Bull	Toro	M	Bik	M	20
Ladybug	Cocinella	F	Bubamara	F	21
Moon	Luna	F	Mjesec	M	22
Eye	Occhio	M	Oko	N	23
Monkey	Scimmia	F	Majmun	M	24
Flower	Fiore	M	Cvijet	M	25
Pear	Pera	F	Kruška	F	26
Ear	Orecchio	M	Uho	N	27
Candle	Candela	F	Svijeća	F	28
Sky	Cielo	M	Nebo	N	29
Wing	Ala	F	Krilo	N	30
					31
					32
					33
					34
					35
					36
					37
					38
					39
					40
					41
					42
					43
					44
					45
					46
					47
					48

49 Table 6: Distribution of the test materials across the genders in the two languages

1
2
3
4

All the images were printed in black and white on thick paper. Some of the example images are provided below.



5
6
7

Image 2: Some of the images used in the task: Tree- M in It, N in Cro; Snail. F in It, M in Cro; and Cat- M in It, F in Cro

8
9

5.3 Procedure

10 All the participants were interviewed individually. Interviews were held in a slightly
11 isolated area of the kindergarten in order to avoid acoustic interferences. Interviews
12 lasted from five to fifteen minutes.

13 Each child was asked to describe what was represented in the image. The children
14 were explicitly asked before the testing to say whether the referents of the images were
15 *beautiful, good, small*¹² etc. until it was clear to the participant that the description had
16 to include an expression of a quality of the shown object. After making sure that the
17 children understood the task, the testing proceeded. The images were shown in a random
18 order, which was accomplished by taking them from a bag, and the participant was asked
19 to describe them. The interviews were recorded with a digital recorder Panasonic RR-
20 US430 either held in the interviewer's hand or laid on the desk. The uttered adjectives

¹² All of these adjectives agree in gender in Italian

1 were manually transcribed on a paper during the interview, and this was then cross-
2 referenced with the recording.

3 When the child produced incorrect agreement, the question was repeated in order
4 to establish if that was just a simple distraction or if it was a non-target-like gender
5 production. If the child made a correct agreement the second time, the response was
6 considered correct, otherwise it was scored as incorrect.

7

8 6. Results

9 First, we will look at the distribution of correct responses in the data divided per group.
10 An answer was counted as correct if the adjective gender matched the gender of the
11 noun, it was incorrect if it did not. The non-applicable (NA) category includes the
12 following cases: no response (n=155), no adjective produced (n=54), and use a non-
13 gendered adjective (only for Italian children) (n=35).

18

Target gender	Correct	Incorrect	NA	Total
M	165 (73%)	/	60 (27%)	225
F	166 (74%)	4 (2%)	56 (25%)	225

19 Table 7: Distribution of correct/incorrect answers in the Italian younger group

20

Target gender	Correct	Incorrect	NA	Total
M	205 (91%)	/	20 (9%)	225
F	201 (89%)	/	24 (11%)	225

21 Table 8: Distribution of correct/incorrect answers in the Italian older group

22

Target gender	Correct	Incorrect	NA	Total
M	123 (82%)	8 (5%)	19 (12%)	150
F	117 (78%)	8 (5%)	25 (16%)	150
N	110 (73%)	17 (11%)	23 (15%)	150

23 Table 9: Distribution of correct/incorrect answers in the Croatian younger group

24

Target gender	Correct	Incorrect	NA	Total
M	137 (91%)	6 (4%)	7 (5%)	150
F	145 (97%)	1 (0,06%)	4 (3%)	150
N	135 (90%)	9 (6%)	6 (4%)	150

25 Table 10: Distribution of correct/incorrect answers in the Croatian older group

26 We can see that both language groups had a high accuracy rate and we can safely assume
27 that they are aware of the nominal category of gender and use it accordingly. However,

1 it is not possible to see whether the displayed results match our acquisition criteria of
2 95% correctness since the NA data is also present in the tables.

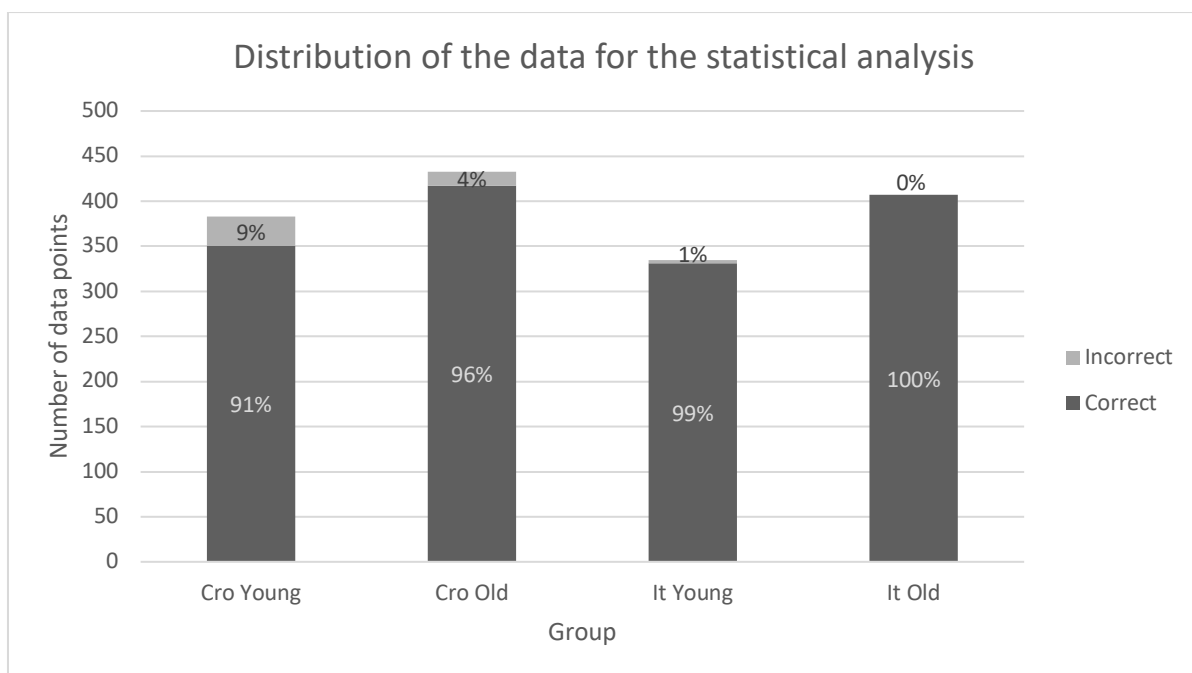
3 At first glance, it seems that the Croatian groups produce more errors. Subsequent
4 statistical analyses will show whether this difference is significant. The Italian children
5 have more NA data than the Croatian groups, however this is not due to their low
6 response rate, but to the fact that some adjectives in Italian do not agree in gender and
7 these errors amount to a total of 35 items in the dataset.

8 In the following sections we will address the research questions laid out in section
9 4 more specifically. However, for the rest of the analyses we will focus on the binary
10 distinction consisting of correct/incorrect gender agreement, and we will thus not
11 consider the NA responses.

12

13 6.1 Timing of gender mastery

14 In figure 3 the data points used in the subsequent statistical analyses are displayed. The
15 raw number of occurrences is displayed on the y-axis, but the stacked columns contain
16 a percentage of the raw data.



17

18 Figure 3: Distribution of the data for the statistical analysis

19

20 Looking purely at the percentiles and taking into account the 95% correctness criterion,
21 it seems that the Italian children have acquired gender already in the younger group

1 whereas the Croatian children have not. Nonetheless, figure 3 does not tell us the
 2 significance of these different error ratios and does not show the error ratio per gender
 3 value. Thus, a statistical analysis is in order.

4 We have conducted a linear regression using Jamovi (Love, Dropmann, & Selker, 2018).
 5 Here, group (young vs. old), language (Croatian vs. Italian), and target gender (M, F, and
 6 N¹³) were set as factors. The outcome of the analysis is presented in table 11.

7

Predictor	Estimate	SE	t	p
Intercept	0.95704	0.00504	190.006	< .001
old – young	0.03154	0.00904	3.489	< .001
croatian – italian	-0.03623	0.01004	-3.610	< .001
F – M	0.00138	0.00992	0.139	0.890
N – M	-0.05309	0.01411	-3.763	< .001

Table 11: Linear regression of all groups

8 The intercept was set to the values of *young*, *Italian*, and *M*. The fact that the Intercept
 9 shows a highly significant value means that the distribution of correct and incorrect
 10 responses is significant, and from the raw data in tables 7-10, it is obvious that correct
 11 responses were predominant in the task. Furthermore, the significance at the group and
 12 at the language level shows that the four groups we tested show significant differences in
 13 their responses, however the nature of these differences is not yet evident from the current
 14 analysis. The fact that there is no difference in the responses to the F and M gender means
 15 that the children master it to the same extent in both languages; the difference between
 16 the N and M refers only to the Croatian group and we can see that the children are
 17 significantly less accurate with one of the genders. In light of the raw data, it is evident
 18 that N agreement is less accurate, but more precise statistical analysis will reveal whether
 19 this difference is significant.

¹³ Note that N was a value only available in the Croatian groups.

1 The next step in our analysis is to look at the language groups by conducting a
 2 linear regression on the Italian and Croatian groups separately. This separation will shed
 3 light on the source of the significant difference in table 11. We will first look at the Italian
 4 group (table 12) and then proceed to the Croatian group (table 13).

5

Predictor	Estimate	SE	t	p
Intercept	0.9940	0.00269	369.49	< .001
old – young	0.0118	0.00538	2.19	0.29
F – M	-0.106	0.00536	1.98	0.48

6 Table 12: Linear regression of the Italian groups

7

8 Apart from the intercept; the statistical analysis does not find any significant differences
 9 in the distribution. This means that there are no significant differences between the two
 10 age groups of Italian children and that both genders are acquired equally well. This
 11 means that the Italian children master gender by the age of 2;6, which is the age of the
 12 youngest participant taken into consideration. The task does not offer insight into gender
 13 mastery prior to 2;6.

14 The same analysis was conducted for Croatian children and is presented in table
 15 13.

Predictor	Estimate	SE	t	p
Intercept	0.9384	0.00826	113.652	<.001
old – young	0.0491	0.01652	2.971	0.003
F – M	0.0171	0.02017	0.846	0.398
N – M	0.0624	0.02022	3.084	0.002

16 Table 13: Linear regression of the Croatian groups

17

18 As in the previous analyses, the significant difference in the intercept means that there
 19 are significantly more correct answers, which was already evident from the raw data. The
 20 Croatian group also has a significant age difference which means that the correct answers
 21 increase significantly with age ($p=0.003$). Again, there is no difference between M and

1 F, but there is a significant difference between M and N ($p=0.002$). This suggests that N
 2 is acquired later than the other two genders in Croatian. In order to check for this, we
 3 have to make further analyses. We have thus tested another dataset from which we
 4 excluded N (table 14). If the difference in the age groups is no longer significant, it would
 5 mean that the low accuracy of N is the sole reason for the observed difference in table
 6 13.
 7

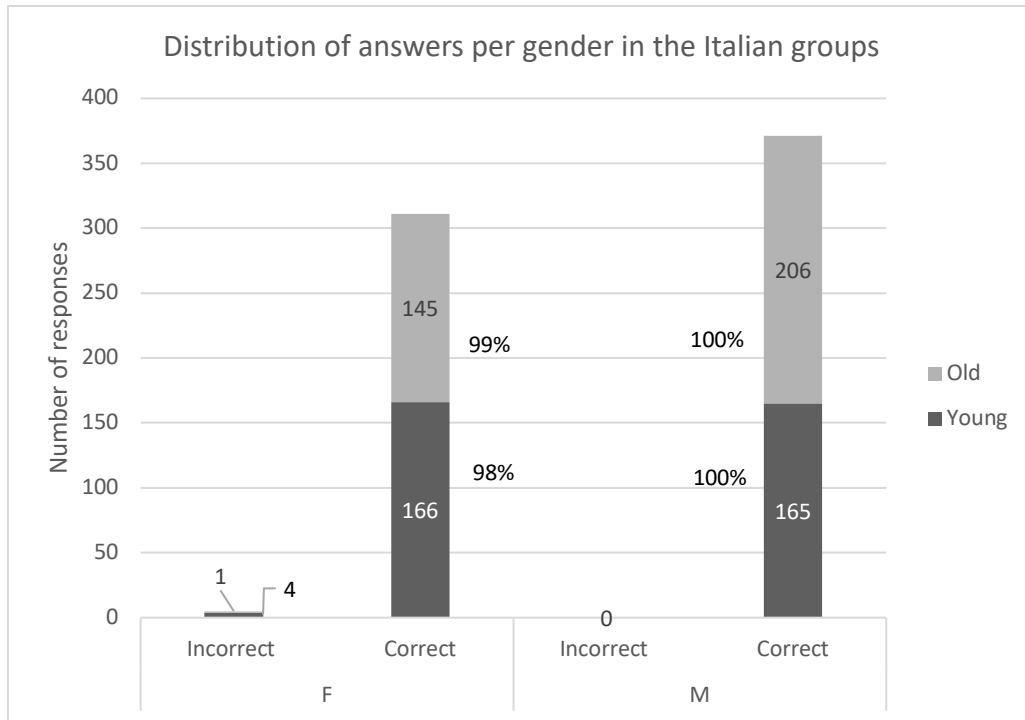
Predictor	Estimate	SE	t	p
Intercept	0.9567	0.00860	111.18	< .001
old – young	0.0380	0.01721	2.21	0.028
F – M	0.0172	0.01718	1.00	0.316

Table 14: Linear regression of the Croatian groups with no neuter gender

8 The analysis showed that the difference between the younger and older children is still
 9 significant, but to a lesser degree, when N is not accounted for. This means that N
 10 strongly contributes to the age difference of Croatian children, but it is not the only factor
 11 and Croatian children acquire gender at a slower pace than Italian children do (statistical
 12 difference between the two age groups not observable for the Italian group).

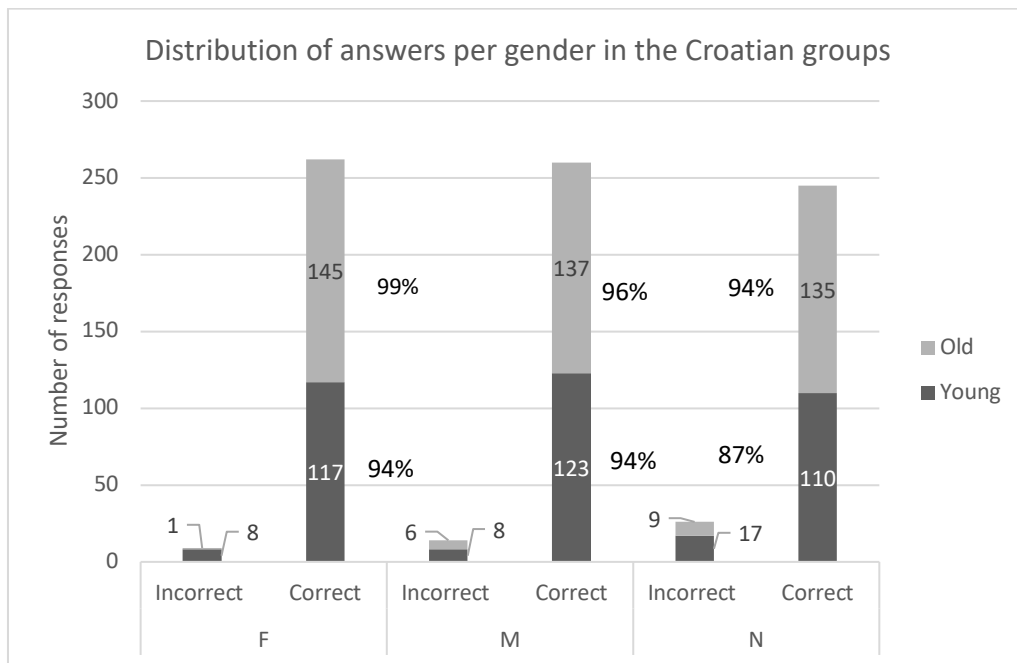
13 6.2 The acquisition of F

14 Due to a greater regularity of F throughout the paradigm, both in assignment and
 15 agreement in the two languages, we have hypothesized that it will be the first gender to
 16 be mastered. Figures 4 and 5 show the distribution of the correct/incorrect responses
 17 divided by gender in the two languages and display the percentage of correct answers
 18 per gender and per group.



1

2 Figure 4: Distribution of answers per gender in the Italian groups



3

4 Figure 5: Distribution of answers per gender in the Croatian groups

5 We can see that the Italian children are above the 95% correctness ratio for both genders
 6 even in the young group, but nevertheless the difference between F and M is evident as
 7 the latter is error-free across the task. This does not fit the previous findings for the article

1 system, but it fits the explanation from Pérez-Pereira (1991) that M is the unmarked
 2 gender in Spanish.

3 Our prediction holds for Croatian since F is the first gender to become error-free in
 4 our task. For the other two genders, the errors are still present in the older group, even if
 5 diminished. The possible reasons for this are discussed in the next section where we look
 6 at the error distribution in the two language groups. Also, from the percentiles presented
 7 in figure 5, we can see that the younger group is below the 95% acquisition threshold
 8 for all three genders. Based on this, we could argue that gender agreement in Croatian is
 9 not fully acquired in the younger group, i.e. by age 3;4 (age of oldest participant).

10 We have conducted ANOVAs on the distribution of answers for each gender to see
 11 how it changes with age. Tables 15-18 display the results of each ANOVA. It was not
 12 possible to conduct an ANOVA for M in Italian as there were only correct answers.

	Sum of Squares	df	Mean Square	F	p
Group	0.0510	1	0.0510	4.82	0.029
Residuals	3.9059	369	0.0106		

Table 15: ANOVA age group comparison of responses in the Italian feminine

13

	Sum of Squares	df	Mean Square	F	p
Group	0.220	1	0.2200	6.98	0.009
Residuals	8.481	269	0.0315		

14 Table 16: ANOVA age group comparison of responses in the Croatian feminine

15

	Sum of Squares	df	Mean Square	F	p
Group	0.0250	1	0.0250	0.512	0.475
Residuals	13.2597	272	0.0487		

16 Table 17: ANOVA age group comparison of responses in the Croatian masculine

17

	Sum of Squares	df	Mean Square	F	p
Group	0.344	1	0.3436	3.99	0.047
Residuals	23.162	269	0.0861		

Table 18: ANOVA age group comparison of responses in the Croatian neuter

1 The statistical results from above need to be discussed by considering the distributions
2 shown in figures 3 and 4. For Italian it is obvious that F is not the gender that is mastered
3 first as there are no errors with M, whereas for F a significant improvement can be noted.
4 When it comes to Croatian, the results have revealed a significant difference, i.e.
5 improvement of F, likely due to the reduction of the errors from 8 to 1. The difference is
6 not present for M as it reflects that M errors are still present in a similar proportion. When
7 it comes to N, the group difference is significant, which means that the children have
8 improved their agreement with N, but it still remains the gender with the most errors.

9 6.3 Error patterns

10 In this section, we will look a bit more closely in the errors that the children make, more
11 precisely, which gender is used instead of the target gender. The answer is straightforward
12 for the Italian children as no mistakes are made with M and since the language has only
13 two genders this means that all the errors made were M agreement on a F target.

14 The error pattern in Croatian might reveal different factors at play in the two age groups.
15 The responses are summarized in table 19, the shaded cells marking a target-like
16 response.

		Response gender		
Target gender	group	F	M	N
F	old	144	0	1
	young	117	3	5
M	old	2	137	4
	young	1	123	7
N	old	3	5	134
	young	3	14	109

17 Table 19: Distribution of gender responses in Croatian children

1 F has a similar distribution of errors among M and N. Within the errors of the other two
2 genders we can see that there is tendency to mistake N with M and vice versa, but F
3 realization for both genders is also present. This is likely due to the syncretism that the
4 two genders have across the case paradigm. With respect to the age groups, M and F
5 proceed together at an early stage, with the N lagging in accuracy. Whereas at a
6 subsequent stage, F is acquired while M and N are at the same level. The possible reasons
7 for this will be brought forth in the discussion.

8 7. Discussion

9 With this task we have strived to reveal whether Italian and Croatian children differ in
10 the time course of the acquisition of their gender system and whether individual gender
11 values are acquired at a different pace due to the differences present in the two systems.
12 In this section we will outline the results in connection to the literature presented
13 throughout the paper and our predictions in order to identify the possible implications
14 for the field of gender acquisition related to transparency.

15 We have predicted that the Italian children will have a higher accuracy rate due to
16 a higher degree of transparency which is provided by (i) less ambiguous declension
17 classes and (ii) presence of an obligatory gender-marked article which acts as a gender
18 cue in case the noun itself does not provide it. The results confirmed this prediction as
19 there were no significant differences between the two age groups in Italian, but there
20 were in Croatian. This entails that the Italian children have acquired the gender system
21 at age 2;6. This is in line with previous research regarding the acquisition of gender in
22 Italian (Kupisch et al., 2002; Pizzuto & Caselli, 1992). Thus, from a theoretical
23 perspective, the fact that the Italian gender system is morphologically transparent and
24 has a salient gender cue on the article, makes the Italian gender system very easy to
25 acquire. For Croatian, we cannot provide a confirmation of the acquisition of the full
26 gender system by age 3;6 due to the amount of correct responses within N still being
27 slightly below 95%. The obtained result indicates that the transparency level of the gender
28 system is related to its acquisition and it can thus be used to make predictions related to
29 the timing and ease of acquisition.

1 The results that we have obtained for Croatian are more central for the discussion,
2 as no previous study has investigated gender acquisition in Croatian from an agreement
3 perspective. We found that Croatian children also make few errors, which is expected as
4 the system is transparent. However, the errors that the children made reveal that the
5 Croatian gender system is acquired in at least two stages. These stages will be discussed
6 in relation to our third research question below.

7 We have also speculated that F will be the first gender to be mastered in both
8 languages, due to its stronger regularity when compared to the other genders. This has
9 indeed shown to be true for Croatian as the agreement patterns for F reach ceiling level
10 in the older group, whereas this does not happen for the M and N. A possible reason for
11 this finding might be (i) the syncretism of M and N in oblique cases and (ii) the
12 considerably lower frequency of N in the input for which children require more time and
13 exposure to grasp. This prediction does not hold for Italian, even though the literature on
14 the acquisition of articles would suggest otherwise. There are two possible reasons for
15 this: (i) the Italian adjectival agreement does not have irregularities with M as the article
16 system does; and (ii) the adjective is acquired at a later stage than the article which
17 according to Bottari et al. (1993) includes the insertion of the right morpheme.

18 Finally, we wanted to see if the differences in the gender systems resulted in
19 differences in the acquisition of individual genders. The Italian children are basically at
20 ceiling for their adjective production at the age we tested. On the other hand, we can
21 see that the Croatian children go through at least two stages which could be observed
22 with the included age groups. The first stage consists of similar error rates with F and M,
23 but significantly higher error rates with N. We can summarize the stage like so: $(F=M) < N$.
24 This is likely due to the lower frequency of N: 6% in child directed speech (Kovačević et
25 al., 2009). The second stage consists of F being at ceiling and the error rates with M and
26 N being similar; this also entails that the agreement accuracy for N has significantly
27 improved since it was much more error prone than M in the younger group. This stage
28 can be summarized like so: $F < (M=N)$. The improvement of N agreement is probably due
29 to a longer exposure to N nouns and their patterns. A plausible reason for M not
30 improving as much as F is the syncretism between M and N. As the child's exposure to
31 and usage of the case paradigm increase, the similarity between M and N becomes more

1 evident. If we were to test an older group, it is likely that these difficulties due to
2 syncretism would resolve.

3 What the results suggest is that the degree of transparency of the gender system
4 matters. We cannot look only at a manifestation of gender in isolation, but at the full
5 agreement paradigm to make more accurate predictions of how a gender system might
6 be acquired. However, the full Croatian paradigm was not investigated here as the
7 elicitation proceeded in NOM, yet we see that the effects of syncretism in oblique cases
8 were reflected in the agreement error rates. The Croatian gender system is acquired more
9 gradually when compared to Italian. Nevertheless, the errors made by Croatian children
10 are quite low in both age groups, which means that gender is grasped quite easily.

11 This study, among others, shows how transparency has to be placed on a
12 continuum and the full paradigms of the agreeing elements have to be taken into
13 consideration to assess how transparent a gender system is. Frequent and clear cues
14 contribute greatly to a fast mastery of the gender system (i.e. the Italian article), while low
15 frequency in the input (Croatian N) hinders this process.

16 8. Conclusions

17 This study has found differences in the time course of acquisition which can crucially be
18 attributed to the different degrees of transparency present in the two languages. Both
19 languages have transparent gender systems and are acquired easily. However, the
20 gender-marked article in Italian, the syncretism of M and N in Croatian as well as the
21 low frequency of N, make it so that the Italian system is more transparent and thus more
22 easily acquired.

23 Consequently, Italian children master gender in adjectives early, by age 2;6, and
24 simultaneously. Croatian children have overall more errors in adjectival agreement and
25 we can recognize two stages with distinct error patterns: in the first stage they are equally
26 accurate with F and M, and significantly less accurate with N; in the second stage F is
27 error-free, whereas the accuracy of the M and N is roughly the same. We have attributed
28 this to the low frequency of N, as well as the syncretism of M and N across the case
29 paradigm. This is not a confirmation of the mastery of the full gender system in Italian,

1 but the results nevertheless show how degree of transparency matters and how it reflects
2 on the acquisition of gender values.

3 This study thus contributes to the research on formal cues in gender acquisition by
4 considering transparency on a scale. It shows how even in two gender systems that are
5 considered transparent; gender mastery does not proceed at an equal pace. The data
6 presented here supports quite clearly that transparency is more than a binary feature
7 between transparent and opaque and it suggests how we must account for a transparency
8 scale and place the languages that we are researching on a continuum. As we have
9 shown here, in order to detect a language's place on this continuum, full paradigms of
10 the agreeing arguments should be taken into consideration.

11 This study also a valuable endeavor to the acquisition of Croatian since it is an
12 understudied language when it comes to language acquisition and it reveals how also
13 the full case paradigm can affect how each gender is acquired. The different times of
14 mastery of each gender value could represent a decisive starting point for future research
15 that could include a wider age range of the participants and, more importantly, testing
16 agreement patterns and accuracy on the full case paradigm.

17
18
19
20
21
22
23

References

- 24 Aljović, N. (2002). Long adjectival inflection and specificity in Serbo-Croatian.
25 *Recherches linguistiques de Vincennes*(31), 15.
- 26 Audring, J. (2014). Gender as a complex feature. *Language Sciences*, 43, 5-17.
- 27 Barić, E., Lončarić, M., Malić, D., Pavešić, S., Peti, M., Zečević, V., & Znika, M. (2005).
28 Hrvatska gramatika. In (4th ed.). Zagreb: Školska knjiga.
- 29 Belletti, A., & Guasti, M. T. (2015). *The Acquisition of Italian*: John Benjamins Publishing.
- 30 Bottari, P., Cipriani, P., & Chilosi, A. M. (1993). Protosyntactic devices in the acquisition
31 of Italian free morphology. *Language Acquisition*, 3(4), 327-369.
- 32 Caselli, M. C., Leonard, L. B., Volterra, V., & Campagnoli, M. G. (1993). Toward mastery
33 of Italian morphology: a cross-sectional study. *Journal of Child Language*, 20(2),
34 377-393. doi:10.1017/S0305000900008333
- 35 Chini, M. (1995). *Genere grammaticale e acquisizione: aspetti della morfologia nominale*
36 *in italiano L2* (Vol. 14): Franco Angeli.

- 1 Chini, M. (1998). Genuserwerb des Italienischen durch deutsche Lerner. *Wegener,*
2 *Heide: Eine zweite Sprache lernen: empirische Untersuchungen zum*
3 *Zweitspracherwerb. Tübingen, Narr, 39-60.*
- 4 Corbett, G. G. (1991). *Gender.* Cambridge: Cambridge University Press.
- 5 Costa, A., Kovačić, D., Franck, J., & Caramazza, A. (2003). On the autonomy of the
6 grammatical gender systems of the two languages of a bilingual. *Bilingualism:*
7 *Language and Cognition, 6(3), 181-200.* doi:10.1017/S1366728903001123
- 8 De Marco, A. (2005). *Acquisire secondo natura: lo sviluppo della morfologia in italiano*
9 (Vol. 47): FrancoAngeli.
- 10 Gudmundson, A. (2010). *L'acquisizione del genere grammaticale in italiano L2: Quali*
11 *fattori possono influenzare il grado di accuratezza.* Department of French, Italian
12 and Classical Languages,
- 13 Karmiloff-Smith, A. (1981). *A functional approach to child language: A study of*
14 *determiners and reference* (Vol. 24): Cambridge University Press.
- 15 Kovačević, M. (2004). *Croatian Kovacevic Corpus* [Corpora]. Retrieved from:
16 <https://childes.talkbank.org/access/Slavic/Croatian/Kovacevic.html>
- 17 Kovačević, M., Palmović, M., & Hržica, G. (2009). The acquisition of case, number and
18 gender in Croatian. *Development of nominal inflection in first language*
19 *acquisition: A cross-linguistic perspective, 153-177.*
- 20 Kupisch, T., Akpınar, D., & Stöhr, A. (2013). Gender assignment and gender agreement
21 in adult bilinguals and second language learners of French. *Linguistic Approaches*
22 *to Bilingualism, 3(2), 150-179.*
- 23 Kupisch, T., Geiß, M., Mitrofanova, N., & Westergaard, M. (2018). *Gender cues in L1*
24 *Russian children acquiring German as an early L2.* . Paper presented at the
25 EuroSLA.
- 26 Kupisch, T., Müller, N., & Cantone, K. F. (2002). Gender in monolingual and bilingual
27 first language acquisition: Comparing Italian and French. *Lingue e linguaggio, 1(1),*
28 107-150.
- 29 Levy, Y. (1983). It's frogs all the way down. *Cognition, 15(1-3), 75-93.*
- 30 Love, J., Dropmann, D., & Selker, R. (2018). Jamovi project (Version jamovi version 0.9).
31 Retrieved from <https://www.jamovi.org>
- 32 Matthews, P. (2007). *The concise Oxford dictionary of linguistics* (2nd ed.). Oxford:
33 Oxford Univeristy Press.
- 34 Mihajlović, M. (2014). A computational morphology approach to Croatian noun
35 inflection—the case of gender assignment. *Wiener Linguistische Gazzette*(Special
36 Issue 78A), 120-127.
- 37 Noccetti, S. (2002). *Pre-and Protomorphology in Language Acquisition: an Italian case*
38 *study:* Edizioni Plus srl.
- 39 Pérez-Pereira, M. (1991). The acquisition of gender: what Spanish children tell us. *Journal*
40 *of Child Language, 18(3), 571-590.* doi:10.1017/S0305000900011259
- 41 Pizzuto, E., & Caselli, M. C. (1992). The acquisition of Italian morphology: Implications
42 for models of language development. *Journal of Child Language, 19(3), 491-557.*
- 43 Rodina, Y. (2008). Semantics and morphology: The acquisition of grammatical gender in
44 Russian.
- 45 Rodina, Y., & Westergaard, M. (2015). Grammatical Gender in Norwegian: Language
46 Acquisition and Language Change. *Journal of Germanic Linguistics, 27(2), 145-*
47 *187.* doi:10.1017/S1470542714000245

- 1 Rodina, Y., & Westergaard, M. (2017). Grammatical gender in bilingual Norwegian–
2 Russian acquisition: The role of input and transparency. *Bilingualism: Language*
3 *and Cognition*, 20(1), 197-214. doi:10.1017/S1366728915000668
- 4 Smoczyńska, M. (1985). *The acquisition of Polish*: Lawrence Erlbaum Associates, Inc.
- 5 Trenkić, D. (2004). Definiteness in Serbian/Croatian/Bosnian and some implications for
6 the general structure of the nominal phrase. *Lingua*, 114(11), 1401-1427.
- 7 Voghera, M., Iacobini, C., Savy, R., Cutugno, F., De rosa, A., & Alfano, I. (2014). *VoLIP:*
8 *a searchable Italian spoken corpus, in Complex Visibles Out There*. Paper
9 presented at the Olomouc Linguistics Colloquium: Language Use and Linguistic
10 Structure., Olomouc: Palacký University.
- 11 Vuletić, D. (1991). *Istraživanje govora*: Fakultet za defektologiju Sveučilišta u Zagrebu.
12