Overt versus Zero Pronouns in Mandarin Chinese [*]

Ruyue Bi

University of California, Berkeley

Abstract

The distribution of Mandarin overt and zero pronouns in donkey sentences is compatible with what has been found in Japanese. Most cases can be accounted for by a distinction of binding methods: specifically, overt pronouns must be dynamically bound, and zero pronouns could be either dynamically bound or interpreted via the *E*-type strategy. However, in both languages, the classic "every farmer who owns a donkey beats it" sentence behaves unexpectedly. To resolve this inconsistency, an additional criterion, unique versus anaphoric definites, is introduced. This approach also sheds light on the syntactic representation of pronouns in Mandarin.

1   Introduction

It is well-established that many East Asian languages, such as Mandarin and Japanese, exhibit radical 'pro' drop – given the proper discourse context, both the subject and the object pronoun of a sentence can be freely omitted. Unlike ordinary "pro-drop" languages, for example Spanish and Italian, that usually show rich verbal agreement systems, radical 'pro' drop languages, interestingly, often do not have verbal inflection. Hence, very little direct evidence can be utilized to infer the status (person, number, gender, etc.) of the omitted pronoun. Huang 1984 discusses the subject-object asymmetry of the null pronoun, specifically in terms of which nominal expressions are permissible as the antecedent. Consider the following set of examples:

(1) a. Zhangsan$_i$ shuo [ta$_{i/j}$ bu renshi Lisi.]
         Zhangsan say   he   not know Lisi
         'Zhangsan said that he did not know Lisi'

     b. Zhangsan$_i$ shuo [$\varnothing_{i/j}$ bu renshi Lisi].
         Zhangsan say   not know Lisi
         'Zhangsan said that (he) did not know Lisi'   [Huang 1984:537]

     c. Zhangsan$_i$ shuo [Lisi bu renshi ta$_{i/j}$.]
         Zhangsan say   Lisi not know him
         'Zhangsan said that Lisi did not know him'   [Huang 1984:538]

     d. Zhangsan$_i$ shuo [Lisi bu renshi $\varnothing_{*i/j}$].
         Zhangsan say   Lisi not know
         'Zhangsan said that Lisi did not know (him)'   [Huang 1984:537]

In (1-a) and (1-c) where an overt pronoun occupies the subject and the object position, respectively, of the embedded clause, *ta* 'he/him' can refer to either the subject of the matrix clause *Zhangsan* or someone else in the discourse context. According to Huang, this is also the case for (1-b) in which the subject of the embedded clause is replaced by a null pronoun. However, the reading of *ta* referring to Zhangsan is not admissible in (1-d), when the null pronoun occurs in the object position. This inconsistency clearly suggests that a zero pronoun is not the simple deletion of its overt counterpart.

Moreover, I would like to make a new observation that in (1-b), although both readings are accessible, the subject null pronoun in the embedded clause more naturally refers to the matrix subject *Zhangsan* than to the contextually salient individual.

To attempt an explanation of the above-mentioned asymmetries, I closely examine the differences in distribution between Mandarin overt and zero pronouns in this paper. Section 2 provides basic background discussion of pronouns and their binding strategies. Section 3 replicates the examples in Kurafuji 1998 on Japanese overt versus zero pronouns, and demonstrates that the same basic conclusion holds in Mandarin: overt pronouns can only be dynamically bound, while both a dynamic binding and an *E*-type strategy are available to zero pronouns. However, this criterion alone cannot account for all example sentences. To resolve the issue raised by classic "every farmer who owns a donkey beats it" sentence, I incorporate the analysis of Mandarin definiteness in Jenks 2017 (which in turn builds from Schwarz 2009, 2013), and draw parallels both between overt pronouns and familiarity definiteness, and between zero pronouns and uniqueness definiteness. Section 4 concludes the paper and raises additional questions to be studied.

2 Pronominal Binding

2.1 Types of Pronouns

Assuming the existence of a linguistic antecedent, Nouwen 2014 categorizes pronouns into three distinct types: co-referential, bound and *E*-type.

Co-referential pronouns have type e antecedents, and can refer across sentence boundaries. In (2), the pronoun *she* refers to Mary, and *him* refers to John.

(2)   a.   Mary says she is friends with John.
         b.   Mary is friends with John. She invited him to the party.

Bound pronouns, not considering reflexives, have quantificational DPs (type $<<e,t>,t>$) as their antecedents, and their referents co-vary with their antecedents. In (3), it hardly makes sense to ask who specifically the pronoun *they* refers to.

(3)   Few girls say they are friends with John.

Proving the existence of the third category, *E*-type pronouns, is considerably more subtle. Evans 1980 is among the first to analyze them in detail.

(4)   [Evans 1980:339]

      a.   John owns some sheep and Harry vaccinates them in the Spring.
      b.   Some sheep are such that John owns them and Harry vaccinates them in the Spring.

In (4-a), the pronoun *them* is neither co-referential nor bound. The pronoun not being co-referential is obvious as its antecedent, the quantificational DP *some sheep*, is not of type *e*. Note that in (4-b), the second part of the conjunction in (4-a) is moved within the scope of the quantifier *some*. The pronoun cannot be bound by *some sheep*; otherwise, (4-a) would have been equivalent to (4-b), which is a false prediction since (4-a) entails Harry vaccinates all the sheep that John owns, while (4-b) does not.

Similarly, we can easily think of a scenario where (5-b) holds but (5-a) does not. Suppose five girls are friends with John, and one of these girls invited him to the party.

(5)   a.   Exactly one girl is friends with John. She invited him to the party.
        b.   Exactly one girl is such that she is friends with John and she invited him to the party.

The referent of an *E*-type pronoun is described with respect to the content of the antecedent sentence. In (5-a), *she* refers to, instead of an individual, the linguistic expression *the girl who is friends with John*, a definite DP of type *e*.

Now that we have established different behaviors of distinct types of pronouns, two binding strategies are introduced to license their occurrence.

2.2   Binding Strategies

Two major approaches to intersentential anaphora have been proposed in the literature – dynamic binding and an *E*-type strategy. According to Chierchia 1992, the former, building on the idea of "active discourse referents," hypothesizes that indefinites are existentially quantified with scope across sentence boundaries. The latter, as the name suggests, is closely related to *E*-type pronouns, where the pronoun is analyzed as the definite description reconstructed from the material of an indefinite in the context.

3   Mandarin Donkey Sentences

3.1   Pronouns and Their Corresponding Binding Strategies

Kurafuji 1998 presents evidence in Japanese illustrating the distinct distribution of overt and zero pronouns in donkey anaphora contexts. He argues that the differences in their distributions are caused by differences in strategies available for anaphora resolution: overt pronouns must be dynamically bound, and zero pronouns could be either dynamically bound or interpreted via the *E*-type strategy. The same set of sentences used by Kurafuji is replicated in Mandarin to test whether this conclusion can be extended.

The first set of sentences (6)-(8) exemplifies cases where both overt and zero pronouns can occur with equal level of acceptability:

(6) **Narrative Sequence**[1]
    Zhangsan you che$_i$. ta$_i$/∅$_i$ zai cheku li.
    Zhangsan have car   (it)   in garage inside.
    'Zhangsan has a car. It is in the garage.'

(7) **Conditional**
    Zhangsan ruguo mai shu$_i$, jiu   hui du-wan   ta$_i$/∅$_i$.
    Zhangsan if   buy book then will read-finish (it)
    'If Zhangsan buys a book, he finishes reading it.'

(8) **Conditional with Symmetric Interpretation**
    ruguo qiangdao$_i$ yudao   ling-yi-ge   qiangdao$_j$, ta$_i$/∅$_i$ hui deng yi-yan ta$_j$/∅$_j$.
    if   bandit   encounter another-one-CL bandit   (he)   will glare one-CL (him)
    'If a gangster passes by another gangster, he will glare at him.'

Both dynamic binding and an E-type strategy can be used to explain (6) and (7). In (6) the antecedent of the subject pronoun, either overt *ta* or zero, is the bare NP *che* 'car', the object of the preceding sentence; in (7) the antecedent of the object pronoun is the bare NP *shu* 'book', the object of the dependent clause. Note in (6), it has been reported that the pronoun referring to Zhangsan, the subject of the preceding sentence, is also a valid reading.

In (8) the subject pronoun of the main clause, either overt *ta* or zero, refers to the subject of the dependent clause, and the object pronoun refers to the object. As Kurafuji points out, previous literature, including Kratzer 1995, has argued that an E-type strategy is inadequate for interpreting symmetrical cases such as (8). Hence, dynamic binding needs be available for both overt and zero pronoun interpretation.

The second set of sentences (9)-(11) demonstrates cases where only zero pronouns can occur. They are also contexts where Chierchia 1992 claims an E-type binding strategy is needed:

(9) **Narrative Sequence**
    mei-ge   chanpin$_i$ dou bei   zixi   jiancha guo. ranhou ∅$_i$/*ta$_i$/?tamen bei
    every-CL product all PASS carefully inspect GUO then   (*it)/?they   PASS
    zhuang-jin xiangzi.
    pack-in   box
    'Every product was inspected carefully. Then they were packed in the box.'

(10) **Paycheck Sentence**[2]
    chule Zhangsan, mei-ge-ren   dou ba gongzika$_i$ gei-le   qizi. Zhangsan ∅$_i$/(ba
    except Zhangsan every-CL-person all BA paycheck give-PERF wife Zhangsan
    *ta$_i$) gei-le   qingfu.
         give-PERF mistress

---

[1]The gloss of Mandarin data mostly follows Huang, Y.-h. A. Li, and Y. Li 2009: BA = *ba* construction expressing "affectedness", CL = classifier, DE = pre-nominal modification marker *de*, DEM = demonstrative, GUO = experiential aspect marker *guo*, PASS = passive, PERF = perfective.

[2]This sentences involves the Mandarin *ba*-construction, where the marker *ba* is mandatory when followed by an overt nominal expression, but has to be omitted when there is a null pronoun.

'Except Zhangsan, everyone gave his paycheck to his wife. Zhangsan gave his to his mistress.'

(11) **Bathroom Sentence**

zhe-dong lou   yaome mei-you weishengjian$_i$, yaome $\varnothing_i$/*ta$_i$ jiu   zai qiguaide difang
this-CL   building either not-have bathroom   or   then in weird   place

'It is the case either that this building does not have a bathroom or that it is in a funny place.'

In (9), the intended antecedent of the subject pronoun, *chanping* 'product,' is universally quantified in the preceding sentence. Note that the passive marker *bei* in the second sentence is optional when following a zero pronoun, but mandatory if the subject is overt. Interestingly, the sentence becomes acceptable, although not as natural, if the subject is the pronoun *tamen* 'they,' 3rd-person plural, referring to "every product" as a collective.

In (10), the intended antecedent of the direct object pronoun, *gongzika* 'paycheck,' is in the scope of a non-c-commanding universal. Note that *ba* is obligatory as part of special *ba* construction denoting "affectedness" in Mandarin when the direct object is overt. We cannot obtain the intended meaning here if the pronoun were overt *ta* 'it'.

In (11), the intended antecedent of the subject pronoun, *weishengjian* 'bathroom,' is in the scope of negation. Kurafuji discusses an alternative account of bathroom sentences (Roberts 1989), which assumes that (11) is equivalent to (12):

(12)   zhe-dong lou   yaome mei-you weishengjian$_i$, ruguo you, $\varnothing_i$/ta$_i$ ye zai qiguaide
this-CL   building either not-have bathroom   if   have (it)   or in weird
difang
place

'It is the case either that there is not a bathroom or that if there is a bathroom, it is in a funny place.'

As it is the case in Japanese, both zero and overt pronouns are now grammatical in (12), which renders the assumption invalid.

Therefore, since zero pronouns in (9) to (11) are interpreted with an *E*-type strategy, we reach the same basic conclusion for Mandarin as Kurafuji did for Japanese: both dynamic binding and an *E*-type strategy should be available for the interpretation of zero pronouns, while only dynamic binding is available for overt pronouns.

However, as Kurafuji points out, there is a seemingly problematic case for this analysis. An overt pronoun is required in the Japanese counterpart of the classic donkey sentence "every farmer who owns a donkey beats it:"

(13)   [$_{\text{Rel}}$Ronbun$_i$-o yon-da]   dono gakusee-mo sore$_i$ hihanshi-ta.
     paper-ACC read-PAST which student-$\forall$ it-ACC criticize-PAST
'Every student that read a paper criticized it.'   [Kurafuji 1998:136]

A similar phenomenon can be observed in Mandarin. In (14), an overt pronoun *ta* has to be present for the sentence to be grammatical:

(14)   mei-ge   you lüzi$_i$ de nongfu dou hui da   ta$_i$/*∅$_i$.
       every-CL have donkey DE farmer all   will beat *(it)
       'Every farmer who owns a donkey beats it.'

Kurafuji resolves this issue by proposing a different syntactic structure for Japanese *-mo* '∀' from the one for English *every*, which leads to a QR movement of 'every student' out of relative clause. This proposal is based on the observation that when a quantificational adverb *taitee* 'mostly' is added in front of the pronoun *sore* 'it', (15-a) can have the reading in (15-b), i.e., "every student who reads papers criticized most of the papers read:"

(15)   a.   [$_{Rel}$ Ronbun$_i$-o yon-da]   dono gakusee-mo taitee sore$_i$-o/∅$_i$ hihanshi-ta.
                 paper-ACC read-PAST which student-∀   mostly it-ACC   criticize-PAST
            'Every student that read a paper criticized most of the papers he read.'

                                                                    [Kurafuji 1998:137]
       b. $\forall x[\text{student}'(x) \to \text{MOST}y[\text{paper}'(y) \wedge \text{read}'(y)(x)][\text{criticize}'(y)(x)]]$

                                                                    [Kurafuji 1998:138]

Note that English sentence (16) does not have the reading in (15-b):

(16)   Every student who read a paper mostly criticized it.   [Kurafuji 1998:139]

This analysis does not seem to have a straightforward application in Mandarin. The universal quantifier *mei* appears to be more aligned with English *every*. Therefore, to resolve the issue put forward by (14), an additional criterion proposed by Jenks 2017 is discussed in the following section.

3.2   Unique and Anaphoric Definites

Jenks 2017, building from the concepts of German 'weak' and 'strong' definite articles proposed in Schwarz 2009, demonstrates a distinction between the definite expressing uniqueness (unique definite) and the one expressing familiarity (anaphoric definite) in Mandarin. Specifically, the former is realized with a bare noun, and latter with a demonstrative, except in subject position. Consider the following donkey sentences:

(17)   a.   mei-ge   you lüzi$_i$ de nongfu dou hui da   [*(na-tou) lüzi$_i$].
            every-CL have donkey DE farmer all   will beat that.DEM-CL donkey
            'Every farmer who owns a donkey beats the donkey.'

       b.   mei-ge   you lüzi$_i$ de nongfu dou hui da   lüzi$_{*i}$.
            every-CL have donkey DE farmer all   will beat donkey
            'Every farmer who owns a donkey beats donkeys.'

In (17-a), the demonstrative DP *na-tou lüzi* refers to the specific donkey owned by the farmer, while we can only get the reading "all the donkey-owners beat donkeys in general" in (17-b). Notice that (17-a) is essentially equivalent to (14), repeated below:

(18) (=(14)) mei-ge   you lüzi$_i$ de nongfu dou hui da   **ta$_i$/*∅$_i$**.
          every-CL have donkey DE farmer all   will beat *(it)
          'Every farmer who owns a donkey beats it.'

There appears to be a correspondence between the overt pronoun *ta* and the demonstrative, and one between the zero pronoun and the bare noun. Two earlier examples are re-examined to further establish the connection, and to ensure the comparability of the Kurafuji 1998 and Jenks 2017 theories.

The first example is where both overt and zero pronouns can occur:

(19) a. (=(7)) Zhangsan ruguo mai shu$_i$, jiu   hui du-wan   **ta$_i$/∅$_i$**.
              Zhangsan if   buy book then will read-finish (it)
              'If Zhangsan buys a book, he finishes reading it.'

      b. Zhangsan ruguo mai shu$_i$, jiu   hui du-wan   [**na-ben   shu**].
          Zhangsan if   buy book then will read-finish this.DEM-CL book
          'If Zhangsan buys a book, he finishes reading the book.'

      c. Zhangsan ruguo mai shu$_i$, jiu   hui du-wan   **shu**.
          Zhangsan if   buy book then will read-finish book
          'If Zhangsan buys a book, he finishes reading the book.'

Both (19-b) with a demonstrative DP and (19-c) with a bare noun have the intended meaning as in (19-a).

The second example is where only a zero pronoun is permissible:

(20) a. (=(10))
          chule Zhangsan, mei-ge-ren   dou ba gongzika$_i$ gei-le   qizi. Zhangsan
          except Zhangsan every-CL-person all   BA paycheck give-PERF wife Zhangsan
          ∅$_i$/(ba ***ta$_i$**) gei-le   qingfu.
                        give-PERF mistress

          'Except Zhangsan, everyone gave his paycheck to his wife. Zhangsan gave his to his mistress.'

      b. *chule Zhangsan, mei-ge-ren   dou ba gongzika gei-le   qizi. Zhangsan ba
          except Zhangsan every-CL-person all   BA paycheck give-PERF wife Zhangsan BA
          [**na-zhang   gongzika**] gei-le   qingfu.
          that.DEM-CL paycheck   give-PERF mistress
          'Except Zhangsan, everyone gave his paycheck to his wife. Zhangsan gave his paycheck to his mistress.'

      c. chule Zhangsan, mei-ge-ren   dou ba gongzika gei-le   qizi. Zhangsan ba
          except Zhangsan every-CL-person all   BA paycheck give-PERF wife Zhangsan BA
          gongzika gei-le   qingfu.
          paycheck give-PERF mistress
          'Except Zhangsan, everyone gave his paycheck to his wife. Zhangsan gave his paycheck to his mistress.'

In (20-b), we cannot obtain the intended meaning – it is Zhangsan's paycheck that he is giving to his mistress. Instead, *na-zhang gongzika* 'that paycheck' refers to a contextually salient individual previously mentioned in the discourse. On the other hand, (20-c) is equivalent to (20-a).

The table below summarizes the three examples discussed:

(21)   Distribution in non-subject position

	ta	demonstrative	∅	bare noun
book sentence (7)	✓	✓	✓	✓
paycheck sentence (10)	✗	✗	✓	✓
donkey-owner sentence (14)	✓	✓	✗	✗

In conclusion, the overt pronoun *ta* appears to have the same distribution as a demonstrative, and the zero pronoun has the same distribution as a bare noun, when they are not in the subject position of a matrix clause.

It is considerably more complicated when pronouns occur in subject position, where, as Jenks 2017 argues, both anaphoric and unique definites can be realized with a bare noun. Then our current theory would suggest that, given the proper context, zero pronouns should always be acceptable in subject position. Further investigation is needed in this regard.

4   Conclusion: Initial Questions (Partially) Answered

Now let's return to the initial questions that motivated this paper, specifically why an object zero pronoun in the embedded clause referring to Zhangsan is not permissible in (22), and why the reading of a subject zero pronoun referring to Zhangsan, on the other hand, is somehow more natural in (23):

(22)   (=(1-d)) Zhangsan$_i$ shuo [Lisi bu renshi ∅$_{*i/j}$].
              Zhangsan say   Lisi not know
              'Zhangsan said that Lisi did not know (him)'   [Huang 1984:537]

(23)   (=(1-b)) Zhangsan$_i$ shuo [∅$_{i/j}$ bu renshi Lisi].
              Zhangsan say   not know Lisi
              'Zhangsan said that (he) did not know Lisi'   [Huang 1984:537]

First, consider the following set of examples addressing the issue raised in (22):

(24)   a.   Zhangsan shuo Lisi bu renshi [**ta** (**zhe-wei   xuesheng**)].
              Zhangsan say   Lisi not know him (this.DEM-CL student)
              'Zhangsan$_i$ said that Lisi did not know him$_{i/j}$.'

        b.   Zhangsan shuo Lisi bu renshi [**zhe-wei   xuesheng**].
              Zhangsan say   Lisi not know this.DEM-CL student
              'Zhangsan$_i$ said that Lisi did not know him$_{?i/j}$.'

        c.   Zhangsan shuo Lisi bu renshi [(**xuesheng**)].
              Zhangsan say   Lisi not know student

'Zhangsan$_i$ said that Lisi did not know him$_j$.'

The reading of the object nominal referring to a contextually salient individual, who is not Zhangsan, is available in all three cases. It referring to Zhangsan is completely licit in (24-a); the judgment varies in (24-b), largely depending on intonation; and the reading is unanimously rejected in (24-c). This observation seems to suggest that the overt pronoun in Mandarin has the representation shown on the left, where DemP, sister to Pro, is deleted, and that the zero pronoun has the representation on the right, where NP is deleted:
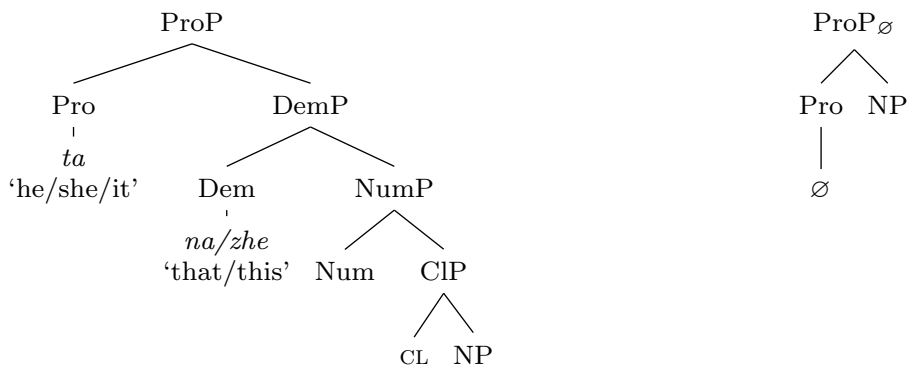
Secondly, in terms of (23) where a zero pronoun occurs in the subject position, *Zhangsan* is its preferred reading, since the demonstrative interpretation is more straightforward. This hypothesis corresponds directly with Jenks 2017's conclusion that an demonstrative interpretation of the bare noun is prioritized when possible.

In conclusion, both criteria, dynamic versus *E*-type binding and unique versus anaphoric definites, are needed to predicate the distribution of Mandarin pronouns in donkey anaphora contexts. In my subsequent studies, I hope to explore the possibility of unifying the two and to give a semantic account of how meanings are composed in this type of sentences.

Interestingly, there is a potentially problematic case, extended from a set of examples listed in Cheng and Huang 1996 and in Jenks 2017, that I have yet to figure out how to account for. This inconsistency may be explained by specific properties of *wh-* words in Mandarin.

(25)   a.   ni   jiao shei jin-kai, wo dou jian ta.
                you ask who enter, I   all   see him/her
                'Whoever you ask to come in, I'll see him/her.'   [Cheng and Huang 1996:130]

        b.   ni   jiao shei jin-kai, wo dou jian na-ge   ren.
                you ask who enter, I   all   see that.DEM-CL person
                'Whoever you ask to come in, I'll see that person.'

                                                  [Cheng and Huang 1996:130]

        c.   ni   jiao shei$_i$ jin-kai, wo dou jian $\varnothing_i$.
                you ask who enter, I   all   see him/her
                'Whoever you ask to come in, I'll see him/her.'   [Cheng and Huang 1996:130]

        d. *ni   jiao shei jin-kai, wo dou jian ren.
                you ask who enter, I   all   see person
                'Whoever you ask to come in, I'll see person.'   [Jenks 2017:12]

More questions remain to be explored. For example, why does plurality improve the acceptability of an overt pronoun in the narrative sequence where the intended antecedent is universally quantified?

References

Chierchia, Gennaro (1992). "Anaphora and dynamic binding". In: *Linguistics and Philosophy* 15.2, pp. 111–183.

Evans, Gareth (1980). "Pronouns". In: *Linguistic inquiry* 11.2, pp. 337–362.

Huang, Cheng-Teh James, Yen-hui Audrey Li, and Yafei Li (2009). *The syntax of Chinese*. Cambridge University Press Cambridge.

Jenks, Peter (2017). "Articulated definiteness without articles". In: *Linguistic Inquiry*. To appear.

Kratzer, Angelika (1995). "Individual-level predicates". In: *The generic book* 125.

Kurafuji, Takeo (1998). "Dynamic binding and the E-type strategy: Evidence from Japanese". In: *Semantics and Linguistic Theory*. Vol. 8, pp. 129–144.

Nouwen, Rick (2014). "E-type pronouns: congressmen, sheep and paychecks". In: *The Semantics Companion*.

Roberts, Craige (1989). "Modal subordination and pronominal anaphora in discourse". In: *Linguistics and philosophy* 12.6, pp. 683–721.

Schwarz, Florian (2009). *Two types of definites in natural language*. University of Massachusetts Amherst.

— (2013). "Two kinds of definites cross-linguistically". In: *Language and Linguistics Compass* 7.10, pp. 534–559.