

# ***HAVE*-verbs, indefinite article omission and definiteness in SLA**

[name]	[name]
[address]	[address]
[address]	[address]
[address]	[address]
[e-mail]	[e-mail]

## **Abstract**

We zoom in on how learners with an L1 without articles acquire an L2 with articles (see e.g. Hakuta 1976; Huebner 1983; Tarone & Parrish 1988; Thomas 1989; Ionin et al. 2004; Trenkic 2008; García-Mayo & Hawkins 2009; Ko et al. 2010). We use recent insights from theoretical semantics about the interaction between verbs and their objects to generate radically new predictions about L2 article acquisition and provide Learner Corpus evidence showing that these predictions are borne out.

## **1 Introduction**

One of the fundamental properties of language is that it allows us to refer, i.e. to establish a relation between a linguistic expression and something in the world. In a language like English, reference typically involves the use of a noun and a determiner (e.g. *a book, the book, this book*). Reference is however a multi-faceted area of investigation and involves much more than just nouns and determiners: linguistic and non-linguistic context, saliency, world knowledge, information structure, etc. Languages furthermore vary strongly in how they express referential properties, in particular in their inventory of definite and indefinite articles. The domain of referentiality is consequently a busy and interdisciplinary one bringing together linguistics, philosophy and psychology (see Hawkins 1978; Kamp 1981; Heim 1982; Givón 1983; Nunberg 1993; Gundel et al. 1993; Longobardi 1994; Haspelmath 1997; Bickel 2003; and Diessel 1999 for examples in linguistics; see Clark & Murphy 1982; Ariel 1990; Gundel et al. 1993; Gordon & Hendrick 1998; Arnold 1998, 2010 for examples in psychology; see Frege 1892; Russell 1905; Strawson 1950; Donnellan 1966; Burge 1973; Kripke 1972; Kaplan 1977; Evans 1982 for examples in philosophy).

In this short paper, we zoom in on how learners with an L1 without articles acquire an L2 with articles (see

e.g. Hakuta 1976; Huebner 1983; Tarone & Parrish 1988; Thomas 1989; Ionin et al. 2004; Trenkic 2008; García-Mayo & Hawkins 2009; Ko et al. 2010). We use recent insights from theoretical semantics about the interaction between verbs and their objects to generate radically new predictions about L2 article acquisition and provide Learner Corpus evidence showing that these predictions are borne out.

## **2 The acquisition of L2 articles by learners with an articleless L1**

In the literature on the L2 acquisition of articles by learners with an articleless L1 (like Korean, Russian, Mandarin, etc.), two empirical phenomena have been at the center of study:

- (1) My uncle gave me for my birthday **very strange present**.
- (2) In May in our club we had **the football match**.

Data taken from the Cambridge Learner Corpus

(1) illustrates article omission: the verb *give* requires its (singular) object to take a determiner that is omitted by the learner. The standard explanation is that (1) exhibits a case of negative transfer: the learner does not have articles in her L1 and omits them in her L2 (e.g. Huebner 1983). In this paper, we will only be concerned with indefinite and not with definite article omission. (2) illustrates definite article overproduction: in the context, no previous reference was made to a football match and the definite article is pragmatically odd. The standard explanation for definite article overproduction is that learners associate the definite article with other notions like specificity or objective identifiability rather than with definiteness (e.g. Ionin et al. 2004; Trenkic 2008).

## **3 Advances in the theoretical semantics of articles**

Recent theoretical work on the semantics of articles and verb types in languages like English and Spanish has led to two major findings. We introduce them here and show how they lead to new insights about L2

indefinite article omission and definite article overproduction in section 3.

The first finding we take from theoretical semantics is that the default entry of a number of verbs selects predicates rather than objects and comes with a built-in existential quantifier (see e.g. Borthen 2003; Dobrovie-Sorin et al. 2006; Espinal & McNally 2011; Cyrino & Espinal 2011; Alexandropoulou 2013; Le Bruyn, de Swart & Zwarts 2016).<sup>1</sup> The immediate upshot of this is that these verbs allow their objects to refer without an explicit spell-out of indefinite articles, even in an article language like Spanish:

- |     |    |                                |     |
|-----|----|--------------------------------|-----|
| (3) | a. | Juan llevaba <b>sombrero</b> . | SPA |
|     |    | Juan wore hat                  |     |
|     | b. | Tengo <b>casa</b> .            | SPA |
|     |    | I-have house                   |     |
|     | c. | Recibí <b>carta</b> .          | SPA |
|     |    | I-received letter              |     |

Article omission as in (3) occurs with verbs that have been qualified as *HAVE*-verbs (Borthen 2003) in Spanish, Catalan, Norwegian, Greek, Brazilian Portuguese and Rumanian. *HAVE*-verbs typically include verbs of possession/ownership (*have*), usage (*wear*), transfer of possession (*receive*, *give*, *buy*), creation (*build*) and consumption (*eat*). One of the major contributions of Le Bruyn, de Swart & Zwarts (2016) is to argue that *HAVE*-verbs have a special default entry cross-linguistically. They attribute the ungrammaticality of the English word-by-word translations of the paradigm in (3) to the syntactic generalization of determiners in regular argument position rather than to a semantic need.

The second major finding in the recent theoretical literature on articles and verb types is the unequivocal evidence for the existence of non-presuppositional definites in (5) (Coppock & Beaver 2015; Le Bruyn, de Swart & Zwarts 2016). Whereas a standard definite is typically taken to mark uniqueness and presuppositionality (4), the definite in (5) marks uniqueness but is non-presuppositional.

- (4) Liz didn't see **the only Armani dress**.  
 (5) Liz didn't wear **the only Armani dress**.

Negation is a standard way to test for presuppositions. (4) denies the existence of an event in which Liz saw the unique Armani dress but does not deny the existence of this dress. The existence of a unique Armani dress is consequently a presupposition. (5) is different in that its negation denies the existence of a unique Armani dress: the standard reading of (5) is that Liz's Armani dress was not the only one and there is thus no unique Armani dress. The definite

<sup>1</sup> We are aware of the fact that the way the existential quantifier is formalized varies from one proposal to the next. We follow the simplest proposal here, *viz.* the one presented in Van Geenhoven (1998).

article in (5) can consequently not be presuppositional in the same way as the one in (4).

(5) provides unequivocal evidence for the existence of non-presuppositional definite arguments in natural language. The fact that definites are typically presuppositional in English and only allow for non-presuppositional readings in contexts like (5) deserves further qualification, in particular as a better understanding of the constraints will give us a good handle on our L2 predictions. We assume with Coppock & Beaver (2015) and Le Bruyn, de Swart & Zwarts (2016) that presuppositional and non-presuppositional definites pragmatically compete with each other as soon as the two variants exist in a given language. The presuppositional variant being the stronger one will then typically end up being preferred. To bring out the non-presuppositional reading, several factors come into play, one of the major ones being verb type.<sup>2</sup> The minimal pair in (4) and (5) shows how a *HAVE*-verb like *wear* allows non-presuppositional definite objects whereas regular verbs like *see* do not. Le Bruyn, de Swart & Zwarts (2016) relate this to the fact that *HAVE*-verbs, on their default entry, come with a built-in existential quantifier and consequently come with a strong preference for non-presuppositional (indefinite) objects.

## 4 From theoretical semantics to SLA

The theoretical semantics literature summarized in Section 2 leads to two radically new predictions about the L2 acquisition of articles.

### 4.1 Indefinite article omission

Given that *HAVE*-verbs come with a built-in existential quantifier cross-linguistically, we predict L2 learners of English with an articleless L1 to treat the input in (6) and (7) in a different way:

- (6) Mary wore **a hat**.  
 (7) Mary saw **a hat**.

The prediction we make is that learners will initially consider the presence of the indefinite article optional in (6) and obligatory in (7). This is due to the fact that *wear* is a *HAVE*-verb and has a default entry with a built-in existential quantifier whereas *see* is a regular verb with no existential quantifier built-in. The empirical prediction we make is that (indefinite) article omission will be higher with *HAVE*-verbs than with regular verbs.

To our knowledge, this prediction is the first theory-driven attempt at organizing what has been traditionally considered a monolithic domain of

<sup>2</sup> The other two factors involved are the presence of *only* and the presence of a modifier. We refer the reader to Le Bruyn, de Swart & Zwarts (2016) for discussion.

transfer-driven L2 errors. Our prediction does not involve negative transfer at the DP level but rather the influence of a cognitive (lexical) universal at the VP level.

## 4.2 Definite article overproduction

The existence of both presuppositional and non-presuppositional definites in natural language leads to the prediction that learners will have to decide to which classes definites in English belong. We expect L2 learners to treat input involving *HAVE*-verbs differently from input involving regular verbs:

- (8) Mary wore **the hat**.
- (9) Mary saw **the hat**.

Even though we assume learners will be able to use the input they get to arrive at the conclusion that definites in English are typically unique and non-presuppositional, we expect them to arrive at this conclusion more quickly for definites occurring as objects of regular verbs than for those occurring as objects of *HAVE*-verbs. The rationale is that the default entry of *HAVE*-verbs requires non-presuppositional objects whereas there is no such requirement for regular verbs. The empirical prediction we make is that the overproduction of non-presuppositional definites will be higher after *HAVE*-verbs than after regular verbs.

The availability of presuppositional and non-presuppositional definites in natural language sheds a whole new light on earlier findings in the L2 literature on definite article overproduction. Even though factors like specificity and objective identifiability have been found to influence article choice in recent (offline) experimental paradigms (e.g. Ionin et al. 2004; Ko et al. 2010; Trenkic 2008), work looking into more naturally occurring production has failed to come to any stable theory-driven conclusions on definite article overproduction (e.g. Huebner 1983; Master 1987; Thomas 1989). We take this to be due to the fact that earlier non-experimental studies have only looked at the DP level and have failed to recognize the role of the VP. In Section 5, we will show that looking into naturally occurring production does give rise to clear conclusions as soon as one looks at the data through the right glasses.

## 5 A Learner Corpus study

Learner Corpora consist of spoken/written texts of L2 learners and do not focus on articles but are the result of the process of real language production in which all aspects of creating sentences and discourses are combined. The main limitations of doing semantic Learner Corpus Research are the same as the ones we find in semantic corpus research more generally, *viz.* no direct access to or control over the meaning that is conveyed, no direct access to the production process itself and sparsity of relevant data. Depending on the

predictions that the researcher wants to check, these limitations need however not be an impediment. In our particular case, our two predictions do not require us to have access to the meaning that is conveyed (unlike research on specificity, objective identifiability, etc.), do not require us to have access to the production process and are concerned with fairly frequent phenomena that can easily be studied on some of the bigger Learner Corpora available.

### 5.1 Data

We used the coded version of the Cambridge Learner Corpus (CLC). This is a 27 million word collection of texts written by learners taking an (English) Cambridge exam. The corpus is tagged for parts of speech, lemmas, errors, corrections, etc. It comes equipped with an online Sketch Engine search interface. We restricted our attention to the Chinese section, the biggest section of learners with an articleless L1s (approx. two thirds of the whole corpus).

We selected 10 high-frequency regular transitive verbs and 11 high-frequency *HAVE*-verbs. We used the list of *HAVE*-verbs in the literature to decide which verbs belong in which category. The full list of verbs is given in (10):

- (10) *Regular verbs*  
Ask, hear, help, know, leave, like, love, meet, show, visit
- HAVE-verbs*  
Buy, get, give, have, need, offer, provide, receive, take, use, wear

At our request, the data limits that were originally imposed to guarantee the stability of the search interface were removed. This allowed us to include the relevant data on the most frequent verb in the list (*have*) and explains why we have one more *HAVE*-verb than we have regular verbs.

We collected the following data per verb and per exam:

- (i) the number of missing determiners (error code *MD*) immediately following the verb that were corrected to *a*;
- (ii) the number of wrong determiners (error code *RD* 'replace determiner') immediately following the verb that were *the* and were corrected to *a*;
- (iii) the number of occurrences of *a* immediately following the verb.

Exams function as a proxy for learners, (i) as a proxy for indefinite article omission, (ii) as a proxy for definite article overproduction and (iii) as a proxy for (correct) indefinite article production.

## 5.2 Results

The results we present on indefinite article omission and definite article overproduction were calculated separately and are restricted to those exams that contain the relevant errors.

Table 1 reports the indefinite article omission ratio and the definite article overproduction ratio. The former is calculated by comparing indefinite article omission to the number of correct indefinite articles plus the number of omitted indefinite articles. The latter is calculated by comparing definite article overproduction to the number of correct indefinite articles plus the number of overproduced definite articles.

	Existential verbs	Regular verbs
Indefinite article omission ratio	1722 out of 6089 (28%)	226 out of 2952 (7,5%)
Definite article overproduction ratio	1402 out of 2244 (62%)	179 out of 608 (29%)

Table 1

In order to assess whether the difference in indefinite article omission between verb types is significant, we went beyond the coarse-grained ratios in Table 1 and used a generalized mixed effects model, using the ordinal software package (Christensen, 2015) in R (R Core team, 2016). We modeled random intercepts for learner and verb, and verb type and frequency as fixed effects. The results from this model confirm the significance of the influence of verb type ( $\beta = -4.05$ ,  $p < .01$ ) on article omission, while controlling for frequency. The negative coefficient for regular verbs indicates that, compared to *HAVE*-verbs, a higher indefinite article omission rate is less likely. We ran the same type of model to assess whether the difference in definite article overproduction is significant. The results from this model confirm the significance of the influence of verb type ( $\beta = -1,999$ ,  $p < .024$ ) on definite article overproduction, while controlling for frequency. The negative coefficient for regular verbs indicates that, compared to *HAVE*-verbs, a higher definite article overproduction rate is less likely.

## 5.3 Discussion

On the basis of the recent findings in the semantics literature we reported in section 2, we made two new predictions for the second language acquisition of articles in section 3. The first was that indefinite article omission is more frequent with *HAVE*-verbs than with regular verbs. The results we reported in 5.2. clearly show that this prediction is borne out. We present a few illustrative examples of indefinite article omission with *HAVE*-verbs in (11) to (15):

- (11) I decided not to buy **second hand computer** [...]
- (12) The school made no efforts to offer **social programme** [...]
- (13) I am looking forward to receiving **free book** [...]
- (14) We are happy to use **remote control** [...]
- (15) Please don't wear **red dress** [...]

The second prediction we looked into was that definite article overproduction is more frequent with *HAVE*-verbs than with regular verbs. The results in 5.2. show that this prediction is also borne out. Some illustrative examples of definite article overproduction with *HAVE*-verbs are given in (16) to (20):

- (16) People only get **the car** when they really need [...]
- (17) We were given **the free evening** straight afterwards [...]
- (18) Uncle Joe has had **the third baby** [...]
- (19) We understand that our town needs **the new face** [...]
- (20) They can offer **the wide range of drinks** [...]

Even though some further refinements of the data collection would be possible, the results are stable enough to compensate for the noise that is still in the data. This noise concerns – among other things – the complication of ditransitive verbs (e.g. *ask*, *give*) that can be followed by their indirect objects and not only by their direct objects, the fact that not all verbs are immediately followed by their objects (see e.g. example (1)) and the danger of having frequent collocates in the data. We controlled to some extent for the noise caused by frequent collocates by selecting verbs that showed a sufficiently broad array of object nouns in their Word Sketch, a useful tool within the Sketch Engine interface. A further study should also look into possible interactions between indefinite article omission and definite article overproduction on the one hand and attempt a cross-sectional interpretation of the data to try and get more insight in developmental patterns on the other hand. Optimally, the data reported here should also be replicated with an experimental paradigm.

## 6 Conclusion

Inspired by recent advances in theoretical semantics on the interaction between reference and verb type, we proposed two new ways of looking at the L2 article acquisition of learners with an articleless L1. The first was to look at indefinite article omission as correlating with verb type: the default entry of *HAVE*-verbs cross-linguistically comes with a built-in existential quantifier and we showed how this translates into a higher indefinite article omission ratio in the L2 article production of learners with an articleless L1. This is a new perspective on article

omission that complements in a fine-grained way the more traditional negative transfer hypothesis.

Our second contribution was to follow-up on the idea that natural language definites can be both presuppositional and non-presuppositional. By comparing definite article overproduction in the object position of *HAVE*-verbs to the object position of regular verbs, we were able to argue that L2 learners with an articleless L1 assume definites can be non-presuppositional. This is a radically new perspective on definite article overproduction that raises interesting new questions about second language pragmatics and the relation between non-presuppositional definiteness, specificity and identifiability.

We hope to have shown how research on referentiality should go beyond the DP and look at interactions with the VP and how general linguistics approaches allow us to generate exciting new avenues for research in SLA that cannot only be traced in experimental paradigms but even show up in more naturally occurring production.

## References

- Alexandropoulou, S. 2013. *The lexical restrictions on verbs with bare nominal complements in Greek and an analysis of such constructions with have-verbs* (Master's thesis, Utrecht University).
- Ariel, M. 1990. *Accessing NP antecedents*. London: Routledge.
- Arnold, J.E. 1998. Reference form and discourse patterns. PhD dissertation, Stanford University.
- Arnold, J.E. 2010. How speakers refer. The role of accessibility. *Language and Linguistics Compass* 4(4):187–203.
- Bickel, B. 2003. Referential density in discourse and syntactic typology. *Language* 79(4):708–736.
- Borthen, K. 2003. *Norwegian bare singulars* (Doctoral dissertation, Norwegian University of Science and Technology).
- Burge, T. 1973. Reference and proper names. *The Journal of Philosophy* 70(14):425–439.
- Christensen, R. H. B. (2015). Ordinal - Regression Models for Ordinal Data. R package version 2015.6-28.
- Clark, H. H. & Murphy, G.L. 1982. Audience design in meaning and reference. *Language and Cognition* 9:287–299.
- Coppock, E., & Beaver, D. (2015). Definiteness and determinacy. *Linguistics and Philosophy*, 377-435.
- Diessel, H. 1999. *Demonstratives. Form, Function and Grammaticalization*. Amsterdam: John Benjamins.
- Dobrovie-Sorin, C., Bleam, T., & Espinal, M. T. 2006. Bare nouns, number and types of incorporation. *Non-definiteness and plurality*, 51-79.
- Donnellan, K. 1966. Reference and definite descriptions. *The Philosophical Review* 75(3):281–304.
- Espinal, M. T., & McNally, L. 2011. Bare nominals and incorporating verbs in Spanish and Catalan. *Journal of Linguistics*, 47, 87-128.
- Evans, G. 1982. *The Varieties of Reference*, Oxford: Clarendon Press.
- Frege, G. 1892. Über Sinn und Bedeutung. *Zeitschrift für Philosophie und Philosophische Kritik* 100:25–50.
- García Mayo, M. del Pilar & Hawkins, R. 2009. *Second language acquisition of articles: Empirical findings and theoretical implications*. Amsterdam: John Benjamins.
- Givón, T. 1983. *Topic Continuity in Discourse: A Quantitative Cross-Language Study*. Amsterdam: John Benjamins.
- Gordon, P. C. & Hendrick, R. 1998. The representation and processing of coreference in discourse. *Cognitive Science* 22(4):389–424.
- Gundel, J., Hedberg, N., & Zacharski, R. 1993. Cognitive status and the form of referring expressions in discourse. *Language* 69(2):274–307.
- Hakuta, K. (1976). A case study of a Japanese child learning English as a second language. *Language learning*, 26(2), 321-351.
- Haspelmath, M. 1997. *Indefinite pronouns*. Oxford: Oxford University Press.
- Hawkins, J. 1978. *Definiteness and Indefiniteness. A Study in Reference and Grammaticality Prediction*. London: Croom Helm.
- Heim, I. 1982. The semantics of definite and indefinite noun phrases. PhD dissertation, University of Massachusetts, Amherst.
- Huebner, T. (1983). *A longitudinal analysis of the acquisition of English*. Ann Arbor: Karoma Publishers.
- Ionin, T., Ko, H., & Wexler, K. (2004). Article semantics in L2 acquisition: The role of specificity. *Language Acquisition*, 12(1), 3-69.
- Kamp, H. 1981. A theory of truth and semantic representation. In *Formal Methods in the Study of Language*, J. Groenendijk, T. Janssen & M. Stokhof (eds), 277–322. Amsterdam: Mathematical Centre.
- Kaplan, D. 1977. Demonstratives. An essay on the semantics, logic, metaphysics, and epistemology of demonstratives and other indexicals. In *Themes from Kaplan*, J. Almog, J. Perry & H. Wettstein (eds), 481–564. Oxford: Oxford University Press.

- Ko, H., Ionin, T., & Wexler, K. (2010). The role of presuppositionality in the second language acquisition of English articles. *Linguistic inquiry*, 41(2), 213-254.
- Kripke, S. 1972. *Naming and Necessity*. Cambridge: Harvard University Press.
- Le Bruyn, B., De Swart, H., & Zwarts, J. 2016. From *HAVE* to *HAVE*-verbs: Relations and incorporation. *Lingua*, 182, 49-68.
- Longobardi, G. 1994. Reference and proper names. A theory of n-movement in syntax and logical form. *Linguistic Inquiry* 25(4):609–665.
- Nunberg, G. 1993. Indexicality and deixis. *Linguistics and Philosophy* 16(1):1–43.
- R Core Team (2016). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.
- Russell, B. 1905. On denoting. *Mind* 14(56):479–493.
- Strawson, P.F. 1950. On referring. *Mind* 59(235):320–344.
- Tarone, E. & Parrish, B. (1988). Task-related variation in interlanguage: The case of articles. *Language learning* 38, 21-44.
- Thomas, M. (1989). The acquisition of English articles by first-and second-language learners. *Applied psycholinguistics* 10, 335-355.
- Trenkic, D. (2008). The representation of English articles in second language grammars: Determiners or adjectives? *Bilingualism: Language and Cognition* 17, 237-257.
- Van Geenhoven, V. (1998). *Semantic incorporation and indefinite descriptions: Semantic and syntactic aspects of noun incorporation in West Greenlandic*. CSLI publications.