Constraining Morphological Variation and Development: Agreement in L2 Spanish

Corrine McCarthy
George Mason University

1. Introduction

Structure building in language acquisition involves the adding of successive nodes in a hierarchical structure; its application is seen in the realm of syntax for both L1 (e.g. Clahsen, Penke & Parodi, 1993; Vainikka, 1993) and L2 acquisition (e.g. Vainikka & Young-Scholten, 1994, 1996), and to some extent within phonology as well (Brown, 1997). Although researchers in morphological theory argue that hierarchical structure should be built into the organization of morphological features (e.g. Harley & Ritter 2002; Cowper 2005), the application of structure building in the domain of morphology has received minimal attention within L1/L2 acquisition. The purpose of this paper is to explore to what extent feature geometric models of morphology may be applicable to the domain of L2 agreement: do L2 learners acquire structure gradually in the domain of morphology?

Within a generative framework, morphological development has been tied to syntactic development (Vainikka & Young-Scholten 1994, 1996). The successful usage of person and number agreement was taken as a diagnostic for the acquisition of phrases such as AgrP (Agreement Phrase), and the successful usage of past tense morphology as a diagnostic for the acquisition of the TP (Tense Phrase). Compare (1), which the authors analyze as an unraised infinitive in a bare-VP tree, with (2), which shows both agreement and raising to a higher syntactic projection.

(1) Eine Katze Fisch alle essen
A cat fish entire eat.INF
'A cat ate the entire fish.' (Vainikka & Young-Scholten, 1994:280)

(2) Trinkst du Cola?
drink.2SG you cola?
'Are you drinking cola?' (Vainikka & Young-Scholten, 1994:286)

A problem with this approach, however, is that speakers may omit morphology, while manifesting other properties associated with AgrP or TP. An example of nominative case (in the form of the pronoun *they,* which is commonly associated with TP), in the absence of tense morphology, is shown in (3), which comes from Lardiere's case study of Patty (1998). Patty also shows robust evidence for a CP projection (4); the presence of a CP should entail the presence of all the projections below it, yet in the same session she fails to consistently produce TP-associated tense morphology.

(3) Yesterday they open until five. (Lardiere 1998:16)

(4) I think [$_{CP}$ that we are so lucky] (Lardiere 1998: 19)

A solution offered by Prévost and White (2000), the Missing Surface Inflection Hypothesis, is that the absence of appropriate tense/agreement morphology is one of lexical access (see also Haznedar & Schwartz, 1997). Learners have the syntactic projections associated with tense/agreement, but fail to

* I would like to acknowledge the SPLLOC project as the source of data for this study. Thank you to the project for making this resource available to researchers. I would also like to thank the audience members at GASLA 2011 for their comments on the presented version of this paper. Although I have not attempted to answer all their questions in this paper, some of them are addressed in a forthcoming paper (McCarthy, to appear). Any errors and inconsistencies remain my own.

 Proceedings of the 11th Generative Approaches to Second Language Acquisition Conference (GASLA 2011), ed. Julia Herschensohn and Darren Tanner, 76-82. Somerville, MA: Cascadilla Proceedings Project.

access inflectional morphology due to "processing reasons or to communication pressure" (129). This proposal allows for morphology and syntax to develop along separate paths, which is a starting point for the current investigation. However, their proposal does not offer any specific predictions about what path the acquisition of morphology might take. The goal of the present paper is to take steps toward constructing a theory of morphological development in which morphology develops independently of syntax. The only assumption regarding syntax is that there must be "enough syntax" in order for syntactic operations such as movement to take place, and for the morphology to have a host, should it happen to appear (though there is no guarantee that it will).

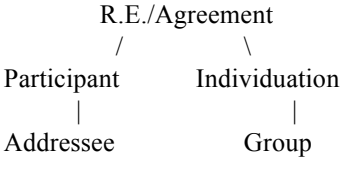
In addition to looking into the path of development, I will briefly address the issue of morphological variation. In previous work I argued that morphological variation is systematic rather than random, in that L2ers favored one systematic default form, which was argued to be the underspecified, representationally-simpler form (McCarthy 2007, 2008). In this paper, I explore the idea that these representationally-simpler forms are not only the defaults, but also the forms that successfully emerge first.

2. Representing person and number

Using typological data from a wide variety of languages, Harley and Ritter (2002) provide a feature geometric model of person and number morphology. For a justification of why the use of geometries is advantageous within morphological theory, and why the tree looks like it does, see Harley and Ritter (2002).

In the geometry, features are represented via underspecification, and using privative values rather than binary ones, e.g. [+plural], [-plural]. These authors assume that the unmarked value is underspecified, and that there is a positive correlation between markedness and the amount of structure required. There are two sub-trees: one for person (the left half of the partial tree shown below) and one for number (the right half). For person, this structure encodes markedness relations: 1st person is marked relative to 3rd, and bears an additional feature of [Participant]. 1st and 2nd person are further distinguished by the feature [Addressee], making 2nd the most highly specified and most marked person. The 1st/2nd contrast will not be examined here, due to a lack of 2nd person contexts in the dataset. For number, plural is marked relative to singular, and bears an additional feature of [Group]. (5) is a partial geometry including the features that are relevant to the present analysis.

(5) Feature geometry (partial), based on Hanson et al. (2000); R.E. = Referring Expression (pronoun)



I assume that finite is more highly specified relative to nonfinite, which lacks the feature Finite (following Cowper, 2005). Spanish, unlike some dialects of Portuguese, does not have "inflected infinitives," i.e., infinitives that show overt person and number agreement. Features related to tense and finiteness are not assumed to be a part of the person/number tree; nevertheless, the successful representation of finiteness can be seen as a precursor to representing the features in (5).

3. Spanish agreement and L2 acquisition

The Spanish L2er must acquire the features that correspond to person and number agreement. Spanish is a language that is richly inflected, as shown in (6-8). Peninsular Spanish and Latin American Spanish differ in their 2nd person plural agreement; since the participants are from the UK, the target is assumed to be Peninsular Spanish. In the simple present (6) and past preterite (7), each person-number paradigm cell is filled with a unique form. In the past imperfect tense (8), 1st and 3rd singular are syncretic, but no other syncretisms are found. In some analyses (e.g. Montrul, 2004) the

3rd person suffix is treated as part of the stem, whereas in others it is treated as a unique affix (e.g. Oltra-Massuet & Arregi, 2005); for present purposes, I treat 3rd person morphology as a unique suffix.

(6) Spanish simple present indicative of -ar/-er/-ir verbs

	1st person	2nd person	3rd person
singular	o/o/o	as/es/es	a/e/e
plural	amos/emos/imos	áis/éis/ís	an/en/en

(7) Spanish past preterite of -ar/-er/-ir verbs

	1st person	2nd person	3rd person
singular	é/í/í	aste/iste/iste	ó/ió/ió
plural	amos/imos/imos	asteis/isteis/isteis	aron/ieron/ieron

(8) Spanish past imperfect of -ar/-er/-ir verbs

	1st person	2nd person	3rd person
singular	aba/ía/ía	abas/ías/ías	aba/ía/ía
plural	ábamos/íamos/íamos	abais/íais/íais	aban/ían/ían

Previous studies of adults acquiring L2 Spanish have found that speakers are very accurate overall on person and number agreement, but sometimes adopt (what appears to be) a 3rd singular default suffix in non-3rd person contexts (Bruhn de Garavito 2003a, 2003b; McCarthy 2007). In an elicited production task, Bruhn de Garavito (2003a, 2003b) found that beginning learners made substitution errors in person and finiteness at a rate of 10.1 percent. The most frequent error, accounting for 66.3 percent of errors, was the substitution of 3rd person in place of another person. It should be noted that Bruhn de Garavito analyzes these tokens as infinitives missing an –r suffix, in line with a claim of the Missing Surface Inflection Hypothesis (Prévost and White 2000): if inflection is present, it is accurate. The next most frequent was the substitution of the infinitive for finite (13.5 percent), followed by 1st person for another person (12.4 percent). McCarthy (2007) reports similar patterns, with 3rd-for-1st and singular-for-plural accounting for the majority of errors; very few errors in finiteness were attested, perhaps due to their relatively high proficiency level. In a small-scale study of initial-state L2 Spanish learners, Mezzano (2003) reports errors in finiteness, such that bidirectional errors in finiteness (finite-for-nonfinite and nonfinite-for-finite) are widely attested in the first session. By the second session, finite-for-nonfinite substitutions had decreased, with the majority of errors in finiteness involving nonfinite verbs used in finite contexts, consistent with the Missing Surface Inflection Hypothesis, which predicts nonfinite defaults. For person, Mezzano reports a weak tendency to favor 3rd person in non-3rd contexts: 57 percent of person errors in the second session involved 3rd for non-3rd substitutions.

If L2 acquisition of morphology consists of the gradual adding of nodes in a feature geometry, the following prediction can be made: the consistent use of the marked/more highly-specified feature will come developmentally later than the consistent use of the unmarked/underspecified feature. In order to test this prediction, I look at two proficiency groups, and compare how each proficiency group realizes the opposing features.

4. Method

The data come from a corpus of L2 Spanish collected as part of phase two of the Spanish Learner Language Oral Corpora (SPLLOC; 2010). The SPLLOC project consists of various experimental tasks; the present analysis comes from the guided interview tasks. The first part of this task involved participants being shown pictures of famous people (including John Lennon and Princess Diana), and talking about what they knew about them. The second part was the "autobiography" in which participants were asked to describe past experiences, including their first memory and what they normally did at various times of their life. The task was designed to elicit past tense, but past tense was not necessarily produced by the speakers.

The participants were native English speakers living in the UK who learned Spanish in instructed

settings. Two proficiency levels were examined: low-intermediate and intermediate; higher-level speakers were excluded because the goal was to examine error production. Proficiency was assessed based on their level in the English (UK) educational system. The low-intermediate group was aged 14-15 and had approximately 240 hours of Spanish; the intermediate group was aged 17-18 and had approximately 750 hours of instruction. A total of 37 speakers are examined here: 20 from the low-intermediate group and 17 from the intermediate group. All of the transcripts that were available between June and July 2010 were analyzed.

All verbs were coded for accuracy in person, number, and finiteness. Additionally, any error that occurred was also coded for its type, e.g., 3rd for 1st, singular for plural, nonfinite for finite. Since Spanish is a null subject language, any ambiguous utterances lacking a clear indication of the subject were not counted. In addition, any formulaic utterances such as *no sé* ('I don't know'), *no lo sé* ('I don't know it/that'), and *me gusta* ('I like it') were excluded. Following Lardiere (1998), utterances that were followed by self-correction were excluded; the final, corrected forms were included. Because the majority of copular verbs occurred in 3rd person contexts, they were excluded from the analysis. Only lexical main verbs are therefore included. Verbs that took on the morphology of another verb class (such as -ar instead of –er) were retained. The files were coded twice by the same person (a speaker of L2 Spanish), with a gap of approximately two months in between each coding pass; differences that arose were then settled by that same coder. 960 lexical main verbs are reported in the present analysis.

5. Results

Accuracy rates on each feature are shown for person in Table 1, number in Table 2, and finiteness in Table 3. Speakers who produced fewer than three tokens of a particular context were not included in the calculation of accuracy rates.

For person, there is an increase in accuracy on the marked feature (1st person) but not the unmarked one (3rd) as proficiency increases. The difference between levels is significant for 1st person ($t=2.378$, $df=18.222$, $p=.029$, equal variances not assumed) but not 3rd person ($t(28)=-.460$, $p=.649$). 3rd person accuracy is consistently fairly high across proficiency levels, whereas 1st person shows an increase from low-intermediate to intermediate. This pattern is consistent with the claim that the marked or most highly specified feature is acquired after the unmarked or underspecified feature.

Table 1. Accuracy rates on each person feature by level, with standard deviation in parentheses, followed by number of speakers counted

	3	1
Low-intermediate	91.9 (13.6); n=13	69.8 (28.2); n=15
Intermediate	93.9 (10.0); n=17	88.4 (11.7); n=17
Totals	93.0 (11.5); n=30	79.7 (22.8); n=32

For number, there is no proficiency effect for either singular or plural (singular: $t(29)=.281$, $p=.541$; plural: $t(18)=.120$, $p=.878$); the accuracy rate for both features remains fairly stable regardless of proficiency level. Very few errors in singular contexts are attested, and speakers perform at ceiling throughout the data set. Errors in plural are more frequent overall, with learners substituting singular agreement for plural. Contrary to predictions, proficiency level appears not to have a very strong effect on the development of number features.

Table 2. Accuracy rates on each number feature value by level, with standard deviation in parentheses, followed by number of speakers counted

	Singular	Plural
Low-intermediate	99.5 (1.9); n=15	91.4 (13.6); n=9
Intermediate	99.2 (3.2); n=17	90.7 (13.7); n=11
Totals	99.3 (2.7); n=31	91.0 (13.3); n=20

Both the low-intermediate and intermediate groups performed at ceiling on finite. The effect of group on finite was significant (t(20.857)=-3.544, p=.002, equality of variances not assumed), indicating that accuracy in finiteness increases along with proficiency level. This pattern is consistent with the predictions, because accuracy in nonfinite morphology emerges developmentally prior to accuracy in finite morphology.

Table 3. Accuracy rates on finiteness by level, with standard deviation in parentheses, followed by number of speakers counted

	Nonfinite	Finite
Low-intermediate	100 (0); n=5	75.4 (28.0); n=20
Intermediate	100 (0); n=12	98.2 (5.7); n=17
Totals	100 (0); n=17	85.9 (23.7); n=27

Looking briefly at the question of variation, are systematic defaults employed as the outcome of substitution errors (as suggested in McCarthy 2007, 2008)? For finiteness and number, only two feature values exist: finite vs. nonfinite; singular vs. plural. This means that all errors in finiteness and number were instances in which the target was replaced by the other feature value. For these features, the relationship between defaults and accuracy is clear: the default form (nonfinite and singular) is the one that is overall most accurate. For these features, when variation occurs, it is systematic and the output of substitution errors is the representationally-simpler default form.

The situation with person is more complicated because three options exist: 1st, 2nd, and 3rd. 3rd person is more accurate than 1st person, but are those errors in 1st person instances in which 3rd person is produced as the default? In this case, the answer is yes: out of the 32 speakers who produced sufficient contexts of 1st person, only three produced something other than 3rd person (namely, 2nd person). 3rd person is more accurate, and it is the favored default in 1st person contexts.

6. Discussion

The goal of this paper was to assess whether a feature-geometric model might provide a good starting point for predicting patterns in development of L2 morphology. Specifically, it was hypothesized that 3rd person, singular, and nonfinite would emerge prior to their more marked, highly-specified counterparts: 1st, plural, and finite. For person and finiteness, this hypothesis appears to be supported; the number results do not support this hypothesis.

Accuracy in 3rd person remains stable regardless of level, whereas accuracy in 1st person increases from lower proficiency to higher. Where errors in 1st person occur, by and large, they are substitutions of 3rd person morphology in a 1st person context. This is consistent with the proposal that 3rd person is the default form, as proposed by McCarthy (2007). Bruhn de Garavito (2003a, 2003b) analyzes 3rd person defaults as instances of nonfinite verbs that are missing an infinitival –r suffix. Although it is beyond the scope of this paper to take issue with this analysis, it should be noted that many of these default 3rd person forms were in the past tense, as in (9), where the irregular verb *ir* 'go' is used.

(9) Interviewer: Y también fuiste a España con tu clase de español. ('And you also went to Spain with your Spanish class.')
Participant: Sí. El semana antes. Sí. **Fue** a Madrid.
Yes. The week before. Yes. **Went-3sg** to Madrid
'Yes. The week before. Yes. I went to Madrid.'

For a more detailed description of these errors, and the extent to which they are compatible with a missing –r analysis, see McCarthy (to appear).

For finiteness, accuracy in finite emerges after accuracy in nonfinite; when errors in finiteness occur, they are nonfinite verbs occurring in finite contexts, as predicted by the Missing Surface Inflection Hypothesis (Prévost & White, 2000). One type of error that has been previously documented is the substitution of finite verbs in finite contexts (Mezzano, 2003). This error did not occur in the

data set, presumably because the learners were at a higher level than Mezzano's, who were at the initial state of L2 Spanish. Further investigation is warranted in this domain: at what point do errors become unidirectional? How can the transition from bidirectional errors to unidirectional ones be explained? One route is suggested in McCarthy (2008): prior to the acquisition of a marked feature value (such as [feminine] as a dependent of Gender), features are not represented in an asymmetrical fashion. This would mean that L2ers who are in the process of acquiring a category such as gender have not yet acquired the markedness relations of the L2; until they manage to acquire the marked feature, there is no preferred default form. For finiteness, there may be a parallel: prior to the acquisition of the marked value [finite], early-stage learners do not yet represent the finiteness features asymmetrically. This would predict that, for a period at the early stage of acquisition, substitution errors are bidirectional. Once the marked feature is acquired, substitution errors are unidirectionally nonfinite-for-finite. Though the details are still speculative, a feature geometric model offers a productive way of modeling developmental patterns in variation.

The acquisition of number is problematic for this proposal, because no development was observed in the marked feature of plural. L2ers at all levels substitute singular for plural. Plural agreement may remain problematic until more advanced levels. Further testing among more advanced speakers will help shed light on this issue.

A shortcoming of the present study is the lack of 2nd person contexts. Unfortunately, this is a product of the nature of the task; participants were asked questions and answered them. A follow-up study could force speakers to ask questions directed at the addressee, where 2nd person contexts would be forced.

To conclude: a feature-geometric model offers L2 researchers a way to explain variation and development in a principled way, and it places morphology within the domain of linguistic competence (see also Lardiere, 2005, 2009). By adopting this model or one like it, it is possible to talk about the development of morphology in a way that is independent of syntactic structure and operations.

References

Brown, Cynthia. (1997). The acquisition of segmental structure: consequences for speech perception and second language acquisition. Unpublished PhD dissertation, McGill University.

Bruhn de Garavito, Joyce. (2003a). The (dis)association between morphology and syntax: The case of L2 Spanish. In Silvina A. Montrul & Francisco Ordóñez (Eds.), *Linguistic Theory and Language Development in Hispanic Languages*, (pp. 398-417). Somerville MA: Cascadilla Press.

Bruhn de Garavito, Joyce. (2003b). Learners' competence may be more accurate than we think: Spanish L2 and agreement morphology. In Juana M. Liceras, Helmut Zobl & Helen Goodluck (Eds.), *Proceedings of the 6th Generative Approaches to Second Language Acquisition (GASLA 2006),* (pp. 17-23). Somerville, MA: Cascadilla Proceedings Project.

Clahsen, Harald, Martina Penke, & Teresa Parodi. (1993/1994). Functional categories in early child German. *Language Acquisition* 3(4), 395-429.

Cowper, Elizabeth. 2005. The geometry of interpretable features: Infl in English and Spanish. *Language* 81(1), 10-46.

Hanson, Rebecca, Heidi Harley, & Elizabeth Ritter (2000). Underspecification and universal defaults for person and number features. In Proceedings of the 2000 Canadian Linguistics Association. *University of Toronto Working Papers in Linguistics.* Electronic resource: http://linguistics.arizona.edu/~hharley/PDFs/HarleyRitterCLA2000.pdf (3 December 2010).

Harley, Harley & Elizabeth Ritter (2002). Person and number in pronouns: A feature-geometric analysis. *Language* 78(3), 482-526.

Haznedar, Belma & Bonnie D. Schwartz (1997). Are there optional infinitives in child L2 acquisition? In Elizabeth Hughes, Mary Hughes, & Annabel Greenhill (Eds.), *Proceedings of the 21st Annual Boston University Conference on Language Development (BUCLD), (*pp. 293-306). Somerville, MA: Cascadilla Press.

Lardiere, Donna (1998). Case and tense in the 'fossilized' steady state. *Second Language Research* 14(1), 1-26.

Lardiere, Donna (2005). On morphological competence. In Laurent Dekydtspotter, Rex A. Sprouse, & Audrey

Liljestrand (Eds.), *Proceedings of the 7th Generative Approaches to Second Language Acquisition (GASLA 2004),* (pp. 178-192). Somerville, MA: Cascadilla Proceedings Project.

Lardiere, Donna (2009). Some thoughts on the contrastive analysis of features in second language acquisition. *Second Language Research* 25(2), 173-227.

McCarthy, Corrine (2007). Morphological variability in second language Spanish. Unpublished PhD dissertation, McGill University.

McCarthy, Corrine (2008). Morphological variability in the comprehension of agreement: An argument for representation over computation. *Second Language Research* 24(4), 459-486.

McCarthy, Corrine (to appear). Modeling morphological variation and development: Person and number in L2 Spanish. To appear in *Linguistic Approaches to Bilingualism*.

Mezzano, Gina (2003). The development of Spanish verbal inflection in early stages of L2 acquisition. Undergraduate honors thesis, University of Illinois at Urbana-Champaign.

Montrul, Silvina A. (2004). *The Acquisition of Spanish: Morphosyntactic Development in Monolingual and Bilingual L1 Acquisition and Adult L2 acquisition*. Amsterdam: John Benjamins.

Oltra-Massuet, Isabel, & Karlos Arregi (2005). Stress-by-structure in Spanish. *Linguistic Inquiry* 36(1), 43-84.

Prévost, Philippe, & Lydia White (2000). Missing Surface Inflection or Impairment in second language acquisition? Evidence from tense and agreement. *Second Language Research* 16(2), 103-133.

Spanish Learner Language Oral Corpora (SPLLOC). (2010). SPLLOC2 Guided Interview task. Retrieved from <http://www.splloc.soton.ac.uk/splloc2/dataset/Interview/Year+10/>, <http://www.splloc.soton.ac.uk/splloc2/dataset/Interview/Year+13/> (9 July 2010).

Vainikka, Anne (1993/1994). Case in the development of English syntax. Language Acquisition 3(3), 257-325.

Vainikka, Anne, & Martha Young-Scholten (1994). Direct access to X'-theory: Evidence from Korean and Turkish adults learning German. In Teun Hoekstra & Bonnie D. Schwartz, (Eds.), *Language Acquisition Studies in Generative Grammar,* (pp. 265-316). Amsterdam: John Benjamins.

Vainikka, Anne, & Martha Young-Scholten (1996). Gradual development of L2 phrase structure. *Second Language Research* 12(1), 7-39.

Proceedings of the 11th Generative Approaches to Second Language Acquisition Conference (GASLA 2011)

edited by Julia Herschensohn
and Darren Tanner

Cascadilla Proceedings Project Somerville, MA 2011

ISBN 978-1-57473-445-4 library binding

Ordering information

Orders for the library binding edition are handled by Cascadilla Press.
To place an order, go to www.lingref.com or contact:

Cascadilla Press, P.O. Box 440355, Somerville, MA 02144, USA
phone: 1-617-776-2370, fax: 1-617-776-2271, sales@cascadilla.com

Web access and citation information

This entire proceedings can also be viewed on the web at www.lingref.com. Each paper has a unique document # which can be added to citations to facilitate access. The document # should not replace the full citation.

This paper can be cited as:

McCarthy, Corrine. 2011. Constraining Morphological Variation and Development: Agreement in L2 Spanish. In *Proceedings of the 11th Generative Approaches to Second Language Acquisition Conference (GASLA 2011)*, ed. Julia Herschensohn and Darren Tanner, 76-82. Somerville, MA: Cascadilla Proceedings Project. www.lingref.com, document #2547.