

How is inflectional morphology learned?

Roumyana Slabakova

University of Iowa

This article considers recent explanations of variability in the second language (L2) comprehension of inflectional morphology. The predictions of five accounts are spelled out: the emergentist account, the Feature Assembly Hypothesis, the Contextual Complexity Hypothesis, the Morphological Underspecification Hypothesis and the Combinatorial Variability Hypothesis. These predictions are checked against the results of an experimental study on the L2 acquisition of inflectional morphology (based on an extension of Slabakova and Gajdos 2008). English-native learners of German at beginning and intermediate proficiency levels took a multiple-choice test where they had to supply appropriate missing subjects. The predictions of the Morphological Underspecification Hypothesis and the Combinatorial Variability Hypothesis were largely supported by the experimental findings. It is argued that only accounts looking at mental representation of lexical features adequately explain L2 morphological variability.

1. Introduction

Variability in the production and comprehension of inflectional morphology is widely attested and well documented in L2 acquisition. Researchers investigating the nature of morphological variability have reported that L2 learners (both children and adults) often fail to produce target-like morphological forms, using instead what seem to be default forms in the respective paradigms. Although there is relative consensus on the reality of morphological errors, the sources of this variability are disputed. This article will probe further into possible sources of morphological variability in language comprehension, as captured by the choice of agreement endings in context, in English-German interlanguage. I open this discussion by presenting several theoretically diverse accounts of morphological variability.

Within the emergentist, or connectionist, account of second language (L2) acquisition (Ellis 2008; O'Grady 2005, as well as articles in special issues of two

EUROSLA Yearbook 9 (2009), 56–75. DOI 10.1075/eurosla.9.05sla
ISSN 1568–1491/E-ISSN 1569–9749 © John Benjamins Publishing Company

journals: *Lingua* 2008, 118, 4 and *Applied Linguistics* 2006, 27, 4) it is proposed that L2 learners are capable of extracting mental representations of grammatical phenomena by induction from input observation. Thus properties of the input encountered by L2 learners play a decisive role in the type of representations they establish. Associative (contingency) learning is mainly based on type and token frequency of forms in the input (Bybee 2001; Ellis 2002, 2006a, b). Order and accuracy of acquisition of inflectional morphology is determined by how robust the particular construction is in the ambient input. High frequency forms (including irregular forms) have lexically stronger representations than low frequency forms. In addition, other factors that can determine order and accuracy of inflectional morphology acquisition are: perceptual salience, semantic complexity, morpho-phonological regularity, and syntactic category (Goldschneider & DeKeyser 2001).

While the explanations of the emergentist approach to L2 morphological variability are abundantly clear and concrete, formalist generative approaches give a variety of explanations. According to the Missing Surface Inflection Hypothesis (Prévost & White 2000), variable production of the morphology may be due to learners' imperfect mapping of specific morphological forms to abstract categories. In such cases (which could be due to processing difficulties and more often attested under communication pressures) learners resort to defaults, forms that are underspecified in some features. However, the authors argue that L2 learners have acquired the relevant features of the terminal nodes in the syntax (from the L1 or on the basis of L2 input constrained by UG, as the case may be) and it is only the imperfect lexical access to the whole set of morphosyntactic features that brings forward L2 learners' variability in production.

Lardiere looks for the sources of morphological variability in the complexity of the morphosyntactic knowledge that has to be acquired by L2 learners. Her Feature Assembly Hypothesis (Lardiere 2007, 2008) maintains that the ways in which grammatical features are morphologically combined and conditioned present formidable learning problems in L2 acquisition. For example, primitive interpretable features are often clustered differently in different languages and knowledge of these form-to-meaning mappings constitutes a kind of morphological competence that has to be acquired by learners. This morphological competence involves figuring out how to reconfigure features into new or different formal configurations or remap native features onto new functional morphology. The prediction of this account is that the more re-assembly of features the L2 learner must do, the more difficulty she will face, and such morphemes (and meanings) will take longer to acquire. With respect to the particular paradigm whose acquisition is being tested in this experiment, the prediction is that members of the copula paradigm whose features are exactly the same in German and

English would be learned easier and chosen with greater accuracy as compared to verbal forms whose feature specifications are mismatched in the two languages.

Hawkins and Casillas' Contextual Complexity Hypothesis (Hawkins & Casillas 2008) is another recent attempt to provide a principled explanation of variability in emerging L2 morphosyntax. The authors propose that while native speakers have vocabulary entries specified in bundles of features at the point of lexical insertion (the terminal nodes), non-native speakers at early stages of development have vocabulary entries specified for the terminal nodes that co-occur with the exponent of that item. For example, to represent the *-s* of 3rd person singular agreement morphology, native speakers associate the form /s/ with any terminal node having features [V, -past, +singular, 3 person], while learners have something like: "insert /s/ in the context of a verb which is in the context of a non-past T, itself in the context of a 3rd person singular N" (p8). The more sister terminal nodes specified in a morpheme's environment, the more costly its lexical access will be. Underlying this proposal is the assumption that L2 learners do not always have access to uninterpretable features; specifically, when these same uninterpretable features do not come from their native language.

Another recent account of variability, the Morphological Underspecification Hypothesis (McCarthy 2007, 2008) argues that morphological variation comes from representational deficits; not from deficits in syntactic but in morphological representation. Therefore, L2 morphology errors in this account involve the systematic substitution of underspecified, representationally-simpler forms across comprehension and production. McCarthy found morphological variability in the comprehension of gender, at least for speakers at the intermediate level of proficiency. In her experimental study, masculine clitics surfaced as defaults in both production and comprehension. She argues against a strictly computational account of variability that places the source of variability in communication problems. Instead, she proposes that variability is better understood as involving a representational cause in the morphology, but not the syntax. A representational difference between morphological features like masculine and feminine (Harley and Ritter 2002) accounts for the difference in accuracy rates across comprehension and production. More specifically, since the masculine gender has a simpler morphological representation as compared to the feminine, it is predicted to be used as a default at beginning stages of acquisition.

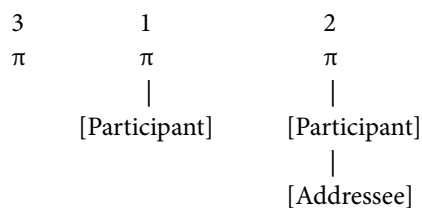
In this article, I will present in more detail two formal proposals of possible morphosyntactic variation. The first account is the Morphological Underspecification account based on the feature geometries of Harley and Ritter (2002), as adopted by McCarthy (2007, 2008). As discussed in the preceding paragraph, this account gives a principled reason why some forms become defaults in acquisition and others don't. The second theoretical account, the Combinatorial Variability

Hypothesis (CVH), Adger (2006), explains intra-personal morphosyntactic variation as arising from the combinatorial mechanisms of language itself. It offers an evaluation metric of uninterpretable feature combinations, predicting which forms are going to surface more, and may even become defaults, in the acquisition of inflection. I shall present the results of an experimental study on the comprehension of German subject-object agreement by native speakers of English who are beginning and intermediate learners of German (an extension of Slabakova and Gajdos 2008). I will argue that none of the L2 acquisition models trying to account for morphosyntactic variation I presented briefly in this introduction can do justice to the experimental results and to the generalizations that arise from them; however, the Morphological Underspecification account and the Combinatorial Variability account do better than the others. I will address McCarthy's (2008) claim that morphological variability seems to be a representational issue, and not just due to communication pressures or lexical access problems in production, as the Missing Surface Inflection Hypothesis would have it.

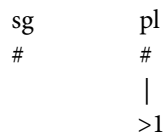
2. Morphological feature geometries applied to pronouns

Harley (1994), Bonet (1995), Harley & Ritter (2002), Cowper (2005), and Cowper & Hall (2004) have argued that morphological features are not represented as unstructured bundles. Instead, there are theoretical advantages to viewing the morphological representation of feature bundles as structured into a feature geometry of privative features. Feature geometries define natural classes for morphological features and a notion of complexity: the more structure, the more complex a representation is. Natural class nodes are used as organizing nodes. The feature geometries in (1) and (2) represent the structure of person and number features.

(1) Person features:



(2) Number features:



UNCORRECTED PROOFS

© JOHN BENJAMINS PUBLISHING COMPANY

Features are monovalent and only appear if they have a positive value. Markedness is encoded via a node-counting metric. The more marked a given feature combination is, the more nodes will be required to represent it. Singular is default in number, and 3 is default in person. However, Harley and Ritter (2002) postulate the following defaults:

- (3) a. Speaker for Participant – therefore 1st person acquired early.
- b. Minimal for Individuation – therefore singular acquired early.

The feature geometry explains language universals (Greenberg 1963) in terms of dependency of some features on others. For example, his Universal 32 postulates that ‘Whenever a verb agrees with a nominal subject or object in gender it also agrees in number.’ Universal 34 says that ‘No language has a dual [number] unless it has a plural.’ (Greenberg 1963:94). Such dependencies can be captured with a feature geometry. The feature geometry also predicts language acquisition orders, see above. Concrete predictions of this account are discussed in Section 5.

3. The Combinatorial Variability Hypothesis (Adger 2006)

According to current minimalist assumptions (e.g., Adger 2003) lexical items (LI) are built up of combinations of features in unstructured sets. Bivalent features are assumed by this theory. How does the learner decide on which features are instantiated in her language? The learner has access to the Conceptual Structure of human thought, which provides a range of possible features that have to be semantically motivated: e.g., number, participant in the discourse, etc. in pronouns. The particular forms of the LIs in the input tell the learner which available contrasts are actually marked in her (L1 or L2) language.

We now turn to the features necessary to classify personal pronouns. The feature [\pm singular] marks the number of the pronoun. The feature [\pm participant] marks whether the pronoun refers to a participant in the speech act. *I*, *we*, and *you* are [+participant]. The feature [\pm author] distinguishes between addressee and speaker. *I* and *we* are [+author]. The Feature Co-occurrence Restriction posits that a lexical item is specified for [\pm author] iff it is specified for [+participant]. This restriction comes from the semantics of English pronouns: there is no sense in which one can be an author of an utterance if one is not a participant. It follows that third person pronouns do not need to specify [\pm author].

UNCORRECTED PROOFS

© JOHN BENJAMINS PUBLISHING COMPANY

(4)	$\left[\begin{array}{l} \text{singular: +} \\ \text{participant: +} \\ \text{author: +} \end{array} \right]$	I	$\left[\begin{array}{l} \text{singular: -} \\ \text{participant: +} \\ \text{author: +} \end{array} \right]$	we
	$\left[\begin{array}{l} \text{singular: +} \\ \text{participant: +} \\ \text{author: -} \end{array} \right]$	you	$\left[\begin{array}{l} \text{singular: -} \\ \text{participant: +} \\ \text{author: -} \end{array} \right]$	you
	$\left[\begin{array}{l} \text{singular: +} \\ \text{participant: -} \end{array} \right]$	he/she/it	$\left[\begin{array}{l} \text{singular: -} \\ \text{participant: -} \end{array} \right]$	they

Certain lexical items carry features that are purely formal in nature. Their job is to establish syntactic dependencies. They are called uninterpretable (Chomsky 2001) and are marked *uF*. In a sentence like *He is tired*, the person features [\pm singular], [\pm participant], [\pm author] are interpretable on the pronoun. They value and then check the uninterpretable features on the *verb* in an agreement chain (Pesetsky & Torrego 2001).

The notion of agreement chain plays an important part in this theory. An agreement chain is defined as a pair of lexical items where the uninterpretable features of one item can be only a subset of the interpretable features of the other. Full Interpretation requires every uninterpretable feature to be in an agreement chain. (5) presents an agreement chain, where “...” stands for c-command.

(5)	$\left[\begin{array}{l} \text{singular: +} \\ \text{participant: +} \\ \text{author: +} \end{array} \right]$...	$\left[\begin{array}{l} \text{usingular: +} \\ \text{uparticipant: +} \\ \text{uauthor: +} \end{array} \right]$
-----	---	-----	--

Next, I shall give a schematic overview of how the theory works. Imagine that a lexical item, let's call it LI_1 , has three interpretable features as exemplified below in (6). This lexical item is able to combine with a range of other items bearing different subsets of uninterpretable features. For example, it can agree with a lexical item LI_2 that bears only uninterpretable F_1 , among other options. The phonetic form of LI_2 could be “x.”

(6)	$LI_1 \{F_1, F_2, F_3\}$...	$LI_2 \{uF_1\}$	→	PF (LI_2) = x
			$LI_3 \{uF_2\}$	→	PF (LI_3) = y
			$LI_4 \{uF_3\}$	→	PF (LI_4) = z

It is important to notice that all the uninterpretable features in LI_2 , LI_3 , and LI_4 correspond to some of the three interpretable features in LI_1 , even though their phonetic forms are different. “What allows the variability is the possibility that particular lexical items may be underspecified for the uninterpretable agreement features that they contain. This underspecification is irrelevant to the semantic system, since these features are not interpreted.” (Adger 2006: 510).

This system also predicts frequencies: if there are two ways for the grammar to output a specific PF, as is the case for *x* in the example in (7), but only one way to output a *y*, then *x* will appear more often than *y*, roughly in proportions 66% vs. 33%.

$$(7) \quad \begin{array}{l} \text{LI}_1 \{F_1, F_2, F_3\} \dots \text{LI}_2 \{uF_1\} \\ \text{LI}_3 \{uF_2\} \\ \text{LI}_4 \{uF_3\} \end{array} \quad \begin{array}{l} \rightarrow \\ \rightarrow \\ \rightarrow \end{array} \quad \begin{array}{l} \text{PF} (\text{LI}_2) = x \\ \text{PF} (\text{LI}_3) = x \\ \text{PF} (\text{LI}_4) = y \end{array}$$

Here is a concrete example of how variability arises. Consider the uninterpretable features that make the matrices of the present forms of *to be*. A child acquiring the features of the pronominal subject determines them on the basis of interpretation. The question arises of what the features of the auxiliary are. In order to check agreement, the child needs to distinguish between *is*, *am* and *are*. For example, the child could go for the complete specification, as given in (8).

$$(8) \quad \begin{array}{l} \left[\begin{array}{l} \text{usingular: +} \\ \text{uparticipant: +} \\ \text{uauthor: +} \end{array} \right] \text{ am} \\ \left[\begin{array}{l} \text{usingular: +} \\ \text{uparticipant: +} \\ \text{uauthor: -} \end{array} \right] \text{ are} \\ \left[\begin{array}{l} \text{usingular: +} \\ \text{uparticipant: -} \end{array} \right] \text{ is} \end{array} \quad \begin{array}{l} \left[\begin{array}{l} \text{usingular: -} \\ \text{uparticipant: +} \\ \text{uauthor: +} \end{array} \right] \text{ are} \\ \left[\begin{array}{l} \text{usingular: -} \\ \text{uparticipant: +} \\ \text{uauthor: -} \end{array} \right] \text{ are} \\ \left[\begin{array}{l} \text{usingular: -} \\ \text{uparticipant: -} \end{array} \right] \text{ are} \end{array}$$

The matrix in (8) offers fully specified LIs, but there is a large amount of homonymy and hence, generalizations are missed while learning becomes more cumbersome. For example, all the [*uauthor: -*] forms get spelled as *are*, and so do all the [*usingular: -*] forms. The algorithm of generalization is based on the premise that LIs with fewest features are best. Going through the whole paradigm and applying underspecification of the feature content of the verb, we get the following features and spellout form combinations:

$$(9) \quad \begin{array}{l} [\text{uauthor: -}] \text{ are} \\ [\text{usingular: -}] \text{ are} \\ [\text{usingular: +, uauthor: +}] \text{ am} \\ [\text{usingular: +, uparticipant: -}] \text{ is} \end{array}$$

The system also uses some additional filters (Adger 2006: 518). The following simultaneous constraints work in the lexicon:

- (10) a. Reject Optionality (an LI is kept if there is always a matching form)
- b. Reject Synonymy (an LI is kept if it does not create synonyms)
- c. Minimize the size of the Lexicon

The reader is referred to the original publication for details on how these are implemented. In the next section, I apply this underspecification procedure to calculate the minimally necessary features for the German copula *sein*.

4. German copula *sein* ‘be’ in the present tense

The uninterpretable features of the German verb *sein* ‘be’ in the present tense were analyzed using the evaluation metric and various filters described above. (11) illustrates the paradigm of *sein* in the present tense.

- | | | | | |
|------|--------|------------|--------|----------|
| (11) | 1. sg. | Ich bin | 1. pl. | wir sind |
| | 2. sg. | du bist | 2. pl. | ihr seid |
| | 3. sg. | er/sie ist | 3. pl. | sie sind |

To come up with the lexical specifications of features that are maximally generalized, one should first go through the unary feature specifications, e.g., [*usingular*: +]. However, none of them is sufficient to describe exactly one LI. Then we consider the binary specifications:

- | | | |
|------|----|---|
| (12) | a. | [<i>usingular</i> : +, <i>uauthor</i> : +] <i>bin</i> |
| | b. | [<i>usingular</i> : +, <i>uauthor</i> : -] <i>bist</i> |
| | c. | [<i>usingular</i> : +, <i>uparticipant</i> : +] <i>bist</i> , <i>bin</i> REJECT |
| | d. | [<i>usingular</i> : +, <i>uparticipant</i> : -] <i>ist</i> |
| | e. | [<i>usingular</i> : -, <i>uauthor</i> : +] <i>sind</i> |
| | f. | [<i>usingular</i> : -, <i>uauthor</i> : -] <i>seid</i> |
| | g. | [<i>usingular</i> : -, <i>uparticipant</i> : +] <i>sind</i> , <i>seid</i> REJECT |
| | h. | [<i>usingular</i> : -, <i>uparticipant</i> : -] <i>sind</i> |
| | i. | [<i>uparticipant</i> : +, <i>uauthor</i> : -] <i>seid</i> , <i>bist</i> REJECT |
| | j. | [<i>uparticipant</i> : +, <i>uauthor</i> : +] <i>sind</i> , <i>bin</i> REJECT |

When a two-feature specification is responsible for two lexical items, the evaluation metric rejects it under (10b), reject synonymy. One form, namely *sind*, is output by two underspecified combinations of features. The rest of the forms are uniquely identified.

- | | | |
|------|--|------------------|
| (13) | [<i>usingular</i> : -, <i>uauthor</i> : +] | → PF <i>sind</i> |
| | [<i>usingular</i> : -, <i>uparticipant</i> : -] | → PF <i>sind</i> |

Thus we arrive at feature specifications of copula *sein* present forms which are capable of predicting variability within a single system through a “non-deterministic choice of form” (Adger 2006: 505).

5. Predictions for accuracy in the beginning stages of L2A

In this section, I spell out the L2 acquisition predictions of the various syntax-theoretic and L2-theoretic accounts discussed in the introduction and Sections 2 and 3.

The Combinatorial Variability Hypothesis predicts that the copula form *sind* would account for a larger proportion of the errors in acquisition. Since *sind* is output by two different feature combinations (see (13) above), there are two agreement chains that it participates in, while all the other forms of *sein* are only output by one agreement chain. Therefore, *sind* is more complex to calculate. If L2 learners are guided by the same universal feature evaluation metric as German acquiring children are, then we expect English learners of German to demonstrate evidence of overusing *sind* in the process of learning the target agreement morphology, and hence, making more errors with it.¹

The Morphological Underspecification hypothesis predicts that either 1st sg or 3rd sg agreement could appear with highest accuracy in the grammar of the learners, but agreement with 2nd sg will be less accurate. Harley and Ritter (2002) propose that Speaker is the default interpretation of a bare [participant] node; however, they also stipulate that 3rd person singular can be a default, too. The authors survey ten L1A studies, and find that (i) the pronouns which appear initially are 1st person singular or 3rd singular neuter/inanimate; and (ii) 2nd person appears after 1st person. All of the plural agreement should be even less accurate.

From the emergentist point of view, if it is robustness of presence in the input that gives rise to stronger representations, hence less errors, then we would expect the following hypothetical error rates as in Table 1, based solely on frequency. *A Frequency Dictionary for German*, Randal Jones and Erwin Tschirner, Routledge (2006) was used to get the frequency of the LIs. The dictionary is based on the Leipzig/BYU Corpus of 4.2 million words of spoken and written German, which contains one million each of spoken language, literature, newspapers, and academic texts, and 200,000 words of instructional language. The frequencies presented in Table 1 are per one million words. If the form *ist* appears about 10200 times per million words and we stipulate a minimal error rate of 1.5 to 2%, then the form *sind* which appears three times less often would have an error rate of 4.5 to 6%, etc. If frequency were the sole source of error rates, then these rates would be in inverse proportion to the factor of the frequency differences.

UNCORRECTED PROOFS
© JOHN BENJAMINS PUBLISHING COMPANY

Table 1. Frequency of verb forms in German corpus and hypothetical error rates

Form	Frequency	Expected error %
ist	10,229	1.5–2
sind	3,835	4.5–6
bin	1,025	15–20
bist	260	60–80
seid	36	> 100

The Feature Assembly Hypothesis would predict no insurmountable difficulty for English native speakers in learning the German copula *sein* forms, since the L1 and the L2 in this particular case make use of exactly the same features, so no re-assembly across categories is necessary. However, although the features are the same, they are not combined in the same way. If we apply the under-specification calculus to the same three features involved: [author], [singular] and [participant], we come up with different feature specifications, at least in four of the six members of the copula. The feature comparisons are given in (14); the mismatches between the feature specifications are given in bold.

(14)	<u>English forms of <i>to be</i></u>	<u>German forms of <i>sein</i></u>
	1sg. [usingular: +, uauthor: +] <i>am</i>	[usingular: +, uauthor: +] <i>bin</i>
	2sg. [uauthor: -] <i>are</i>	[usingular: +, uauthor: -] <i>bist</i>
	3sg. [usingular: +, uparticipant: -] <i>is</i>	[usingular: +, uparticipant: -] <i>ist</i>
	1pl. [usingular: -] <i>are</i>	[usingular: -, uauthor: +] <i>sind</i>
	2pl. [usingular: -] <i>are</i>	[usingular: -, uauthor: -] <i>seid</i>
	3pl. [usingular: -] <i>are</i>	[usingular: -, uparticipant: -] <i>sind</i>

If a beginning learner is simply mapping morphemes (1st singular to 1st singular, 2nd singular to 2nd singular; that is, *am* to *bin*, *are* to *bist*, etc.), she will run into some perfect matches and some mismatches. Within the Feature Assembly Hypothesis, it is reasonable to predict that the forms in which there is some re-assembly of features involved would pose a greater difficulty. Thus, it could be expected that the 1st and 3rd singular forms would present least difficulty, with the rest of the copula forms not differentiated in difficulty.

Finally, the Contextual Complexity Hypothesis predicts no significant difficulty with this learning task, since the number of terminal nodes in whose context the forms of *sein* appear is exactly the same, so we should not expect any difference in error rates. The latter hypothesis predicts that error rates in acquisition of the copula should be exactly the same for all forms. Table 2 lists the predictions; the Missing Surface Inflection Hypothesis is excluded because it makes no concrete predictions about comprehension of inflectional morphology. However, the latter hypothesis will be discussed further in Section 7.

Table 2. Predictions of several linguistic theories for the relative difficulty in copula *sein* forms L2 acquisition by English native speakers

Account	Important factor affecting L2A	Predictions
Emergentism	frequency of morphemes	Most errors with <i>seid</i> , fewest errors with <i>ist</i> (see Table 1)
Feature Assembly (Lardiere 2008)	mismatch between L1 and L2 morpheme feature mappings	Fewest errors with <i>bin</i> or <i>ist</i> , more errors with rest of the forms
Contextual Complexity (Hawkins & Casillas 2008)	terminal nodes co-occurring with morphemes	No difference in error rates for different forms
Morphological Underspecification (Harley & Ritter 2002)	feature geometry errors with <i>sind</i> and <i>seid</i>	Fewest errors with <i>bin</i> or <i>ist</i> , more errors with <i>sind</i> and <i>seid</i>
Combinatorial Variability (Adger 2006)	underspecified (uninterpretable) feature representation of forms	Errors with <i>sind</i> double those with other copula forms

6. The experimental study

6.1 Participants, testing procedure and materials

Twenty-four beginner and 18 intermediate learners of German participated in the experimental study (a preliminary version of the results was also reported in Slabakova and Gajdos 2008). They were for the most part undergraduate students. Participants answered a background questionnaire and took a written test. The learners' proficiency levels were established based on the number of class hours of exposure to German instruction at a US university. At the time of study, the beginners were exposed to roughly 40 hours of German classroom instruction; the intermediate learners, to 140 hours.

The test contained simple sentences with missing subjects. Participants had to choose which subject (out of four options) went well with the provided sentence. They could choose more than one option and they were provided with an example which showed more than one correct choice. About 30% of experimental items were such that more than one of the answers was correct. The test contained six items for each form of *sein* and 10 fillers with other verbal forms, for a total of 40 items. (15) provides an example test item.

UNCORRECTED PROOFS
© JOHN BENJAMINS PUBLISHING COMPANY

- (15) _____ bist ein guter Freund.
 are a good friend
- Moritz
 du 'you' ← the only correct choice
 die Schüler 'the students'
 er 'he'

6.2 Results

It was considered that the test participant made an error if she chose an unsuitable subject for the copula, for example “Moritz” in example (15) above, but also if she neglected to choose a correct subject, that is, if she did not choose “du”. Tables 3 and 4 give the percentage of errors for the beginning and the intermediate learners, respectively.

At this point, it seems that most errors are with *ist* and with *sind* forms. However, it was also noticed that the error rates in choosing pronoun subjects versus choosing full DP subject varied dramatically.² Table 5 illustrates this discrepancy.

Where may this discrepancy come from? In order to understand better what exactly is involved in subject-verb agreement, I discuss the agreement mechanism provided in Adger (2003: 220–222). In outline, the subject in Spec of ν P values the uninterpretable phi feature of T. The features of T, in their turn, value the

Table 3. Percentage errors by verb form for beginning learners

Type of error	<i>bin</i>	<i>bist</i>	<i>ist</i>	<i>seid</i>	<i>sind</i>
Not choosing correct subject	2.1	3.5	8.9	12.5	18.2
Choosing wrong subject	4.82	4.86	9.5	13.05	12.5
Combined errors	3.46	4.18	9.2	12.75	15.35

Table 4. Percentage errors by verb form for intermediate learners

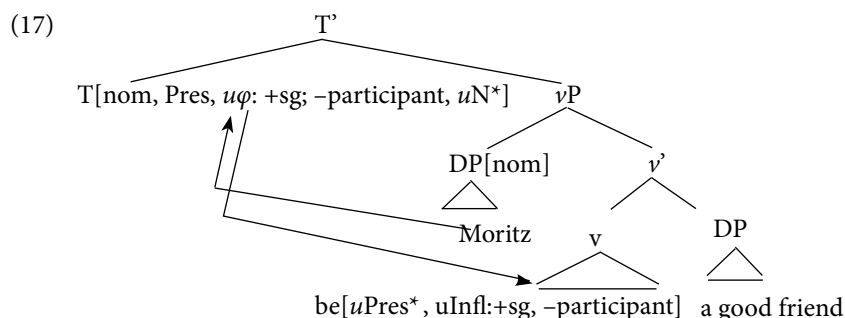
Type of error	<i>bin</i>	<i>bist</i>	<i>ist</i>	<i>seid</i>	<i>sind</i>
Not choosing correct subject	1	3.7	16.7	1.9	15.2
Choosing wrong subject	0.6	1.5	5.5	4.8	5.1
Combined errors	0.8	2.6	11.1	3.35	10.15

Table 5. Percentage errors in all forms of *sein* depending on type of subject

Type of error	Beginners	Intermediate
Errors in choosing correct pronoun subjects	7.50	4.50
Errors in choosing correct DP subjects	20.18	29.80

uninterpretable Infl feature on little v . To take a concrete example with an English sentence as in (16), the following derivation would obtain as in (17):

(16) Moritz is a good friend.



The copula is merged in a version of little v that takes the subject in its specifier and a predicate DP as its complement (Adger 2006: 196–197). The agreement mechanism involves two steps. First, the interpretable features of the subject (in our example (16)) value the uninterpretable phi features on T. As a second step, T values the uninterpretable Infl feature on the copula. Thus the copula receives the pronunciation “is”. Later on, the subject moves to the specifier of TP to satisfy a strong EPP feature [uN^*] and the copula moves to T to satisfy a strong tense feature [$uPres^*$]. The same mechanism would obtain for the equivalent German sentence (see (15)).

Keeping this mechanism in mind, I stipulate that the beginning and intermediate learners in this experiment had trouble calculating the *interpretable* features of the DP subjects in the first step of the agreement mechanism, namely, valuing features on T. Even the intermediate learners were six times more accurate picking pronominal subjects than DP subjects. It seems that the interpretable features on the pronouns were better acquired at the time of the testing.

Another, and related, explanation of the DP versus pronominal subject discrepancy has to do with rote learning. The forms for the 1st and 2nd person singular copula are most often encountered in the input together with equivalent pronouns for *I* and *you-sg*. In a language classroom, they are learned in a paradigm, by rote. That is why, it was conjectured that learners create the easiest lexical associations: *I + am*, *you + are*, etc. On the other hand, the form for 3rd person singular most often appears with overt subject DPs like *the teacher* or *the student*. The probability of *ist* combining with any DP subject instead of *er* or *sie* is much higher (in comparison to 1st and 2nd person forms).³ What is more, the

Table 6. Percentage errors and standard deviations on **pronoun** subjects

Learner group	<i>bin</i>	<i>bist</i>	<i>seid</i>	<i>sind</i>
Beginners	4.82 (0.13)	4.86 (0.16)	7.6 (0.19)	10.4 (0.23)
Intermediate	0.8 (0.09)	2.6 (0.12)	1.9 (0.12)	8.4 (0.23)

paradigm learned by rote does not provide the learner with DP subjects for the forms of the copula, the agreement has to be calculated based on the interpretable feature of the DP subject. But if this paradigm memorization explanation is on the right track, then the errors with *ist* are disproportionately affected by the type of the subject. Therefore, as a final look at the results, it was decided to abstract away from the errors with DP subjects, and therefore errors with *ist*. The final look at the data in Table 6 only calculates the errors in supplying pronoun subjects.⁴

One-way ANOVA reveals a significant difference between accuracy on different forms of *sein*, $F(3, 527) = 3.41, p = 0.02$ for the beginners, $F(3, 395) = 2.84, p = 0.04$. The beginning learners make roughly half the percentage of errors with *bin*, *bist* and *seid*, compared to *sind*, while the intermediate learners improve accuracy on all other forms but *sind*, thereby increasing the accuracy gap.

7. Discussion

I will summarize the interesting observations that arise from the experimental findings. Beginning learners do make errors in choosing the correct form of the copula. Over 15% of errors, the highest percentage, are with the form *sind*. The intermediate learners do not get much better with *sind*, compared to the beginners, after a considerably longer exposure to the language (see Table 6). The intermediate learners demonstrate even worse accuracy than the beginners when they have to combine a DP subject with the copula (see Table 5). On the crucial *sind* forms, the intermediate and the beginning learners are not sufficiently different by two tailed t-test ($P = 0.75$). The error rates with *sind* are roughly double those with *bin* and *bist* when subjects are pronouns. I will consider some possible explanations of these findings and check the predictions of the various morphological variability accounts summarized in Table 2.

Starting with the account that predicts no variation between the different members of the *sein* paradigm, the Contextual Complexity Hypothesis cannot explain these findings, since the number of terminal nodes in whose context the forms of *sein* appear is exactly the same, so we should not expect any difference in error rates. This hypothesis predicts that error rates in acquisition of the copula forms should be exactly the same. Why then the persistent difficulty?

Table 1. Frequency of verb forms in German corpus and hypothetical error rates

Form	Frequency	Expected error %
ist	10,229	1.5–2
sind	3,835	4.5–6
bin	1,025	15–20
bist	260	60–80
seid	36	> 100

Table 7. Frequency of verb forms in German textbook and hypothetical error rates

Form	Token frequency	Expected error %
ist	480	1.5–2
sind	73	6–12
bin	11	48–96
bist	11	48–96
seid	3	< 100

The emergentist explanation is not supported either. Table 1 is repeated here for ease of reference. These frequencies, as calculated in a frequency dictionary, pertain to the representative samples of native German.

But what if the frequency of the *sein* forms is different in the instructional input to which the learners are exposed? In Slabakova and Gajdos (2008), we also counted the occurrences of the forms in the actual textbook which the learners use in class. Table 7 presents the results for the counts from half of the textbook, six out of twelve chapters, the odd number chapters only.

The frequency of the *sein* forms in the textbook came out in roughly the same proportions to each other as in the general language frequency dictionary. That is, *ist* is by far the most frequent, followed by *sind* which is six times less frequent, etc. Note that even if we do not abstract away from DP subjects, as we do in Table 6, frequency alone cannot explain the error rates in Tables 4 and 5. There is not even a tendency of a correlation between frequency and accuracy. For example, *seid* is 160 times less frequent than *ist* but the error rates of beginning learners with these forms are approximately 9% and 12%, respectively. In addition, none of the factors put forward by Goldschneider & DeKeyser (2001), namely, perceptual salience, semantic complexity, morpho-phonological regularity, and syntactic category, are relevant for explaining the error rate discrepancies, as these factors do not distinguish between the forms of *sein*.⁵

The predictions of the Feature Assembly Hypothesis are not strongly supported. According to this account, it was expected that 1st singular *bin* and 3rd singular *ist* would be the most accurate forms, since their feature specifications

match in English and German. However, as one can ascertain in Tables 3 and 4, 1st singular *bin* and 2nd singular *bist* have more or less the same accuracy rates, while 3rd singular *ist* seems to present considerably more difficulty. Of course, as mentioned above, accuracy on *ist* may be skewed by the much lower accuracy on the interpretable features of the subjects (see Section 6.2, specifically discussion around example (16)). Furthermore, accuracy on the rest of the forms, namely, *bist*, *seid* and *sind*, is not roughly equal, although an equal amount of re-assembly would be required for their re-mapping in the English-German interlanguage.

The first prediction of the Morphological Underspecification Hypothesis is supported: 1st person singular agreement is chosen more accurately than 3rd person singular (Tables 3, 4). Its second prediction is not supported: 2nd person singular is not chosen with less accuracy compared to 1st; in fact, 1st and 2nd are equal in accuracy. Its third prediction is also roughly supported, but only by the beginning learners: plural agreement is least accurately supplied by these learners.

The prediction of the Combinatorial Variation Hypothesis is also supported: the error rates with *sind* are roughly double those with *bin* and *bist* when subjects are pronouns. Even if we do not abstract away from DP subjects but consider them in the mix, as in Tables 3 and 4, errors with *sind* are the highest percentages.

The issue of L1 transfer and the related issue of classroom instruction play very interestingly in these results. It would seem that one of the first things that classroom learners are exposed to in a German classroom is the paradigm of *sein* 'to be' in the present tense, right after chunks like 'What is your name?' and 'My name is X.' What is more, the copula paradigm has the same array of phi features participating in the feature matrices. The copula forms are extremely frequent, there is ample exposure to them in the classroom and in the textbook, why then such error rates in a relatively simple test? Even more puzzling is the fact that there does not seem to be much progress in the development of these morphemes, and a hundred more hours of classroom exposure makes little difference with respect to accuracy of choice. Findings such as the ones in this experiment seem to point to the generalization that even morphemes that are very abundant in the input and that are introduced and drilled in language classrooms may pose learning difficulties for learners, and that these difficulties may persist even at the intermediate stages.

Why doesn't simple transfer of the copula paradigm obtain? At a superficial level, nothing could be simpler. English forms are *am*, *is*, *are*, that is, there are three phonological forms. German forms are *bin*, *bist*, *ist*, *sind*, *seid*, that is, five forms. Morphemes could be mapped between the two languages based on person and number specifications. There are no other features to re-assemble, as for example [dual] in Hopi in addition to the [singular] and [plural]. The answer to the puzzle lies in the fact that the copula forms are not just any lexical items

like *hat* or *car*; they are inflectional morphology carrying bundles of features. As I showed earlier in (14), there is no simple morphological mapping available between them, especially when we consider their feature specifications. Applying the underspecification calculus to the same three features involved: [author], [singular] and [participant], we came up with different feature specifications in four of the six members of the copula. The predictions of the Feature Assembly Hypothesis proved to be too broad-brush. On the other hand, the predictions of the two accounts that use linguistic mechanisms for calculating agreement, the Morphological Underspecification Hypothesis and the Combinatorial Variability Hypothesis, were largely supported.

Both accounts that were supported by the data are basically feature underspecification accounts, although they use different evaluation metrics. The Morphological Underspecification Hypothesis relies on a universal feature geometry; the Combinatorial Variability Hypothesis uses a paradigm-internal calculation mechanism that employs universal underspecification rules. The common trait of these accounts is that they use a detailed understanding of the human language system to explain intriguing acquisition facts. Neither simple surface transfer nor language classroom exposure can explain these facts.

The final point I address is one raised by McCarthy (2008: 482–483), namely, whether results such as the ones obtained in this experiment support a representational basis for morphological variability. McCarthy tests comprehension and production of Spanish gender and number agreement in clitics and adjectives in the interlanguage of English native speakers. She argues that her findings point to “L2 speakers’ representations [being], at some cases and particularly at lower proficiency levels, deficient” (p. 483). Her account differs from the Missing Surface Inflection Hypothesis (Prevost & White 2000), according to which “variability in production is attributed to problems of lexical access; difficulties arise when the pressure to communicate is strong” (McCarthy 2008: 464). My findings are compatible with McCarthy’s representational explanation and less compatible with Prevost and White’s lexical access under communication stress explanation.⁶

Learners in the experiment described in this article do not clearly resort to defaults as compared to McCarthy’s learners who choose the masculine gender as default, thereby demonstrating that they have not built the feminine node in their feature geometry. The present experimental task involves sentence comprehension and multiple choice, a relatively easy task in which lexical access is supposed to be less taxing than in free language production. Participants in our experiment were under no communicative pressure. (Nevertheless, some lexical access is always implicated.) There is not much evidence that our learners use any of the copula forms as defaults. However, our L2 learners showed elevated error rates with the forms whose features are more complex to compute. Even more

indicative in this respect is the accuracy discrepancy between DP and pronominal subject choices, attested at both proficiency levels, which is a matter of lexical access and feature computation. In sum, elevated error rates with the featurally more complex forms, in comprehension and under no elevated communicative pressure, may point to lexical representations that are in flux. Thus, a representational account of variable accuracy is supported.

A methodological shortcoming of the Slabakova and Gajdos (2008) experimental study is that the test items contained the copula forms and asked the learners to supply the possible subjects that can go with this form. This was done because it was felt that providing the subject and leaving the copula form blank taps directly into metalinguistic knowledge or chunk learning as reinforced by textbook exercises. The current test design allows for multiple choices in a single test item and calls for an evaluation of each of the options as a separate sentence. The optimal research design would have both versions of the test, with missing copula forms and with missing subjects, and compare the findings. This is left for further research.

In conclusion, I have argued in this article that there may be different underlying sources for variable L2 morphosyntactic acquisition and performance. Variation, however, stems from the grammatical representations within the learners' interlanguage grammars, and is not at all random (see also McCarthy 2007, 2008). Together with prosody and feature re-assembly, combinatorial variability and morphological markedness are also part of the picture. Frequency, however, cannot be not the whole picture.

Acknowledgements

This article discusses some findings of an experiment that evolved out of a generative SLA course project. I am grateful to Johnathan Gajdos for creating the experimental materials, running the subjects, and especially reading through the German textbooks while counting copula forms. The participants in the experimental study are gratefully acknowledged. Portions of this experimental work and its theoretical treatments were presented at the GALA 2007 conference in Barcelona; the SLRF 2007 conference at the University of Illinois, Urbana-Champaign; and at EUROSLA 2008 in Aix-en-Provence. Audiences at all three conferences asked useful and insightful questions. Special thanks go to David Adger for discussing the Combinatorial Variability Hypothesis over email. Last but not least, I am indebted to Leah Roberts and two anonymous reviewers for suggesting clarifications and pointing out implications I had missed. All remaining errors are mine.

Notes

1. While I am not aware of any studies on child German copula acquisition that count the different copula forms, overuse of *sind* would be the prediction of the Combinatorial Variability Hypothesis. At this point in time, this prediction remains just that, a prediction. Thanks go to an anonymous reviewer for asking the question.
2. It was not the case that DP subjects only occurred with 3rd person copula forms. Some full DPs occurred in sentences that featured 2nd person copula.
3. Although this claim is anecdotal, see note 5.
4. An anonymous reviewer questions the decision to lump pronominal subjects *sie* 'she' and *wir* 'we' together in the calculation of the accuracy rate with *sind* on the ground that *sie* is homophonous: the same form means 'she' 3rd person singular feminine as well as 'they', 3rd person plural (see paradigm in (11) above). It was suggested that this ambiguity of the *sie* form may have confused the learners. That is why accuracy was also calculated separately. The beginners were less accurate supplying *wir* 'we' subjects with a 12.5% error rate while they were more accurate with *sie* 'they' subjects, where the error rate was 8.33. In the intermediate learner group, however, accuracy was reversed: they made 5.6% errors with *wir* and 11.2% errors with *sie*. Thus, the ambiguity of *sie* does not appear to have much of an effect on pronoun choice for the copula form *sind*.
5. In addition, one should look at the consistency of the subject-copula form collocation in the German language. For example, how often does *ich* appear with *bin*, *du* with *bist*, etc. To the best of my knowledge, such a collocation count is not available at this time.
6. Of course, in order to definitively tease the predictions of the two accounts apart, one must compare production and comprehension data as McCarthy does. The Slabakova and Gajdos (2008) experimental task does not allow such a comparison.

References

- Adger, D. (2006). Combinatorial Variability. *Journal of Linguistics* 42, 503–530.
- Bonet, E. (1995). Feature structure of Romance clitics. *NLLT* 13, 607–647.
- Bybee, J. (2001). *Phonology and Language Use*. CUP.
- Chomsky, N. (2001). Derivation by phase. In Kenstowicz, M. (Ed.), *Ken Hale: A Life in Language*, pp. 1–52. Cambridge, MA: MIT Press.
- Cowper, E. (2005). The geometry of interpretable features: Infl in English and Spanish. *Language* 81, 10–46.
- Cowper, Elizabeth, en Daniel Currie Hall. (2004). The pieces of π . *Proceedings of the 2004 annual conference of the Canadian Linguistic Association*. In Junker, M.-O., McGinnis, M. and Roberge, Y. (Eds).
- Ellis, N. C. (2002). Frequency effects in language processing: a review with implications for theories of implicit and explicit language acquisition. *Studies in Second Language Acquisition* 24, 143–188.
- Ellis, N. C. (2006a). Language acquisition as rational contingency learning. *Applied Linguistics* 27(1), 1–24.

- Ellis, N. C. (2006b). Selective attention and transfer phenomena in L2 Acquisition: Contingency, cue competition, salience, interference, overshadowing, blocking, and perceptual learning. *Applied Linguistics* 27(2), 164–194.
- Ellis, Nick (2008). The Dynamics of Second Language Emergence: Cycles of Language Use, Language Change, and Language Acquisition. *The Modern Language Journal* 92, 2, Summer, 232–249.
- Goldschneider, J. M. and R. DeKeyser. (2001). Explaining the ‘natural order of L2 morpheme acquisition’ in English: A meta-analysis of multiple determinants, *Language Learning* 51, 1–50.
- Greenberg, Joseph H. (1963). Some universals of grammar with particular reference to the order of meaningful elements. *Universals of language*, ed. by Joseph H. Greenberg, 73–113. Cambridge, MA: MIT Press.
- Hawkins, R. & Casillas, G. (2008). Explaining frequency of verb morphology in early L2 speech. *Lingua* 118(4), 595–612.
- Harley, H. and Ritter, E. (2002). A feature-geometric analysis of person and number. *Language* 78(3), 482–526.
- Jones, R. & Tschirner, E. (2006). A frequency dictionary of German. Routledge.
- Lardiere, D. (2007). *Ultimate Attainment in Second Language Acquisition. A case Study*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Lardiere, D. (2008). Feature assembly in second language acquisition. In Liceras, J., H. Zobl and H. Goodluck (Eds.), *The Role of Features in Second Language Acquisition*. Mahwah, New Jersey: Lawrence Erlbaum Associates.
- McCarthy, C. (2007). *Morphological Variability in Second Language Spanish*. PhD Dissertation, McGill University.
- McCarthy, C. (2008). Morphological variability in the comprehension of agreement: an argument for representation over computation. *Second Language Research* 24(4), 459–486.
- O’Grady, W. (2005). *Syntactic carpentry: An emergentist approach to syntax*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Pesetsky, D. & Torrego, E. (2001). T-to-C movement: causes and consequences. In Kenstowicz, M. (Ed.), *Ken Hale: A Life in Language*, pp. 355–426. Cambridge, MA: MIT Press.
- Prévost, P & White, L. (2000). Missing surface inflection hypothesis or impairment in second language acquisition? Evidence from tense and agreement. *Second Language Research*, 16, 103–133.
- Slabakova, R. and J. Gajdos (2008). “Variable morphology in L2 Acquisition: Facts and Explanations”. In *Selected Proceedings from SLRF 2007*, Melissa Bowles, Rebecca Foote and Silvia Perpiñán (Eds.), Cascadilla Proceedings Project.
- Smith, J. (2000). *Synchrony and diachrony in the evolution of English: Evidence from Scotland*. PhD dissertation, University of York.