Learning datives: The Tolerance Principle in monolingual and bilingual acquisition

Charles Yang
University of Pennsylvania, USA

Silvina Montrul
University of Illinois at Urbana-Champaign, USA

Abstract
We study the learnability problem concerning the dative alternations in English (Baker, 1979; Pinker, 1989). We consider how first language learners productively apply the double-object and *to*-dative constructions (*give the book to library/give the library the book*), while excluding negative exceptions (*donate the book to the library/donate the library the book*). Our solution for first language acquisition is based on The Tolerance Principle, a formal model that detects productivity from the distributional properties of the input data (Yang, 2005, 2016). This principle predicts an acquisition stage where the constructions are productive, followed by a stage where learners retreat from overgeneralization and form more finely grained rules. This work calls for a formally rigorous model of acquisition, which can incorporate input effects and retain the benefits of an abstract generative grammar without resorting to piecemeal learning. We provide an analysis of child-directed speech in the CHILDES database to support the learning proposal for first language acquisition, while considering its potential applicability to second language (L2) acquisition and first language (L1) attrition.

Keywords
dative alternation, English, exceptions L1 acquisition, L1 attrition, L2 acquisition, productivity

Corresponding author:
Charles Yang, University of Pennsylvania, 609 Williams Hall, Philadelphia, PA 19104, USA.
Email: charles.yang@ling.upenn.edu

I Parameters and generalizations

Human language has turned up some of the hardest challenges for learning sciences. For all the progress in natural language processing, backed by an unfathomably large volume of data and computational prowess, current systems are still nowhere near the accuracy of human language learning and use. The feat of language acquisition is even more impressive when considered in a psychological setting; an infinite range of grammatical expressions must be projected from a finite amount of primary linguistic data, giving rise to the logical problem of language acquisition that challenges first language (L1) and second language (L2) learners alike (Hornstein and Lightfoot, 1981; White, 2003).

A unifying approach to linguistic complexity and language acquisition has been the driving force in modern linguistic theories (Chomsky, 1965). The Principles-and-Parameters framework sets out a universal template of language variation, and the language-particular choices can be determined by the primary linguistic data. We will not review the extensive research on parameter setting (Dresher, 1999; Fodor, 1998; Gibson and Wexler, 1994; Niyogi and Berwick, 1996; Sakas and Fodor, 2012; Sakas et al., in press) but simply point to the variational learning model (Legate and Yang, 2007; Yang, 2002), which provides a formal means of integrating the quantitative properties of the input into the development of child language. Furthermore, before the child converges on the target value of a parameter, a nontarget but nevertheless typologically and biologically possible value is likely to be exercised, providing a principled account of children's acquisition errors. The gradualness of language development suggests a natural role for probabilistic learning mechanisms, which helps situate language acquisition in a broader cognitive and evolutionary setting (Chomsky, 2005; Hauser et al., 2002).

In this article, we explore a different problem that has received less attention in the generative study of formal learnability. The problem concerns the acquisition (and loss) of rules that cannot be plausibly built into the innate endowment for language but must be formed by inductive generalization from language-specific input data. The acquisition of morphology and phonology most clearly illustrates the data-driven nature of these problems: not even the most hardened nativist would suggest that the 'add -*ed*' rule for the English past tense is available innately, waiting to be activated by regular verbs such as *walk–walked*. In the domain of syntax, which is our focus, a prominent problem of inductive learning is found in the English ditransitive dative constructions (double-object and *to*-dative constructions), which has become known as Baker's Paradox after a classic paper by Baker (1979).

(1) a. John gave the ball to Bill.
         John gave Bill the ball.
    b. John assigned the problem to Bill.
         John assigned Bill the problem.
    c. John promised the car to Bill.
         John promised Bill the car.
    d. John donated the picture to the museum.
       * John donated the museum the picture.
    e. * John guaranteed a victory to the fans.
         John guaranteed the fans a victory.

Examples such as (1a)–(1c) seem to suggest that the double-object and the *to*-dative constructions are mutually interchangeable, only to be disconfirmed by the counterexamples of *donate* (1d) and *guarantee* (1e). How do children avoid these traps of false generalizations? The learning problem here is an example of the Subset Problem in language acquisition (Berwick, 1985; Pinker, 1989). Specifically, if learners form an ultimately incorrect generalization – namely, verbs such as those in (1) can participate in the dative constructions – how do they ever learn the negative exceptions in (1d) and (1e) in the absence of negative evidence (Brown and Hanlon, 1970)?

In Section II, we briefly review the basic facts of the English dative constructions and their acquisition. In particular, we suggest that innate, universal, and UG-internal factors are unlikely to provide sufficient constraints to guide the child learner, and a principled theory of generalization from data is necessary. Such a theory, dubbed The Tolerance Principle (Yang, 2005, 2016), is summarized in Section III, along with its primary empirical motivation and highlights from several empirical case studies. Section IV puts The Tolerance Principle into use. We analysed data from a five-million-word corpus of child-directed English collected from the public domain to show how the dative constructions can be productively acquired. More interestingly, The Tolerance Principle provides a mechanism for the learner to retreat from overgeneralizations without resorting to (indirect) negative evidence. How does this theory apply to other language learning situations? Section V considers whether and how The Tolerance Principle could also be at work in L2 acquisition and in L1 attrition. We review representative studies of the L2 acquisition and L1 attrition of the dative constructions and lay out a research program that relates formal theories of grammar and learning to these two bilingual situations.

II   Dative constructions and Baker's Paradox

Of course, Baker's Paradox cannot really be a paradox: All English-learning children eventually acquire the dative constructions. As presented in (1), however, the problem of learning does seem intractable: How do children make up their minds on five verbs (*give*, *assign*, *promise*, *donate*, and *guarantee*) with very similar semantic properties but very different syntactic distributions? We suggest that the learning paradox is largely imagined by linguists and psycholinguists: the inductive indeterminacy in (1) will never arise in a realistic language learning situation, and learners in fact have much stronger evidential basis to make linguistic generalizations, under the guidance of a learning theory such as The Tolerance Principle. In this section, we provide the background for the acquisition study with a brief review of the dative constructions and the challenges they pose for both linguistic theory and language learning.

First, observe that the dative constructions are undoubtedly productive in English, and a strictly lexicalized theory of learning that does not go beyond the adult input is untenable (for a review, see Pinker, 1989). When new verbs such as *text* entered the language, both the double-object and the *to*-dative construction became instantly available: *John texted Bill the news* and *John texted the news to Bill*. Another tell-tale sign for productivity comes from children's naturalistic production data. Just as in the case of past tense, where overregularized errors (e.g. *think–thinked, fall–falled*) point to the productivity of the 'add *-ed*' rule, English-learning children spontaneously use verbs in the dative

constructions that are ungrammatical in adult language and thus would never appear in the input. Indeed, the rates of overregularization in datives (Gropen et al., 1989) and past tense (Marcus et al., 1992) are comparable, at about 5% of all tokens. A few naturally occurring examples from Gropen et al. (1989) and Bowerman and Croft (2008) are given below:

(2)   Jay said me no.
      I said her no.
      Don't say me that. (Asking adult not to tell him to put on his socks)
      So don't please … keep me a favor (asking brother not to throw up on a ride)
      Shall I whisper you something?
      You put me just bread and butter.
      Mattia demonstrated me that yesterday.

Additionally, both children and adults are capable of extending novel verbs in dative constructions (for a summary of earlier studies, see Gropen et al. 1989). For instance, Conwell and Demuth (2007) conducted an experiment with three-year-olds, who were asked to observe an action that involved transferring an object from the child to a recipient via a conveyor belt or a catapult; the action was described as *pilk* or *gorp*. The experimenter modeled the use of the novel verb in one of the two dative constructions – *You pilked the cup to Toby* or *I pilked Petey the cup* – and the children were then asked to perform and describe these actions. Conwell and Demuth found that the children were capable of extending both constructions and were not limited to the usage form they were exposed to (contra Akhtar and Tomasello, 1997).

Second, there is general agreement on what makes dative constructions possible across languages (Beck and Johnson, 2004; Goldberg, 1995; Gropen et al., 1989; Hale and Keyser, 2002; Krifka, 1999; Levin, 1993; Pinker, 1999), which may provide guidance for children to acquire these constructions. For instance, Brown (2010), Hovav and Levin (2008), Levin (2008), and others suggest that the double-object construction requires verbs with the semantics of caused possession, and the *to*-dative construction requires verbs of caused motion. The distinction between the two classes can be observed in the following examples (Gropen et al., 1989):

(3)   a.   I sent a package to the boarder/the border.
      b.   I sent the boarder/*the border a package.

Both *boarder* and *border* may be the destination of the motion caused by *sent*; thus both variants are grammatical for the *to*-dative construction (3a). By contrast, only a *boarder*, not *border*, may have the package as the result of caused possession, hence the grammaticality contrast in (3b).

Note that these semantic properties of dative verbs are only necessary but not sufficient conditions on the constructions, or what Gropen et al. (1989) and Pinker (1989) call 'broad-range rules'. Presumably, all languages have verbs of caused possession and caused motion, but whether a specific language allows the productive use of these constructions must be determined on the basis of experience. In some languages, the double-object construction is not productive but limited to a handful of verbs (Harley, 2002;

Jelinek and Carnie, 2003; Jung and Miyagawa, 2004), and languages such as Chamorro lack the construction altogether (Chung, 1998). Furthermore, it is evident that caused possession/motion are too coarse grained to account for the full range of facts; subgrouping of the verbs is necessary via 'narrow-range rules' (Gropen et al., 1989; Pinker, 1989). Consider the contrast between *throw* and *whisper*:

(4) a. I threw the ball to him.
I threw him the ball
   b. I whispered the secret to him.
   * I whispered him the secret.

According to Gropen et al. (1989), the difference between *throw* and *whisper* is due to their membership in different narrow-range rules/classes. Specifically, *throw* belongs to the class of 'instantaneous causation of ballistic motion', which permits the double-object in English, whereas *whisper* falls in the class of 'manner of speaking', which does not. But even the narrow-range rules have different productive status across languages (see Levin, 2008, and the references therein). The *throw*-class, for instance, is productive for the double-object construction in English, Greek, Hebrew, and Warlpiri, but not productive in Fongbe, Icelandic, Mandarin, and Yaqui. Thus, even if the narrow-range classes were innate – hardly an innocuous assumption because these include classes such as 'verbs of instrument of telecommunication' (e.g. *radio*, *telephone*, and *telegraph*) – inductive learning is still necessary for each specific language.

Finally, there is no escape from the fact that even within a single language, the availability of the dative constructions can be lexically arbitrary. This would seem to defy all attempts to derive structurally sufficient and necessary conditions on the datives on the basis of etymology, phonology, and morphology (for a review, see Harley and Miyagawa, 2016). Levin's (1993) encyclopedic study of verb classes gives us a sense of variability in English alone. For instance, Latinate verbs are generally described as exceptions to the double-object construction (5a) but exceptions to the exception can be found in (5b).

(5) a. Unavailable for double-object construction: address, administer, broadcast, contribute, convey, delegate, deliver, demonstrate, denounce, describe, dictate, dispatch, display, distribute, donate, elucidate, exhibit, explain, explicate, express, forfeit, illustrate, introduce, narrate, portray, proffer, recite, recommend, refer, reimburse, remit, restore, return, sacrifice, submit, surrender, transfer, transport
   b. Available for double-object construction: advance, allocate, allot, assign, catapult, cede, concede, extend, grant, guarantee, issue, promise, refund, relay, render, rent, repay, serve

Indeed, it is not difficult to find very similar verbs that pattern completely differently. All the verbs in (6) have the semantics of caused possession but, unlike donate, all are possible in the double-object construction.

(6) a. They assigned the class two problem sets.   (Latinate like *donate*)
   b. They offered the driver fifty dollars.   (Stress-initial like *donate*)
   c. They promised the citizens clean water.   (Both Latinate and stress-initial like *donate*)

Inventing additional narrow-range labels or postulating even finer structural restrictions, which is always a possibility for theorists, amounts to a restatement of the facts and clearly does not help language learners. It is important, then, to recognize that the broad- and narrow-range rules and similar structural characterization of the datives are descriptions of the terminal state of learning, not an explanation of learners get there.

In sum, the English dative constructions show the typical pattern in the acquisition of rules in coexistence with exceptions: Children overgeneralize before learning lexically arbitrary restrictions, while discovering the productive regularity within certain sub-classes (e.g. *fax*, *text*, *email*, etc.). The learning model cannot entirely rely on universal structural constraints, and it must include a reasonable mechanism by which children retreat from overgeneralizations, without negative evidence.[1]

III   How children generalize

The Tolerance Principle (Yang, 2016) was developed to address an apparently simple problem: How does an English-learning child recognize that the 'add *-ed*' rule is the productive rule applicable to an unlimited number of verbs while the irregular patterns are lexically restricted and do not extend (Xu and Pinker, 1995; Yang, 2002)? It follows a strong intuition shared by almost all researchers of linguistic productivity (Aronoff, 1976): A rule is productive if it applies to most items to which it is applicable. But if a rule has too many exceptions, then the learner will decide against its productivity.

The calibration of productivity can be observed during the course of language acquisition, where the most prominent example is, again, the acquisition of English past tense. Children initially are locked into a protracted stage during which the irregular verbs are inflected correctly (Maratsos, 2000; Marcus et al., 1992; Yang, 2002). Then, and rather abruptly, overregularization errors (e.g. *think–thinked, know–knowed*) start to appear. This marks the emergence of the productive regular rule ('add *-ed*'), which can apply to all verbs, affecting even the irregulars that never appear with '-*ed*' in the input.[2] An important question in the study of English past tense is to determine the tipping point at which the 'add *-ed*' rule becomes productive: the child must learn enough regular verbs to overcome the irregulars. As long as rules have exceptions – almost all do, because all grammars 'leak' – the scope of linguistic productivity remains a central issue in the formal study of language and language acquisition.

The Tolerance Principle, which is summarized in Yang (2016), is a theory that provides for the threshold of exceptions for a productive rule. In the case of English past tense, the regular 'add *-ed*' rule becomes productive only if the child has acquired a sufficiently large number of verbs such that the number of exceptions (i.e. the irregulars) falls below the threshold. The motivation for The Tolerance Principle comes from computational efficiency considerations of language use. It holds that when faced with alternative organizations of linguistic data, the child chooses a more efficient, i.e. faster, grammar. For the acquisition of rules, the choice is between a productive rule with a list of exceptions or a fully lexicalized list that dispenses with the rule altogether.[3] By comparing the computational cost of full listing and the cost of a rule plus exceptions, we can derive a threshold value of exceptions below which postulating a productive rule results

in faster processing time. Under general assumptions about word frequencies (i.e. Zipf's law; Zipf, 1949),[4] it is possible to prove the following:

(7) Tolerance Principle: If R is a productive rule applicable to N candidates, then the following relation holds between N and e, the number of exceptions that could but do not follow R:

$$e \lesssim \theta_N \text{ where } \theta_N = \frac{N}{ln\ N}$$

To apply The Tolerance Principle, one must have a well-defined rule that has a precise condition for its applicability. Such rules may be innately given or, more typically, inductively learned from language-specific data. As in any kind of inductive learning, the hypothesis space that the learner considers must be constrained in some suitable fashion; at the minimum, the learner must know the 'primitives' over which generalization can be formed. For instance, to learn the conditions for the 'add -*ed*' rule, the learner attempts to form structural generalizations over verbs that add '-*ed*' in past tense. Because the verbs that add '-*ed*' are phonologically diverse, an inductive learner will quickly discover that the phonological shape of the word places no restrictions on the application of '-*ed*', or '**If** verb **Then** add -*ed*'; see Yip and Sussman (1997) for a particularly elegant implementation of this widely shared idea in many fields (e.g. Berwick, 1985; Mitchell, 1982; Osherson and Smith, 1981). Now The Tolerance Principle can be applied to the '**If** verb **Then** add -*ed*' rule. Here N is the number of verbs in the child's vocabulary, and e is the number of verbs that do not add '-*ed*', i.e. the number of irregular verbs in the child's vocabulary. The learner compares e against the value of θ_N ($N/ln\ N$). If e is lower than the threshold, then the 'add -*ed*' rule is deemed productive; otherwise, 'add -*ed*' is still lexical, and the child does not generalize beyond the list of verbs in the input that attached '-*ed*' in past tense.

The slow growth of the $N/ln\ N$ function suggests that a productive rule must be supported by an overwhelming number of rule-following items to overcome the exceptions. Some values of N and the corresponding thresholds of tolerance are given in Table 1.

It is clear that, as the proportion of N, the total number of relevant lexical items, the tolerable number of exceptions decreases as N increases. This suggests that productive rules are relatively easier to learn when the learner has a smaller vocabulary, a conclusion that may have significant implications for the difference between child and adult language acquisition; for extended discussion, see Yang, 2016.

Yang (2016) presents numerous empirical case studies of morphology, phonology and syntax to show that The Tolerance Principle makes accurate predictions about where productivity arises in language and when it collapses. Here we summarize recent experiments by Schuler et al. (2016), which produced near-categorical support for The Tolerance Principle using artificial language. Children between the age of 5 and 7 were presented with 9 novel objects with labels. The experimenter produced suffixed 'singular' and 'plural' forms of those nouns as determined by their quantity on a computer screen. In one condition, five of the nouns share a plural suffix, and the other four have individually specific suffixes. In another condition, only three share a suffix and the other six are all individually specific. The nouns that share the suffix can be viewed as the regulars, and the rest of the nouns are the irregulars. The choice of 5/4 and 3/6 was by design:

Table 1. The tolerance threshold for rules of varying sizes.

N	θ_N	Percentage
10	4	40.0
20	7	35.0
50	13	26.0
100	23	23.0
200	38	19.0
500	80	16.0
1,000	145	14.5
5,000	587	11.7

The Tolerance Principle predicts the productive extension of the shared suffix in the 5/4 condition because 4 exceptions fall just below the threshold ($\theta_9 = 4.1$), but no generalization in the 3/6 conditions. In the latter case, despite the statistical dominance of the shared suffix as the most frequent suffix, the six exceptions exceed the threshold. In a wug-like test after training, children were presented with novel items in the singular and were prompted to produce the plural form. Nearly all children in the 5/4 condition generalized the shared suffix on 100% of the test items in a process akin to the productive use of English '-*ed*'. In the 3/6 condition, almost no child showed systematic usage of any suffix, much like native speakers when queried with 'gapped' (ineffable) morphological items (e.g. the past participle of *stride*, and numerous other cases in morphologically complex languages; see Baerman et al., 2010; Halle, 1973; Yang, 2016). The results from adult participants, however, are quite different. In both the 5/4 and 3/6 conditions, adults used the suffixes probabilistically, roughly matching their token frequencies in the training data, similar to the findings in previous research (e.g. Hudson Kam and Newport, 2005, 2009). Currently it is not clear why children and adults approach the same learning materials in such contrasting ways, but the results of Hudson Kam and Newport's studies may shed light on the precise mechanisms between L1 and L2 acquisition.

The Tolerance Principle is a parameter-free model in the sense that it is not tuned to fit any linguistic data or experimental results. It makes a categorical prediction about productivity on two input variables (*N* and *e*). We now give a simple example to show how The Tolerance Principle accounts for the emergence of the 'add -*ed*' rule in English past tense; the treatment of similar cases of phonological, morphological, and syntactic acquisition follows the same blueprint.

Suppose an English-learning child knows $e = 120$ irregular verbs, roughly those found in a five-million-word corpus of child-directed English in the public domain (CHILDES; MacWhinney, 2000), which is approximately a year's input data that many English-learning children receive (Hart and Risley, 2003). We can infer the minimum value of *N*, the number of all verbs (regular plus irregular), such that $\theta_N = N/ln\ N \geqslant 120$. A simple calculation shows that the value is about 800. That is, if children have acquired 120 irregular verbs, they will need at least 680 regular verbs to establish the productivity of the 'add -*ed*' rule. This condition is easily met: the five-million-word corpus of child-directed English contains over 900 regular verbs inflected in past tense. Thus, an English

learner will be justified to conclude that the 'add -*ed*' rule is productive and can be extended to novel items (e.g. *rick–ricked*; Berko, 1958).

In fact, the prediction can be made more precisely, potentially at the individual level. After all, productivity is determined by the numerical values of N and e, which are determined by an individual learner's vocabulary of regulars and irregulars. Consider 'Adam', the poster child for English past-tense acquisition (Pinker, 1999). Adam produced the first instance of overregularization error – *What dat feeled like*? – at the age of 2;11. In the transcript of almost a year prior to that point, not a single irregular past tense is incorrect. If we take overregularization as the moment when 'add -*ed*' becomes productive (Marcus et al., 1992), then it must be the case that, by this point, Adam has acquired a sufficiently large number of regulars to overwhelm the irregulars. To test this prediction, we extracted every verb stem in Adam's transcripts until 2;11 (MacWhinney, 2000). There are $N = 300$ verbs in all, of which $e = 57$ are irregulars. This is very close to the predicted $\theta_{300} = 53$, and the small discrepancy may be due to the undersampling of the regular verbs, which tend to be lower in frequency and are more likely to be missed in a finite corpus. The critical point to note here is that Adam needed a majority of regular verbs to acquire the 'add -*ed*' rule: this is strongly consistent with the predictions of The Tolerance Principle.

We now formulate a corollary of The Tolerance Principle that provide a straightforward account of how the English dative constructions are learned by children.

IV Forming dative generalizations

Before turning to how linguistic generalizations are formed, let us consider a hypothetical example, the kind that is likely to arise in any instance of inductive learning problem. Imagine being shipwrecked on a desert island. If you come across 10 exotic species, seven of which are clearly tame and friendly, you'll probably assume the next encounter is harmless: seven of 10 seem pretty good odds. But one of 10? Two out of 10? You'd be well advised to proceed with caution.

Generalizations evidently require the weight of evidence: seven exemplars appear sufficient but one or two probably will not do. This is not to say that the sufficiency of evidence is foolproof: the other three species, with which we have had no direct experience, may well turn out to be dangerous. Nor does sufficiency guarantee permanence: if an additional 20 species have come our way, suddenly seven out of 30 no longer inspires confidence. In the spirit of The Tolerance Principle, we propose the following corollary:

(8) Sufficiency Principle: Let R be a generalization over N candidates, of which M items are attested to follow R. R can be extended to all N items if and only if:

$$N - M \leq \theta_N \text{ Where } \theta_N = \frac{N}{ln \ N}$$

Before the positive evidence is sufficient, that is, M sits below the sufficiency threshold, the learner lexicalizes all M items and does not generalize beyond them. Without any kind of generalization, the problem of overgeneralization does not even arise. That is, Baker's Paradox involving the verb *donate*, as in (1), is a non-issue unless the child has

observed a significant majority of similar verbs attested in the double-object and *to*-dative construction, which, as we will see, probably will not happen during the first few years of language acquisition. Only when *M* crosses the sufficiency threshold relative to *N*, again, an overwhelming majority, does R become a truly productive rule.

The Sufficiency Principle has a built-in mechanism for retreating from overgeneralization. Suppose $N = 50$ and the learner has accumulated $M = 40$ instances to warrant a generalization ($50 - 40 = 10 < \theta_{50} = 12$). But further down the road, N starts to increase again. Suppose that N has grown to 70 with 20 more items, but *M* has been standing still at 40: now the rule will cease to be productive ($70 - 40 = 30 > \theta_{70} = 16$), and the learner will lexicalize all 40 items and the once productive generalization will be abandoned. The dynamics of learning under the Sufficiency Principle is exactly the same as under The Tolerance Principle as illustrated in the acquisition of English past tense, where lexicalized learning abruptly triggers categorical changes in the grammar.

The Sufficiency Principle provides a straightforward solution for the acquisition of datives; for details, see Yang, 2016. For the ease of presentation, we will only consider the acquisition of the double-object construction: children's initial overgeneralization (*she said me no*) and then subsequent retreat from it. To establish an estimate of the English learner's input experience, we examined the distribution of verbs that participate in the double-object construction ('verb NP NP'). Ideally, we would like to use the input data to specific learners, but no such dataset is currently available. Thus, we constructed a five-million-word corpus of child-directed North American English from the CHILDES database, by collapsing the transcripts for all children. The size of this corpus is comparable to roughly a year's linguistic input for working-class children; this is hardly perfect, but close to the best one can manage in the public domain.

A total of 42 verbs were used by adults in the double-object construction. Of these, 38 have a very clearly identifiable semantics of 'caused possession', which we assume is identifiable if the learner is equipped with a suitable set of semantic primitives (Grimshaw, 1990; Jackendoff, 1990; Pinker, 1989). The four exceptions, well below the threshold $\theta_{42} = 11$, do not convey caused possession but are all performative verbs (*call, consider*, *name*, and *pronounce*), which plausibly leads to a subclass with associated productivity. Thus, the semantic condition of caused possession, now widely recognized by many researchers (Gropen et al., 1989; Krifka, 1999; Levin, 1993; Pesetsky, 1995; Pinker, 1989) as a prerequisite for the double-object construction, need not be stated as a UG primitive but can be acquired from the language-specific data under a principled learning model.

(9)   In English, if a verb appears in the double-object construction, then it will have the semantics of caused possession, a tolerably low number of exceptions notwithstanding.

At this point, the child may consider the converse of (9), trying to establish the validity of caused possession as a *sufficient* condition for the double-object condition. Again poring over the five million words of child-directed English, we established the set of caused possession verbs to see if the subset of $M = 38$, which are actually used in the double-object construction, constitutes sufficient evidence for generalization to the entire set. There are an additional 11 verbs in the corpus that have the semantics of

caused possession but these verbs are not used in the double-object construction in the five-million-word corpus:

(10)   address, deliver, describe, explain, introduce, return, transport, ship, mention, report, say

This is an interesting list. For some of the verbs in (10), e.g. *introduce* and *say*, the double-object construction is ungrammatical:

(11)* John introduced the kids a new dish.
    * John said Bill something mean.

Whereas others do allow the double-object construction but did not have the opportunity to do so in the child-directed corpus:

(12)   John shipped Bill his purchase.

But of course children do not know *why* verbs in (10) fail to appear in the double-object construction (ungrammaticality or lack of opportunity) in the speech addressed to them. Nevertheless, a sufficiently large number of verbs in the child-directed corpus, namely $M = 38$, is able to trigger the following generalization for the entire set of $N = 49$ verbs: $11 < \theta_{49} = 12$ satisfies the Sufficiency Principle.

(13)   If a verb has the semantics of caused possession, then it can appear in the double-object construction.

This immediately accounts for the overgeneralization errors in (2) such as *She said me no*, as well as the experimental evidence that children as young as 3;0 have productive usage of the dative constructions upon learning a novel verb with the appropriate semantic properties (Conwell and Demuth, 2007).

But how do children retreat from the (over)generalization in (13)? After an exhaustive search of the child-directed English data in the public domain, we do not believe that the problem in Baker's Paradox is likely to arise for a young child. As noted by Gropen et al. (1989), verbs such as *donate* are far too rare to be learned in the early years of language acquisition. Indeed, Latinate vocabulary, which includes *donate* and other verbs at the heart of the Subset Problem, is generally acquired in a school setting (Jarmulowicz, 2002; Tyler and Nagy, 1989). (The situation with the Latinate verbs may be quite different in L2 acquisition, as we discuss in Section V.)

According to Levin's encyclopedic survey (Levin, 1993: pp. 45-48), there are in fact more caused possession verbs (138) that resist the double-object construction that those that allow it (115): 115 out of 253 certainly does not warrant any productive generalization. However, many of these verbs are rare and probably will not enter into the calibration of productivity for most English learners. By making suitable frequency estimates (for details, see Yang, 2016: Section 6.3), one can 'trim' the verbs to a relatively common set of 92. Still, only 52 can be expected to appear in the double-object construction. But this still falls well short of the threshold imposed by the Sufficiency Principle: a

productive generalization for 92 items requires $92 - \theta_{92} = 72$ positive instances. The generalization in (13) is no longer productive, and the learner will now need to lexicalize every verb that appears in the double-object construction: errors such as those in (2) will be eliminated as the learner, after acquiring a sufficiently large vocabulary, successfully retreats from an overly inclusive hypothesis.

There are additional moves available to the learner. For instance, they may subdivide the 92 caused possession verbs into finer classes such as 'ballistic motion', 'manner of speaking', etc. (Gropen et al., 1989; Levin, 1993; Pesetsky, 1995; Pinker, 1989), assuming that the semantic attributes associated with such categorization, which may include both linguistic and non-linguistic factors, are accessible to the learner. The Sufficiency Principle can be applied recursively, detecting productive generalizations in some of these subclasses. One such class may be verbs of 'telecommunication': *fax*, *phone*, *telegram*, etc., leading to the immediate availability of the construction when verbs such as *email* and *text* entered the English lexicon. Certain phonological and morphological constraints of verbs (Green, 1974; Oehrle, 1976) can be successfully identified; for details, see Yang, 2016: Section 6.3. Let us now consider how The Tolerance Principle may apply in two related aspects of bilingualism: L2 acquisition and L1 attrition.

V   The Tolerance Principle in L2 acquisition and L1 attrition

L2 acquisition is the acquisition of another language after the structural foundations of the first language are in place, and L1 attrition is the other side of the L2 acquisition coin: the loss of aspects of a previously acquired grammar when input and use of the language decreases (as exposure to and use of the L2 increases; Schmidt, 2013). Second language acquisition and language attrition research must explain how generalizations and overgeneralizations are learned and unlearned in the progression and regression of grammar, and how language development may be sensitive and resilient against variation in input and experience. Input and experience are intrinsically linked. The input a bilingual person is exposed to is divided between two languages, and bilinguals use the two languages for different purposes and in different domains of life (Grosjean, 2016).

The L2 learner may be exposed to less input than an L1 learners but faces the same inductive problem as the L1 acquiring child. Like in L1 acquisition, L2 learners must build a grammatical representation of the L2 based on input, and they must also determine which rules are productive and what data are exceptions. If L2 learners also build grammatical representations guided by statistical patterns in the input, The Tolerance Principle could apply in this situation as well. However, there are two other important differences between L1 and L2 acquisition, which may work together or even perhaps mask the operation of The Tolerance Principle at a given stage of development. One difference is that the L2 learner (child or adult) is more cognitively and linguistically mature than the child L1 learner when acquisition starts. The other difference is previous linguistic knowledge. When parsing L2 input, older learners may engage higher order cognitive schemas and domain general mechanisms not yet available to L1 children (Bley-Vroman, 1989), leading to different types of analysis of the data, different hypotheses, and different developmental trajectories. The linguistic representation of the first language is

another source of information at the disposal of L2 learners, especially at lower levels of development, when the L1 is most likely to be a default hypothesis space for parsing the input and assigning a linguistic representation to the L2 (Schwartz and Sprouse, 1996). With sufficient and sustained exposure to the target language and with increased proficiency, many L2 learners are able to eventually overcome L1 influence and converge on the grammar of the L2, at least in some specific domains.

Many children and adults immersed in an L2 environment undergo language attrition. While the evolution of grammar in L1 acquisition leads to the final state of an adult native speaker; language attrition in the context of bilingualism leads to language regression and loss of native speaker status as the L2 grows stronger and the bilinguals shift dominance. L1 attrition affects phonology, lexical knowledge, retrieval and access, and leads to changes ('errors', interference from the other language) in morphosyntax, semantics and syntax–discourse interfaces (Montrul, 2008; Schmidt, 2013). As in L2 acquisition, where there is transfer from the L1 to the L2, in L1 attrition the L2 also becomes an additional source of knowledge for the native speaker. Influence of the dominant L2 onto the L1 is also common in many domains (Montrul and Ionin, 2012). And as in L2 acquisition, there are age effects: the degree and extent of language attrition depends on the age of onset of bilingualism, as many studies have shown that children are more likely to exhibit attrition than adults (Montrul, 2008). In fact, attrition in pre-puberty children can be fast and furious, affecting the integrity of the developing grammar (Flores, 2010). Thus, in both L2 acquisition and L1 attrition factors like age, linguistic environment and knowledge of another language shape the nature of language acquisition and its regression. The question we pose is whether The Tolerance Principle also applies in these situations, and our answer is that it should, albeit in concert with other factors. We review existing work in L2 acquisition and L1 attrition of the dative alternation in English, identifying trends that are suggestive that The Tolerance Principle may be at work in bilingualism as well. We end by proposing a research program that would allow us to test more directly the predictions of The Tolerance Principle in L2 acquisition and L1 attrition.

VI   How datives come and go

The dative alternation has been a poster child of generative approaches to second language acquisition and widely recognized as an instance of a poverty-of-stimulus phenomenon, since the L2 learner must somehow determine which verbs allow alternating syntactic forms and which ones do not from a limited set of data in the input (Juffs, 1996; Perpiñán and Montrul, 2006; White, 2003). There are several studies of the L2 acquisition of the dative alternation in English by L1 speakers of Spanish (Agirre, 2015), Brazilian Portuguese (Zara et al., 2013), French (Hawkins, 1987; Le Compagnon, 1984; Mazurkewich, 1984), Japanese (Bley-Vroman and Yoshinaga, 1992), Chinese (Chang, 2004), Japanese and Chinese (Inagaki, 1997), Korean and Mandarin (Oh, 2006, 2010; Oh and Zubizarreta, 2006), Korean and Japanese (Whong-Barr and Schwartz, 2002), Russian (De Cuypere et al., 2009–10), Turkish (Marefat, 2005), and German (Jäschke and Plag, 2016; Woods, 2015). These studies, most of them experimental, have addressed the role of L1 transfer on the acquisition of the dative alternation English, the

developmental stages L2 learners go through in acquiring the dative alternation, and how L2 learners acquire different subclasses of verbs based on finer-grained semantic and morphological constraints (e.g. goal vs. benefactive verbs, Latinate vs. Anglo-Saxon verbs). We acknowledge that because these studies use different elicitation methods (grammaticality judgments with and without corrections, acceptability judgments and reaction times, elicited production tasks with real and novel verbs) and different populations of L2 learners in terms of age (children and adults) of diverse linguistic backgrounds, these studies are obviously not strictly comparable. Yet, the patterns and trends reported are suggestive of how The Tolerance Principle may also be at work in L2 acquisition. There is evidence, for example, that L2 learners, like L1 learners, also move from a very general productive interlanguage grammar to a more restrictive and lexically constrained grammar.

One common finding in the L2 literature is that the *to*-dative is acquired before the double-object dative and overgeneralized, which is the opposite of what is found in the child L1 acquisition, where young children are more likely to generalize from double objects to *to*-datives than the other way around (Conwell and Demuth, 2007). More overgeneralization of *to*-dative to double objects than of double objects to *to*-dative in L2 acquisition have been reported at initial, intermediate and even advanced stages of development. This may occur because in some studies the L1 of the learners has *to*-datives, and only some languages have double objects. For example, Mazurkewich (1984), Le Compagnon (1984) and Hawkins (1987) examined the acquisition of the dative alternation by French-speaking learners with a focus on the Latinate constraint. French is assumed not to have the double-object construction, and the three studies found that the learners acquired the *to*-dative construction earlier than the double-object construction. Agirre (2015) investigated the L2 acquisition of the double-object and *to*-dative constructions by Spanish-speaking L2 learners of English of beginner, intermediate and advanced proficiency, to see if the L2 learners showed differences in their judgments of *to*-datives and double-object constructions with goal and benefactive verbs, on the one hand, and whether they were sensitive to the morphological (Latinate) and semantic (possessor) constraints, on the other. Syntactically, Spanish appears to have double-object construction as in English.[5] However, these are not subject to the same morphological and semantic constraints as in English. In fact, double objects seem to have a broader distribution in Spanish than in English. Agirre found that beginner level learners assigned similar ratings to grammatical sentences with double objects and *to*-dative verbs in English, while intermediate and advanced learners rated grammatical *to*-datives more acceptable than grammatical double objects (as found by Hawkins, 1987; Mazurkewich, 1984). Because the beginner learners were more accepting of double objects than the intermediate and advanced groups, Agirre interpreted the results of the beginner learners as firmly supporting the Full Transfer/Full Access Hypothesis (Schwartz and Sprouse, 1996). Since double objects and prepositional datives are available in Spanish, the learners transfer the two structures onto English from the beginning. The intermediate and advanced learners' higher acceptability of prepositional datives than of double objects suggests a stage of overgeneralization of *to*-datives to double-object datives. Higher use, acceptability or preference of the *to*-dative over the double-object datives in the studies with French speakers may be because the L1 also facilitates

the acquisition of *to*-datives in English. Since Spanish is assumed to have the equivalent of double objects and *to*-datives, L2 learners do not show a pattern of preference of the *to*-dative over the double-object dative until later (intermediate level), when they may have accumulated more exposure to datives in the input.

There are other studies finding a preference or higher use of *to*-datives over double-object datives in L2 English that cannot be accounted for by L1 influence, however. Chang (2004) found this pattern with Chinese-speaking learners and Führer (2009), De Cuypere et al. (2009–10) with Russian-speaking learners, and Jäschke and Plag (2016) with German-speaking advanced learners. Chinese, Russian and German are assumed to have double-object constructions similar to English (Li and Thompson, 1981; Pechmann et al. 1994). Jäschke and Plag (2016) extended Bresnan and Ford's (2010) methodology with native speakers of American and Australian English to test advanced learners of English. Their goal was to see whether German-speaking L2 learners would prefer the double-object dative over the *to*-dative construction in English, and which contextual and syntactic factors regulated their preferences. Sentences with *to*-datives and double-object datives were presented with a context paragraph, and participants had to decide which sentence was a better continuation of the paragraph. At the groups level, Jäschke and Plag found that the learners were sensitive to many of the same syntactic and discourse factors the American and Australian native speakers reported in Bresnan and Ford (2010) were sensitive to for choosing a given verb in the *to*-dative or in the double-object construction but, interestingly, there was a slight preference for *to*-datives over double-object datives. The individual results showed this pattern to be true of 13 learners; the other 11 showed a preference for the double-object dative over the *to*-datives.

While the role of the L1 has been extensively discussed and cannot be ruled out in some of these studies, patterns due to L1 transfer co-exist with overgeneralization errors that are clearly not related to the L1. None of these studies devotes much space to discuss the role of input in any significant depth, which is what The Tolerance Principle is based on. In general, overgeneralization errors tend to appear during intermediate stages of development in L2 acquisition, and following The Tolerance Principle, these errors may emerge from learners performing an inductive distributional analysis of the input, which keeps track of the number of instances of verbs found in one construction or the other. Not only is the input that L2 learners receive less abundant in quantity than the input an L1 learner receives (in terms of hours per day), but it also differs in quality. L2 learners, who are older than L1 learners, are exposed to different input than child L1 learners, especially if they acquire the language in an instructed setting and are introduced to reading and writing earlier than an L1-acquiring child.

An explanation consistent with our proposal is that, in addition to the L1 having *to*-datives, the input to L2 learners may also contain more verbs in the *to*-dative construction than in the double-object construction, when compared to the conversational input most common to the child L1 learner. Coming back to Agirre's study, it is also possible that, based on the limited input the L2 learners in this study may have been exposed to, they have encountered *to*-dative and double-object verbs in the English input with relatively similar frequency and they have activated both rules. But if the learners are exposed to more literate input, it is possible that they encounter more *to*-dative than double-object verbs, and that is why they may be more accurate recognizing *to*-dative

than double-object constructions in English. Interestingly, Agirre found that intermediate learners, but not beginners, were more accepting of ungrammatical sentences with Latinate verbs in double objects which, in our assessment, may have encountered more double objects than *to*-datives in the input. Agirre, unfortunately, does not discuss how the type of input may relate to the developmental patterns found in her study.

Jäschke and Plag (2016), by contrast, do take into account contextual factors, and suggest that learners only gradually acquire sensitivity toward the linguistic constraints that govern the choice between the two dative constructions. Their study only tested advanced L2 learners. Missing from this study is more discussion of how this sensitivity to distributional constraints might be developed, or on the basis of what. A cross-sectional study including lower level proficiency learners, or a longitudinal analysis of the input their participants may have been exposed to may shed more light on whether some advanced learners encounter higher instances of double-object datives than of *to*-datives while others encounter more *to*-datives than double-object datives, which may explain the preference of one type of dative over the other by their participants.

Beyond the emergence and overgeneralization patterns of double datives and *to*-datives, there is evidence from early studies (e.g. Ard and Gass, 1987) that L2 learners move from a more productive grammar to a more restrictive, lexically-constrained grammar. This would be consistent with learners being sensitive to input distributions, and The Tolerance Principle being somehow active in L2 acquisition as well, although always also masked by potential L1 effects. Tanaka (1987), who tested Japanese learners of English, observed differences by verbs: *to*-datives and double-object constructions were used equally and judged similarly with the verb *give*, but the acceptance of *to*-datives in translation judgments and grammaticality judgment tasks by higher proficiency learners decreased with other verbs. Hawkins (1987), a study with French-speaking learners of English, found that the double-object construction was acquired first with goal (*to*-datives) and later with benefactive (*for*-datives). The final stage was the acquisition of language specific constraints. For example, lower proficiency learners overgeneralized double objects to Latinate verbs more than higher proficiency learners. Oh (2006) and Oh and Zubizarreta (2006) tested Korean and Japanese learners of English and also found higher accuracy and acceptability of double-object sentences with goals than with benefactives. (Japanese does not have double objects, and Korean has some structures comparable to goal and benefactive datives). By contrast, Agirre (2015) did not find an asymmetry for goal and benefactive double objects by Spanish-speaking learners, which she also attributes to the fact that in Spanish the double-object construction has a broader semantic distribution than in English.

Eventually, L2 learners seem to also develop sensitivity to 'narrow-range' rules in the sense of Pinker (1989), even when L2 learners do not yet exhibit native-like behavior. For instance, Bley-Vroman and Yoshinaga (1992) showed that advanced Japanese-speaking learners of English distinguished narrow-range rules for the dative alternation only when real verbs were used (not when made-up verbs were used). In a follow-up study to Bley-Vroman and Yoshinaga (1992) that tested Japanese and Chinese learners of English, Inagaki (1997) investigated knowledge of the double-object construction with *throw*, *push*, *tell* and *whisper*-type verbs. *Throw*-type verbs and *push*-type verbs are syntactically distinct in Japanese, while in Chinese these verb types are not allowed in

double objects. Inagaki further found that the advanced Japanese learners distinguished between *tell* and *whisper*-type verbs in English double-object constructions, even though these verbs do not differ in their ability to appear in dative constructions in Japanese, but were unable to distinguish between *throw*-type and *push*-type verbs when these verb types are not syntactically distinct in Japanese. The results of Bley-Vroman and Yoshinaga and Inagaki could indicate the effect of the frequency of particular verbs appearing in double-object construction in later stages of L2 dative acquisition. At the same time, Inagaki's (1997) findings support the role of L1 influence for the results of the Chinese learners but the role of frequency for the results of the Japanese learners. Together, the L2 findings from Bley-Vroman and Yoshinaga and Inagaki seem consistent with the predictions of The Tolerance Principle, although it is still an open question how far L2 learners can go in delineating the narrow-range verb classes and to what extent they can acquire native-like competence in this domain.

Without a thorough analysis of the input, this question may be hard to answer. Assuming different verb classes are represented in the input and L2 learners cannot delineate narrow range subsets, it is possible that L2 learners do not acquire productive rules based on narrow-range classes and instead rely on the 'lexicalization' strategy that native speakers would also use. Relying more on lexicalization than on productive and proceduralized rules is taxing and costly computationally and psycholinguistically, therefore giving credence to the hypothesis that L1 and L2 acquisition are fundamentally different (Bley-Vroman, 1989). This, however, remains to be tested empirically, with psycholinguistic experiments and techniques that can actually measure the cost of productivity vs. lexicalization in L2 acquisition.

Compared to L2 acquisition, research on the potential loss or unlearning of rules in L1 attrition is much more limited, but the results from an existing study allows us to consider how The Tolerance Principle may apply in situations of language loss as well, when triggered by changes in the input. Seliger (1991) reports on a case study of an American English-speaking child who immigrated with his family to Israel when she was 6 years old. Upon arrival in Israel, the family decided to use Hebrew as the home language and the child began schooling in Hebrew. Therefore, use and exposure to English for this child declined substantially after migration. The child was studied longitudinally for a year when she turned 9 years old (after three years of living in Israel), during which speech samples of her English were elicited every three to four weeks through stories and spontaneous interactions. When the child was 10;6 years old, she was administered a grammaticality judgment task in English.

At age 9, when the naturalistic data collection started, the child made overgeneralization errors with the dative alternation in English, producing sentences like *So she telled it to another girl* and *The school gives the girl that she has the birthday a present*. Hebrew has *to*-datives and no double-object constructions. Sentences with direct and indirect objects have variable order (*Dan natan sefer le-Sarah* 'Dan gave a book to Sarah' and *Dan natan le-Sarah* 'Dan gave to Sarah a book'). The dative preposition *le/la* can mean *to* or *for*. Hebrew has resumptive pronouns in relative clauses. Seliger interpreted some of the patterns in the child's English as transfer from Hebrew (L2 influence on the L1).

To see whether the child had not fully acquired or had lost the constraints on the dative alternation, Seliger designed a grammaticality judgment task with sentences with

to-dative verbs with different word orders and sentences with verbs in the double-object construction (*Dick handed Sally the book. *Ariella answered the teacher the question*). Eight child native speakers of English (four ages 5–7 years and four ages 8–9 years) completed the task and acted as comparison group. The children ages 4–7 years represented the baseline grammatical performance age before onset of bilingualism in the child participant of the case study, and the children ages 8–9 years would be more comparable to the age of the child participant of the case study at the time of testing. The results showed that the child participant of the case study accepted 4 of 5 ungrammatical sentences with the dative alternation (**Ariella answered the teacher the question*), while the native speakers accepted 1 of 5 and rejected 2 of 6 grammatical sentences (*She read the baby the story*). There were no differences between the child native speakers of English based on age (younger children did not accept more errors than older children). The assumption here is that the child of the case study had full knowledge of the dative alternation in English at age 6 years, before moving to Israel, like the 4–7-year-old children who acted as baseline. Since the *to*-dative is the only option in Hebrew while it is one of two options in English, when input in English decreased dramatically the child seemed to have adopted the grammar of Hebrew for English as well.

The data for this case study was collected with oral narratives, and unfortunately there is no mention of the adult speech in English directed to the child. But The Tolerance Principle could easily be applied to L1 attrition as well. Reduced exposure to English as a heritage language in a bilingual environment decreases the exposure to exemplars and the opportunities to use verbs in dative constructions, including the double-object construction. If the English input to the child contains few instances of double objects, and the L2 grammar of the child generates the prepositional *to*-dative construction, the forgetting or 'unlearning' of the double-object construction results from not encountering a minimum threshold of examples to retain the double-object rule.

Naturally, one study is not sufficient to draw any strong conclusions, but there is certainly opportunity to investigate in more detail the potential attrition of the dative alternation in English as a heritage language, in the context of immigration. There is a need for more studies of children and adults undergoing attrition using naturalistic data, and experiments suitable to the age and linguistic abilities of the bilinguals are also needed. By testing children and adults in different countries where English is in contact with different languages and by also collecting robust measures or estimates of input, such studies will allow us to verify the operation of The Tolerance Principle more directly, as well as investigate the role of age of onset of bilingualism, L2 dominance, and quantity and quality of input to trace the progression and regression of productive rules and exceptions. In Section VII we suggest how.

VII Toward a closer look at input in bilingualism

Human languages present rules and exceptions, and we have argued that the child learner must be equipped with some learning mechanism that detects generalizations in the presence of exception. This calls for a theory that marries aspects of nativism with aspects of input-oriented approaches. A concrete model that operates on numerical basis – such as The Tolerance Principle – offers testable predictions on the nature of language across

individuals and populations. It may also be interesting to investigate the possibility that the mechanisms of language acquisition are retained through the adulthood (Slabakova, 2008), which may elucidate the continuity and differences between first and second language acquisition and attrition.

L1 learners, L2 learners and L1 attriters differ in their cognitive maturity and initial linguistic state, but another important difference is the amount and type of input they are exposed to, which is a function of age and experience. The findings in L2 acquisition on the dative alternation appear to be compatible with the general direction of The Tolerance Principle. More importantly, they invite further quantitative research on the nature of the input evidence: how it guides adult learners to form syntactic generalizations, and to what extent this process differs from the process native child learners go through. Many studies suggest influence of the L1 in the acquisition of double-object construction in English, but in order to address the predictions of The Tolerance Principle more directly, ideally, studies should analyse the actual input second language learners receive to consider more directly whether and how the $e < N/ln\ N$ rule may apply. The study of language attrition calls for similar studies.

As explained in Section III, 3-year-old children can productively extend the double-object and *to*-dative construction on novel verbs, as shown by experimental evidence. The 3-year-old children studied by Conwell and Demuth (2007) generalized novel verbs in both constructions, but Conwell and Demuth also noticed a strong preference to extend a verb from double objects to *to*-datives more than from to-datives to double objects (48% vs. 8.9%), despite the fact that *to*-datives are less frequent than double objects in the input (Campbell and Tomasello, 2001; Snyder and Stromswold, 1997). Gropen et al. (1989) found that older children (average age 7;4) were also sensitive to the morphophonological constraints on the dative verbs: polysyllabic nonce verbs were generalized less frequently than monosyllabic nonce verbs (for similar findings, see also Mazurkewich, 1984; for how such constraints may be acquired, see Yang, 2016: Chapter 6). Latinate verbs, which lie at the heart of Baker's Paradox, are almost surely acquired much later when children develop literacy, resulting in retreat from overgeneralization.

Adult L2 learners, by contrast, are generally already literate, and are likely to be exposed to Latinate verbs in the input earlier in the L2 acquisition process than L1 learners of English. And if L2 acquisition starts in elementary school, as in many children who learn English around the world, they are also likely to be exposed to more Latinate verbs early on than 3-year-olds acquiring English as L1. Suppose a 6-year-old child receives English as a foreign language instruction at school two or three times a week for an hour. Assume that the child is not exposed to English outside school. When the child is 9 years old, after three years of instruction, she may encounter the Latinate verb *donate* and other similar ones in the input at school. In that case, the L2 acquiring child will encounter Latinate verbs earlier on the acquisition process than an L1 acquiring child (after 3 years as opposed to after 6 or 7 years in L1 acquisition according to Tyler and Nagy, 1989). But the question is, of course, whether or not a threshold level of Latinate verbs compared to a threshold level of non-Latinate (Anglo-Saxon) verbs is encountered relatively sooner in L2 than in L1 acquisition. This could perhaps be established by examining the teaching materials that an L2 acquiring child may be exposed to longitudinally.

In the case of language attrition, not only would we also obtain output from the participants through naturalistic production or their performance in experimental tasks, but we would also document the changes in the input, the amount of input in each language and obtain samples of what the actual input consists of (samples of directed speech, reading materials, etc.). Analyses of the input would allow us to determine whether the input and input frequencies of the relevant verbs in L1 and L2 acquisition and L1 attrition driving inductive learning may already be very different. We can also document how input changes over time, and how it may relate to the bilingual's development of the two languages. The question is how to do it.

We demonstrated in Section IV how The Tolerance Principle and the Sufficiency Principle can be tested and falsified with corpora containing samples of child and adult speech or with experimental evidence. Quantifying corpora representative of the input for L2 acquisition presents serious challenges because different learners may be exposed to different types of input depending on age and context of acquisition (i.e. naturalistic vs. instructed). There are available corpora of learners' productions (output) but not of the input addressed to the learners. Unlike the field of L1 acquisition where large scale corpora like the CHILDES database has been standard practice for years, large corpora have not been mainstream in studies of L2 acquisition until recently (Granger et al., 2002; Lozano, 2015; Myles, 2005). Most of the available L2 corpora consist of written language (i.e. compositions, essays, answers to exams), and elicited speech (narratives) of learners of different language backgrounds and proficiency levels. (See, for example, the list of learner corpora around the world compiled by the Center for English Corpus Linguistics at the Université Catholique de Louvain in Belgium at www.uclouvain.be/en-cecl-lcworld.html). To date, L2 learner corpus research has been utilized to describe what is frequent and typical in learner language compared to native speaker language. For example, Tono (2004) analysed learner corpora to document how Japanese learners of English acquire the subcategorization patterns of verbs in English when compared to native speakers. And Callies and Szczesniak (2008), who conducted a corpus-based study of the English dative alternation, found that German and Polish learners of English knew the difference between verbs that participated in the double-object datives and those that appeared in the *to*-dative construction.

Formal approaches to second language acquisition have not made use of corpora to test and verify hypotheses. This may be due to the fact that there is a dearth of longitudinal and cross-sectional samples of speakers of similar linguistic backgrounds and ages and of oral data: the available learner corpora is not the most suitable to answer theoretical questions because they are heavily biased towards written samples from advanced learners (Myles, 2005). Furthermore, to investigate the input to learners, we would also need corpora of classroom discourse, where teacher talk could be documented along with learner talk. Studies of classroom input were in vogue in the 1980s and addressed whether simplified input (teacher talk, foreigner talk) was beneficial to second language learners (Håkansson, 1987; Lightbown, 1985; Long, 1983). At present, the field of classroom discourse analysis, where exchanges between teachers and learners are analysed, engages in the collection of that type of data (Seedhouse, 2012), and perhaps these data could be made available to pursue the type of hypothesis-testing studies that we need.

Another way to examine the actual input to L2 learners in the classroom setting is to examine the written teaching materials and compare it to the output L2 learners produce in written and oral discourse. There are some studies of textbook language (Catalán and Francisco, 2008), and especially vocabulary frequency, but none in reference to how dative verbs may be used in English language textbooks, and how these relate to what learners acquire (but see Yang et al., 2000).

Alternatively, experiments with training and novel verbs could be designed to test the predictions of The Tolerance Principle more directly, as in Schuler et al. (2016), especially in light of the contrasting results produced by the L1 children and L2 children and adults. By creating different input conditions, it may be possible to test more directly the dynamics of learning under The Tolerance Principle and see whether and how the Sufficiency Principle may operate in L2 acquisition as well.

In conclusion, investigating the theoretical predictions of The Tolerance Principle in a bilingual context requires that we collect and analyse actual samples of input and see how it relates to the learners' learning and unlearning of rules. We need more studies of learners corpora and more focused experimental studies. This will allow us to test more directly how language development may be sensitive and resilient against variation in input and experience.

Declaration of conflicting interest

Funding

Notes

1. On indirect negative evidence, see Pinker (1989) for a traditional analysis and Yang (2016) in a probabilistic learning framework with critical focus on Bayesian learning methods.
2. For the irregular verbs, the productive 'add -*ed*' rule will be in competition of the irregular past-tense forms, resulting in probabilistic variation (e.g. *holded–held*). We refer the reader to Marcus et al. (1992), Maratsos (2000), and Yang (2002) for detailed analyses of this phenomenon, while noting that, for the present study, the key question is the condition under which the 'add -*ed*' rule becomes productive, rather than its effect on children's irregular morphology once its productivity is established.
3. This approach entails that productivity is a categorical phenomenon, a fact that is unambiguously supported in child language though insufficiently recognized by linguistic researchers (for crosslinguistic review, see Lignos and Yang, 2016; Yang, 2016: Chapter 2).
4. Zipf's law states that the rank-frequency distribution of words in a corpus of natural language is inversely proportional to its rank in frequency, such that the most frequent word will occur approximately twice as often as the second most frequent word, three times as often as the third most frequent word, etc. For example, in the Brown Corpus of American English text (Kučera and Francis, 1967), the word 'the' is the most frequently occurring word, and by itself accounts for nearly 7% of all word occurrences (69,971 out of slightly over 1 million). Second in rank is the word 'of', accounting for slightly over 3.5% of words (36,411 occurrences), followed by 'and' (28,852).

5.	Spanish has *to*-datives (*Pedro entregó el paquete a Ana* 'Pedro gave the package to Ana') and double-object constructions (Cuervo, 2007; Demonte, 1995), when the indirect object argument is doubled by a dative clitic, as in *Pedro **le** entregó el paquete a Ana* 'Pedro dat. cl gave a package to Ana'. When the goal is not doubled by a clitic the *a*-phrase is a prepositional dative, when doubled by a clitic a double object. Benefactive datives lend support to Demonte's analysis because the preposition *para* 'for' appears in the prepositional dative construction (*Sara hizo el almuerzo para su hijo* 'Sara made lunch for her son') but in the double-object construction the preposition 'a' appears (*Sara le hizo el almuerzo a su hijo* 'Sara made her son lunch').

References

Agirre A (2015) The acquisition of the dative alternation in English by Spanish speakers. *Vigo International Journal of Applied Linguistics* 12: 63–80.

Akhtar N and Tomasello M (1997) Young children's productivity with word order and verb morphology. *Developmental Psychology* 33: 952–65.

Ard J and Gass S (1987) Lexical constraints on syntactic acquisition. *Studies in Second Language Acquisition* 9: 235–55.

Aronoff M (1976) *Word formation in generative grammar*. Cambridge, MA: MIT Press.

Baerman M, Corbett G, and Brown D (eds) (2010) *Defective paradigms: Missing forms and what they tell us*. Oxford: Oxford University Press.

Baker CL (1979) Syntactic theory and the projection problem. *Linguistic Inquiry* 10: 533–81.

Beck S and Johnson K (2004) Double objects again. *Linguistic Inquiry* 35: 97–124.

Berko J (1958) The child's learning of English morphology. *Word* 14: 150–77.

Berwick R (1985) *The acquisition of syntactic knowledge*. Cambridge, MA: The MIT Press.

Bley-Vroman R (1989) The logical problem of foreign language learning. *Linguistic Analysis* 20: 1–49.

Bley-Vroman R and Yoshinaga N (1992) Broad and narrow constraints on the English dative alternation: Some fundamental differences between native speakers and foreign language learners. *University of Hawai'i Working Papers in ESL* 11: 157–99.

Bowerman M and Croft W (2008) The acquisition of the English causative alternation. In: Bowerman M and Brown P (eds) *Crosslinguistic perspectives on argument structure and structure: Implications for learnability*. Hillsdale, NJ: Erlbaum, pp. 279–307.

Bresnan J and Ford M (2010) Predicting syntax: Processing dative constructions in American and Australian varieties of English. *Language* 88: 168–213.

Brown R and Hanlon C (1970) Derivational complexity and the order of acquisition in child speech. In: Hayes JR (ed.) *Cognition and the development of language*. New York: Wiley, pp. 11–53.

Bruening B (2010) Double object constructions disguised as prepositional datives. *Linguistic Inquiry* 41: 287–305.

Callies M and Szczesniak K (2008) Argument realization, information status and syntactic weight: A learner corpus study of the dative alternation. In: Walter M and Grommes P (eds) *Fortgeschrittene Lernervarietäten: Korpuslinguistik und Zweitsprachenerwerbsforschung [Advanced learner varieties: Corpus linguistics and research into second language acquisition]*. Tübingen: Niemeyer, pp. 165–87.

Campbell AL and Tomasello M (2001) The acquisition of English dative constructions. *Applied Psycholinguistics* 22: 253–67.

Catalán R and Francisco R (2008) Vocabulary input in EFL textbooks. *RESLA* 21: 147–65.

Chang LS (2004) Discourse effects on EFL learners' production of dative constructions. *Journal of National Kaohsiung University of Applied Sciences* 33: 145–70.

Chomsky N (1965) *Aspects of the theory of syntax*. Cambridge, MA: MIT Press.

Chomsky N (2005) Three factors in language design. *Linguistic Inquiry* 36: 1–22.

Chung S (1998). *The design of agreement: Evidence from Chamorro*. Chicago, IL: University of Chicago Press.

Conwell E and Demuth K (2007) Early syntactic productivity: Evidence from dative shift. *Cognition* 103: 163–79.

Cuervo MC (2007). Double objects in Spanish as a second language. *Studies in Second Language Acquisition* 28: 583–615.

De Cuypere L, De Coster E, and Baten K (2009–10) The acquisition of the dative alternation by Russian foreign language learners. *Phrasis* 2: 187–212.

Demonte V (1995) Dative alternations in Spanish. *Probus* 7: 5–30.

Dresher BE (1999) Charting the learning path: Cues to parameter setting. *Linguistic Inquiry* 30: 27–67.

Flores C (2010) The effect of age on language attrition: Evidence from bilingual returnees. *Bilingualism: Language and Cognition* 13: 533–46.

Fodor JD (1998) Unambiguous triggers. *Linguistic Inquiry* 29: 1–36.

Führer M (2009) The dative alternation in the interlanguage of German learners of English Unpublished bachelor's thesis, Universität Siegen, Siegen, Germany.

Gibson E and Wexler K (1994). Triggers. *Linguistic Inquiry* 25: 407–54.

Goldberg AE (1995). *Constructions*. Chicago, IL: University of Chicago Press.

Granger S, Hung J, and Petch-Tyson S (eds) (2002) *Computer learner corpora, second language acquisition and foreign language teaching*. Amsterdam: John Benjamins.

Green GM (1974) *Semantics and syntactic regularity*. Bloomington, IN: Indiana University Press.

Grimshaw J (1990) *Argument structure*. Cambridge, MA: MIT Press.

Gropen J, Pinker S, Hollander M, Goldberg R, and Wilson R (1989) The learnability and acquisition of the dative alternation in English. *Language* 65: 203–57.

Grosjean F (2016) The complementarity principle and its impact on processing, acquisition and dominance. In: Silva Corvalán C and Treffers-Daller J (eds) *Language dominance in bilinguals*. Cambridge: Cambridge University Press, pp. 66–84.

Hale K and Keyser SJ (2002) *Prolegomenon to a theory of argument structure*. Cambridge, MA: MIT Press.

Halle M (1973) Prolegomena to a theory of word formation. *Linguistic Inquiry* 4: 3–16.

Harley H (2002) Possession and the double object construction. *Language Variation Yearbook* 2: 29–68.

Harley H and Miyagawa S (2016) Ditransitives. In: Aronoff M et al. (eds) *Oxford Research Encyclopedia of Linguistics*. Oxford: Oxford University Press.

Hart B and Risley TR (2003) The early catastrophe: The 30 million word gap by age 3. *American Educator* 27: 4–9.

Håkansson G (1987) *Teacher talk: How teachers modify their speech when addressing learners of Swedish as a second language*. Travaux de l'Institute de Linguistique de Lund. Lund: Lund University Press.

Hauser MD, Chomsky N, and Fitch WT (2002) The faculty of language: What is it, who has it, and how did it evolve? *Science* 298(5598): 1569–79.

Hawkins R (1987) Markedness and the acquisition of the English dative alternation by L2 speakers. *Second Language Research* 29: 20–55.

Hornstein N and Lightfoot D (1981) Introduction. In: Hornstein N and Lightfoot D (eds) *Explanation in linguistics: The logical problem of language acquisition*. London: Longman, pp. 9–31.

Hovav MR and Levin B (2008) The English dative alternation: The case for verb sensitivity. *Journal of Linguistics* 44: 129–67.

Hudson Kam CL and Newport EL (2005) Regularizing unpredictable variation: The roles of adult and child learners in language formation and change. *Language Learning and Development* 1: 151–95.

Hudson Kam CL and Newport EL (2009) Getting it right by getting it wrong: When learners change languages. *Cognitive Psychology* 59: 30–66.

Inagaki S (1997) Japanese and Chinese learners' acquisition of the narrow-range rules for the dative alternation in English. *Language Learning* 47: 637–69.

Jackendoff RS (1990) *Semantic structures*. Cambridge, MA: MIT Press.

Jarmulowicz L (2002) English derivational suffix frequency and children's stress judgments. *Brain and Language* 81: 192–204.

Jäschke T and Plag I (2016) The dative alternation in German–English interlanguage. *Studies in Second Language Acquisition* 38: 485–521.

Jelinek E and Carnie A (2003) Argument hierarchies and the mapping principle. In: Carnie A, Harley H, and Willie M (eds) *Formal approaches to function in grammar*. Amsterdam: John Benjamins, pp. 265–96.

Juffs A (1996) *Learnability and the lexicon: Theories and second language acquisition research*. Amsterdam: John Benjamins.

Jung YJ and Miyagawa S (2004) Decomposing ditransitive verbs. In: *Proceedings of SICGG*, Seoul, pp. 101–20.

Krifka M (1999) Manner in dative alternation. *West Coast Conference on Formal Linguistics* 18: 260–71.

Kučera H and Francis W (1967). *Computational analysis of Modern American English*. Providence, RI: Brown University Press.

Le Compagnon B (1984) Interference and overgeneralization in second language learning: The acquisition of English dative verbs by native speakers of French. *Language Learning* 34: 39–67.

Legate JA and Yang C (2007) Morphosyntactic learning and the development of tense. *Language Acquisition* 14: 315–44.

Levin B (1993) *English verb classes and alternations: A preliminary investigation*. Chicago, IL: University of Chicago Press.

Levin B (2008) Dative verbs and dative alternation from a crosslinguistic perspective. Unpublished manuscript, Stanford University, Stanford CA.

Li CN and Thompson SA (1981) *Mandarin Chinese: A functional reference grammar*. Los Angeles, CA: University of California Press.

Lightbown P (1985) Input and acquisition for second language learners in and out of the classroom. *Applied Linguistics* 6: 263–73.

Lignos C and Yang C (2016) Morphology and its acquisition. In: Hippisley A and Stump G (eds) *The Cambridge handbook of morphology*. Cambridge: Cambridge University Press, pp. 765–91.

Long M (1983) Linguistic and conversational adjustments to nonnative speakers. *Studies in Second Language Acquisition* 5: 177–93.

Lozano C (2015) Learner corpora as a research tool for the investigation of lexical competence in L2 Spanish. *Journal of Spanish Language Teaching* 2: 180–93.

MacWhinney B (2000) *The CHILDES project: Tools for analyzing talk*. 3rd edition. Mahwah, NJ: Lawrence Erlbaum.

Maratsos M (2000) More overregularizations after all: New data and discussion on Marcus, Pinker, Ullman, Hollander, Rosen and Xu. *Journal of Child Language* 27: 183–212.

Marcus G, Pinker S, Ullman MT, Hollander M, Rosen J, and Xu F (1992) *Overregularization in language acquisition: Monographs of the Society for Research in Child Development*. Chicago, IL: University of Chicago Press.

Marefat H (2005) The impact of information structure as a discourse factor on the acquisition of dative alternation by L2 learners. *Studia Linguistica* 49: 66–82.

Mazurkewich I (1984) The acquisition of the dative alternation by second language learners and linguistic theory. *Language Learning* 34: 91–108.

Mitchell T (1982) Generalization as search. *Artificial Intelligence* 18: 203–26.

Montrul S (2008) *Incomplete acquisition in bilingualism: Re-examining the age factor*. Amsterdam: John Benjamins.

Montrul S and Ionin T (2012) Dominant language transfer in Spanish heritage speakers and L2 learners in the interpretation of definite articles. *The Modern Language Journal* 96: 70–94.

Myles F (2005) Interlanguage corpora and second language acquisition research. *Second Language Research* 21: 373–91.

Niyogi P and Berwick R (1996) Formalizing triggers. *Linguistic Inquiry* 27: 605–22.

Oehrle RT (1976) The grammatical status of the English dative alternation. Unpublished PhD thesis, Massachusetts Institute of Technology, Cambridge, MA, USA.

Oh E (2006) A case study of structural transfer: The acquisition of English double objects by adult Korean speakers. Unpublished manuscript, University of Southern California, Los Angeles, CA, USA.

Oh E (2010) Recovery from first language transfer: The second language acquisition of double objects by Korean speakers. *Second Language Research* 26: 407–39.

Oh E and Zubizarreta ML (2006) *Against morphological transfer*. In: Deen KU, Nomura J, Schulz B, and Schwartz BD (eds) *Proceedings of the inaugural conference on Generative Approaches to Language Acquisition: North America*. Honolulu, HI, pp. 261–72.

Osherson DN and Smith EE (1981) On the adequacy of prototype theory as a theory of concepts. *Cognition* 9: 35–58.

Pechmann T, Uszkoreit H, Engelkamp J, and Zerbst D (1994) *Word order in the German middle field: Linguistic theory and psycholinguistic evidence*. Report No. 43. Saarbrücken: Universität des Saarlandes.

Perpiñán S and Montrul S (2006) On binding asymmetries in dative alternation construction in L2 Spanish. In: Klee CA and Face TL (eds) *Selected proceedings of the 7th conference on the acquisition of Spanish and Portuguese as first and second languages*. Somerville, MA: Cascadilla Proceedings Project.

Pesetsky D (1995) *Zero syntax: Experiencer and cascades*. Cambridge, MA: MIT Press.

Pinker S (1989) *Learnability and cognition: The acquisition of argument structure*. Cambridge, MA: MIT Press.

Pinker S (1999) *Words and rules: The ingredients of language*. New York: Basic Books.

Sakas WG and Fodor JD (2012) Disambiguating syntactic triggers. *Language Acquisition* 19: 83–143.

Sakas W, Yang C, and Berwick R (in press) Parameter setting is feasible. *Theoretical Linguistics*.

Schmidt M (2013) *Language attrition*. Cambridge. Cambridge University Press.

Schuler K, Yang C, and Newport E (2016) Testing The Tolerance Principle: Children form productive rules when it is more computationally efficient to do so. In: *The 38th Cognitive Society Annual Meeting*. Mahwah, NJ. Cognitive Science Society, pp. 2321–26.

Schwartz B and Sprouse R (1996) L2 cognitive states and the Full Transfer/Full Access Model. *Second Language Research* 12: 40–72.

Seedhouse P (2012) *Conversation analysis and classroom interaction*. Malden, MA: Wiley-Blackwell.

Seliger H (1991) Language attrition, reduced redundancy, and creativity. In: Seliger H and Vago R (eds) *First language attrition*. Cambridge: Cambridge University Press, pp. 173–84.

Slabakova R (2008) *Meaning in the second language*. Berlin: de Gruyter.

Snyder W and Stromswold K (1997) The structure and acquisition of English dative constructions. *Linguistic Inquiry* X: 281–317.

Tanaka S (1987) The selective use of specific exemplars in second-language performance: The case of the dative alternation. *Language Learning* 37: 63–88.

Tono Y (2004) Multiple comparisons of IL, L1 and TL corpora: The case of L2 acquisition of verb subcategorization patterns by Japanese learners of English. In: Aston G, Bernardini S, and Stewart D (eds) *Corpora and language learners*. Amsterdam: John Benjamins, pp. 45–66.

Tyler A and Nagy W (1989) The acquisition of English derivational morphology. *Journal of Memory and Language* 28: 649–67.

White L (2003) *Second language acquisition and universal grammar*. Cambridge: Cambridge University Press.

Whong-Barr M and Schwartz BD (2002) Morphological and syntactic transfer in child L2 acquisition of the English dative alternation. *Studies in Second Language Acquisition* 24: 579–616.

Woods R (2015) The acquisition of the dative alternation by German–English bilingual and monolingual children. *Linguistic Approaches to Bilingualism* 2: 252–84.

Xu F and Pinker S (1995) Weird past tense forms. *Journal of Child Language* 22: 531–56.

Yang C (2002) *Knowledge and learning in natural language*. Oxford: Oxford University Press.

Yang C (2005) On productivity. *Linguistic Variation Yearbook* 5: 333–70.

Yang C (2016) *The price of linguistic productivity: How children learn to break the rules of language*. Cambridge, MA: MIT Press.

Yang S, Huang YY, and Lee N (2000) Some reflections on English textbook input for Hong Kong students: Based on a case study of tense aspect acquisition problems. *Asia Pacific Journal of Language Education* 3: 1–24.

Yip K and Sussman G (1997) Sparse representations for fast, one-shot learning. In: *Proceedings of the National Conference on Artificial Intelligence*. Washington, DC: American Associate for Artificial Intelligence, pp. 521–27.

Zara JV, Oliveira FL, and Souza RA (2013) Selective transfer in the acquisition of English double object construction by Brazilian learners. *Alfa, Sao Paulo* 57: 519–44.

Zipf GK (1949) *Human behavior and the principle of least effort: An introduction to human ecology*. Cambridge, MA: Addison-Wesley.