The Sino-Korean Light Verb Construction and Lexical Argument Structure

Chung-hye Han[†] and Owen Rambow[‡]

[†]IRCS, University of Pennsylvania
3401 Walnut St., Suite 400
Philadelphia, PA 19104
chunghye@linc.cis.upenn.edu

[‡]ATT Labs-Research, B233
180 Park Ave., PO Box 971
Florham Park, NJ 07932
rambow@research.att.com

Abstract

In Korean, a class of lexemes of Chinese origin exhibit both nominal and verbal behavior. Specifically, they can assign lexically idiosyncratic case, but require a semantically vacuous light verb in order to form a sentence and are themselves marked with accusative case. In this paper, we propose a TAG-based account of this behavior, and propose some generalizations towards a pure representation of lexical argument structure.

1. Linguistic Facts and Issues

In this paper, we provide a syntactic analysis of Sino-Korean light verb constructions (LVC henceforth) that are composed of the light verb *ha* and an activity-denoting noun of Chinese origin.[1] We will refer to this activity-denoting noun as the 'base' of the LVC. The argument structure of LVCs come from the base, and the light verb is semantically vacuous and does not assign any theta roles. This is shown by the fact that although the examples in (1) all contain *ha*, they have different argument structures.

(1) a. John-i   swuhak-ul yenkwu-lul   ha-yess-ta.
       John-Nom math-Acc research-Acc HA-Past-Decl

   'John researched math.'

   b. Kicha-ka Seoulyek-ey   tochak-ul   ha-yess-ta.
      train-Nom Seoul-station-at arrival-Acc HA-Past-Decl

   'The train arrived at Seoul station.'

   c. Kicha-ka Seoulyek-eyse   chwulpal-ul   ha-yess-ta.
      train-Nom Seoul-station-from departure-Acc HA-Past-Decl

   'The departed from Seoul station.'

For instance, the arguments in (1a) are agent and goal, those in (1b) are patient and goal, and those in (1c) are patient and source.

If, however, the theta roles in LVCs are assigned by the base, it is puzzling why the argument NPs are syntactically realized outside of the base NP. The case postpositions such as *Acc*, *-ey* and *-eyse* on the argument NPs indicate that they are daughters of VP, and not the base NP. An

NP that is a daughter of another NP requires genitive or null case postposition in Korean. We will refer to the first kind of case as VERBAL CASE, and the second as NOMINAL CASE. Moreover, as noted by (Grimshaw & Mester, 1988), there are restrictions on argument realization, which can be clearly shown with ditransitive LVCs, as in (2).

(2) a.   John-i   Mary-eykey inhyung-ul senmwul-ul ha-yess-ta.
           John-Nom Mary-to   doll-Acc   gift-Acc   HA-Past-Decl

           'John gave a gift of a doll to Mary.'

     b.   John-i   Mary-eykey inhyung(-uy) senmwul-ul ha-yess-ta.
           John-Nom Mary-to   doll(-Gen)   gift-Acc   HA-Past-Decl

     c.   * John-i   inhyung-ul Mary-eykey-uy senmwul-ul ha-yess-ta.
           John-Nom doll-Acc   Mary-to-Gen   gift-Acc   HA-Past-Decl

     d.   * John-i   Mary-eykey-uy inhyung(-Gen) senmwul-ul ha-yess-ta.
           John-Nom Mary-to-Gen   doll(-Gen)   gift-Acc   HA-Past-Decl

The base *senmwul* ('gift') assigns agent, goal and theme. In (2a), all the argument NPs are realized outside of the base NP. In (2b), the agent and goal arguments are realized outside of the base NP, but the theme argument is realized inside the base. However, it is not possible to realize theme argument outside of the base when the goal argument is realized inside the base, as shown in (2c), and it is not possible to realize both theme and goal arguments inside the base, as shown in (2d).

(Grimshaw & Mester, 1988) (G&M henceforth) summarize the restrictions on argument realization as follows: (i) the subject argument must always be outside the base NP; (ii) at least one argument apart from the subject must be outside the base NP; and (iii) for nouns that take a theme and a goal, if the theme argument is realized outside the base NP, the goal must also be realized outside the base NP. In what follows, we first briefly discuss some previous analyses and their shortcomings, and present our own analysis using the framework of Feature Based Lexicalized Tree Adjoining Grammar. We discuss English data in comparison, and conclude with a discussion of noun phrases.

2. Previous Analyses

According to G&M, a light verb such as *ha* has no argument structure on its own and it occurs with a noun which is 'theta-transparent.' Theta-transparent nouns can transfer some or all of their arguments to the argument structure of the light verb. This mechanism allows the light verb to directly assign theta roles to the argument NPs in syntax and such argument NPs are realized outside the base NP. They further assume (following much previous work) that arguments have a hierarchy according to prominence. For instance, the agent is more prominent than the goal, which is more prominent than the theme. Based on this assumption, they propose that when a theta role is transfered (e.g., the theme), any theta roles that are higher in prominence must transfer as well (i.e, the agent and goal). This explains the ungrammaticality of (2c). G&M also stipulate that the base noun must transfer at least one internal argument in order to be licensed. Otherwise, the theta-criterion is violated, since the base noun does not receive a theta role from anywhere. This is why (2d) is ungrammatical under G&M's system.

G&M wrongly predict that intransitive LVCs do not exist, since there is no internal argument to participate in the transfer. But intransitive LVCs clearly do exist, as shown in (3) and (4). Note that while (4) may be ambiguous between a heavy and light verb reading of *ha*, (3) is not, since the subject is not an agent.

(3) John-i   samang-ul ha-yess-ta.   (4) John-i   swuyuong-ul   ha-yess-ta.
     John-Nom death-Acc HA-Past-Decl   John-Nom swimming-Acc HA-Past-Decl
     'John died.'   'John was swimming.'

For this reason, (Yoon, 1991) rejects G&M's argument transfer theory and proposes 'argument sharing' mechanism. He argues that the light verb is thematically underspecified and so unsaturated. This forces the base noun which has theta structure and the light verb to undergo the operation of Theta Identification, allowing the argument structure of the base noun and that of the light verb to be shared. This sharing is viewed as the unification of the argument structure of the base noun into the underspecified argument structure of the light verb. Yoon's theory predicts that when there are more than one internal arguments, they must all be realized outside of the base NP. But this is an incorrect prediction: in ditransitive LVCs, while the goal argument is realized outside the NP, the theme argument can be realized inside NP, as shown in (2b).

The same problem persists in (Park, 1992). He argues that the categorial status of the base is not a noun, but a verb. Thus, it assigns theta-roles just as any other verbs. The light verb is simply an auxiliary verb that supports inflection. But if the base is simply a verb, then (2b) is wrongly predicted to be ungrammatical.

3. TAG Analysis

The key to our analysis is the assumption that the base is underspecified with respect to word class (verb or noun). We propose that this base is the anchor of an elementary tree with all its arguments and that it acquires a noun status only after the light verb adjoins into the elementary tree. The assumption that the category of the base is unspecified is well-motivated for two reasons: (i) The base form originates from Chinese, in which the same form is used both as a noun and a verb, (ii) there is no consensus in the literature as to what the category of the base is and positing that it is either a noun or a verb leads to difficulties, as discussed in §2. We represent this by using the label X for its category (which projects to XP). We also assume that each node in a tree is associated with a category feature CAT with values such as V(ERB) and N(OUN). The CAT feature of nodes labeled V, VP, or S is necessarily V for both the top and bottom feature structures, while nodes labeled N or NP necessarily have [CAT:N].[2] But the CAT feature of the base of LVC is unspecified. In addition, we assume that nodes in a projection have a full set of morpho-syntactic features. In this paper we use only the binary feature [TENSED:]. We assume that the base is [TENSED:-] (since it carries no tense morphology), that the S node is marked [TENSED:+], and that the TENSED feature is shared among the nodes of a projection. We assume that when a lexeme (of any category) forms a syntactic predication structure it projects to a maximal verbal projection (VP) and we refer to this VP as the PREDICATE. Furthermore, following (Heycock & Lee, 1989), we assume that in Korean, nominative case is assigned by the predicate, not by Infl. (Heycock & Lee, 1989) use as evidence the presence of multiple nominative constructions and the fact that infinitivals can have nominative case-marked subject. As a result, all clausal structures need a VP node as a sister to the subject argument to license nominative case.[3] We also assume that the lexeme projects all of its argument positions in canonical order according to theta hierarchy. That is, the most prominent argument attaches

---

[2]The node labels are not actually used in our analysis, and we could also label all nodes XP. We retain the traditional labels for clarity.

[3]This is compatible with the XTAG analysis of the predicative use of nouns and adjectives in English, the trees for which project from N (or A) to S via NP (AP) and VP (though perhaps the VP is less motivated in English than in Korean because the adjoined auxiliary provides the nominative case in English, not the predication structure itself).

to the highest projection, and the least prominent attaches to the lowest projection.[4] We assume that each lexeme idiosyncratically fixes a case grid for its arguments,[5] which is only realized in appropriate syntactic contexts. (Thus, rather than speak of unified case assignment, we will henceforth speak of case assignment by the lexical head and subsequent *case realization* in a particular syntactic context.)
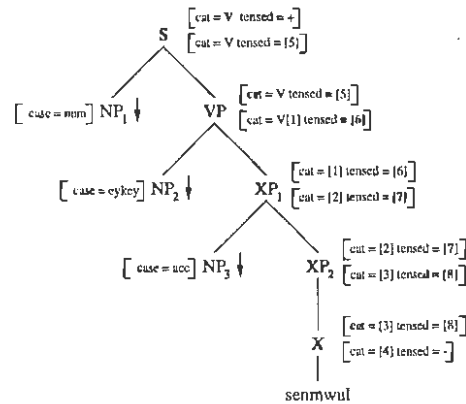

Figure 1: Sino-Korean base lexeme *senmwul* 'gift' projecting to a predicative structure

In Korean, a verbal case such as *Nom, Acc, eykey* is realized when the head has feature [CAT:V], while if the head has feature [CAT:N], *Nom* and *Acc* are realized as *Gen* or null, while any other postpositional case is realized as that postposition followed by *Gen*.[6] As an example, the elementary tree for the base *senmwul* 'gift' is shown in Figure 1, which is a ditransitive structure. Assuming the argument hierarchy agent – goal – theme, as in G&M, agent (as indicated by [case:nom]) is attached to S, goal (as indicated by [case:eykey]) is attached to VP and theme (as indicated by [case:acc]) is attached to XP_1. (The subscripts on nodes are used only for distinguishing different nodes, they play no role in the analysis.)

We now turn to the light verb. Its properties can be best explained in comparison to heavy *ha*. Heavy *ha* (Figure 2, left) is a standard transitive verb: it has two arguments (i.e., theta-marked dependents), to which it assigns nominative and accusative case, respectively. (Nominative case is realized in a syntactic predication environment, while accusative case is realized whenever the lexical head is verbal, which it is by assumption.) The light verb *ha* (Figure 2, right) differs from the heavy *ha* in that the light *ha* loses its ability to assign theta roles: it has no arguments of its own. Furthermore, it has lost its ability to create a predication structure. Thus it can no longer assign nominative case. It therefore does not project to a VP after taking its complement, but only to an XP, with [CAT:V]. However, light *ha* retains its ability to assign accusative case as well as the feature [CAT:N] to its complement. Since there is only one substitution node left, and since both root and substitution node are labeled XP, the tree is optionally an auxiliary tree (as is the case for English predicative auxiliary trees).

---

[4]We do not deal with the issue of optional arguments in this paper.

[5]Alternatively, we could assume each lexeme idiosyncratically chooses a set of theta-roles and then devise a functional mapping that derives the cases of a lexeme from the set of theta roles. Such an approach is only a notational variant of ours, and, as it has no additional content, we do not pursue it here.

[6]In other languages, the mapping between verbal and nominal case may not be as straightforward and each may be marked idiosyncratically from the head.
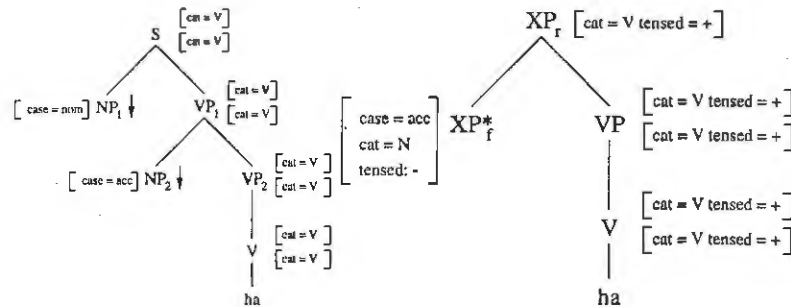
Figure 2: Heavy *ha* (left) and light *ha* (right)

The feature clash between [TENSED:-] on the base and [TENSED:+] at the root of the tree in Figure 1 forces the adjunction of light *ha*. We can adjoin this tree at either of the two XP nodes in the tree in Figure 1: if we adjoin at XP_1, the lowest argument NP_3 is realized with genitive/null case (as in (2b)); if we adjoin at XP_2, the lowest argument is realized with verbal case (as in (2a)). In both cases, NP_1 and NP_2 are realized with verbal cases. Our analysis predicts the pattern of data introduced in §1, while specifically avoiding the less appealing aspect of G&M's and Yoon's analyses, namely the cumbersome mechanisms of argument transfer or theta-identification, and the stipulation that agent and at least one internal argument must be transferred from the base. In fact, our analysis correctly predicts the existence of intransitive LVCs such as (3). The unique argument is the most prominent argument trivially and so it simply attaches to S, and receives nominative case from the predicate VP.[7]

4. Comparison to English Light Verb Constructions

Unlike in Korean, in English LVCs, all the internal arguments are realized within an NP. For instance, in a ditransitive light verb construction such as *make a donation*, the theme and goal arguments are nominal, as shown in (5). The theme *10,000 dollars* requires *of*, indicating that it is a sister of a noun, not a verb. Although it is rather difficult to tell whether the goal *to the charity* is nominal or verbal, it clearly has the possibility of being nominal, as shown in (6).

(5)   a.   * John made a donation 10,000 dollars to the charity last year.

      b.   John made a donation of 10,000 dollars to the charity last year.

(6)   Twenty donations of 10,000 dollars to the charity occurred last year.

In the spirit of (Larson, 1988), we assume the structure given in Figure 3 (left) for ditranstives with a dative NP. We do not postulate a privileged predicate VP for English, in contrast to Korean, since nominative case assignment in English is done through a tensed verb. This is supported by the fact that infinitivals in English cannot have a nominative case-marked subject NP. We propose that the light verb *make* anchors the auxiliary tree given in Figure 3 (right). The light verb tree is similar to the Korean light verb tree in that the root node has [CAT:V] and the foot node has [CAT:N]. This tree can only adjoin to XP1, constrained by the English SVO word

---

[7]In the English XTAG grammar (The XTAG-Group, 1998), the light verb and the base noun are anchors of a single elementary tree. Although this analysis works for English for all practical purposes, extending it to Korean forces us to postulate multiple elementary trees for a single Sino-Korean lexeme, failing to account for the systematic variation in the syntactic realization of its argument structure. In §4, we will show that our analysis can also be extended to English, allowing a uniform anaylsis of LVCs in both Korean and English.

order. Once the adjunction takes place at XP1, the [CAT:] feature on XP2, X2′, X2, X1′ and X1 will all come to have the value N. Thus, NP2 must be realized with a preposition *of*, and X1 must be realized as a nominal form *donation*. Further, the adjoined light verb *make* assigns a nominative case to subject NP₁, and an accusative case to its NP complement *a donation of 10,000 dollars to the charity*.
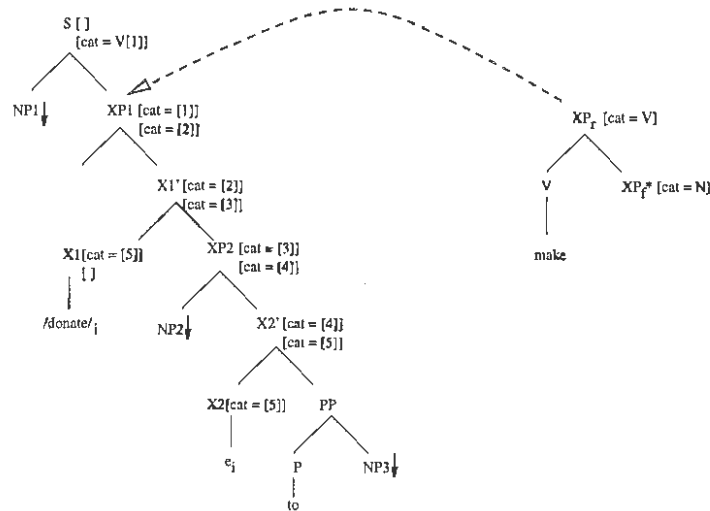

Figure 3: Ditransitive with Dative NP (left) and Light Verb *make* (right)

According to Larson, ditransitive sentences can undergo dative shift deriving a double object construction. If dative shift applied to the ditransitive structure in Figure 3, one might expect to derive from (5b) a string as in (7).

(7)   * John made the charity a donation of 10,000 dollars last year.

If we blindly apply dative shift to the ditransitive structure in Figure 3, abstracting away from details, the case marking *to* on the goal argument NP₃ (*the charity* in (5b)) would disappear, and so it could in principle move up to [Spec, XP1] to receive case. But once the adjunction of the light verb takes place at XP1, all the projections of X2 and X1 would become nominal, disallowing any case assignment to NP₃. This leads us to conclude that dative shift cannot apply to sentences such as (5b), which means that the only way to derive (7) is through a ditransitive full verb *make* as in *John made Mary a cake*. But then, this ends up in semantic conflict between full ditransitive *make* and *donation*. In (7), *make* requires a direct object who is a beneficiary of John's action, but the direct object *the charity* is behaving as a recipient due to the presence of the noun *donation*.

Our analysis on the Korean LVC therefore can be extended to the English LVC, allowing for a unified account of LVCs in both languages as well as accounting for their differences.

5. Towards a Pure Representation of Lexical Argument Structure

In Korean, the Sino-Korean lexemes that we have discussed in §1 can also project to an NP. In this case, all arguments are obligatorily realized using genitive case marking.

(8)   John-uy   Mary-eykey-uy inhyung(-uy) senmwul
       John-Gen Mary-to-Gen   doll(-Gen)   gift

'John's gift of a doll to Mary'

The question arises how these NPs are represented. We assume that the same Sino-Korean lexeme can project to an NP or to an S with the same argument structure, and thus we need a unique representation of lexical argument structure. Our current representation is fixed to project to S. We propose to extend our analysis by assuming that there is an underlying *lexical argument frame* (LAF), a representation of pure argument structure in which syntactic categories (i.e., node labels and the CAT feature) are not yet fixed. Node labels are added (as features) during the lexical phase of a derivation, when an LAF is instantiated with syntactic features prior to the syntactic derivation involving other trees (also see (Chomsky, 1970)).

Specifically, we will now represent *all* node labels as XP, X', or X. The difference between the verbal and nominal node labels will now be represented at all nodes using the feature [CAT:]. But [CAT:] does not yet account for the difference between VP and S, so we need to introduce new features in order to represent our analysis of the LVC (which crucially relies on the VP/S distinction).

As discussed in §3, in Korean, the VP represents an unsaturated syntactic predication structure, the nominal argument is the subject of predication, and the S represents a saturated predication structure. We will capture this analysis with two new binary features, PRED and SUBJPRED, which indicate the presence or absence of an unsaturated predication structure and of a saturated predication structure, respectively. Note that [PRED:- SUBJPRED:+] does not make sense and is assumed not to occur.
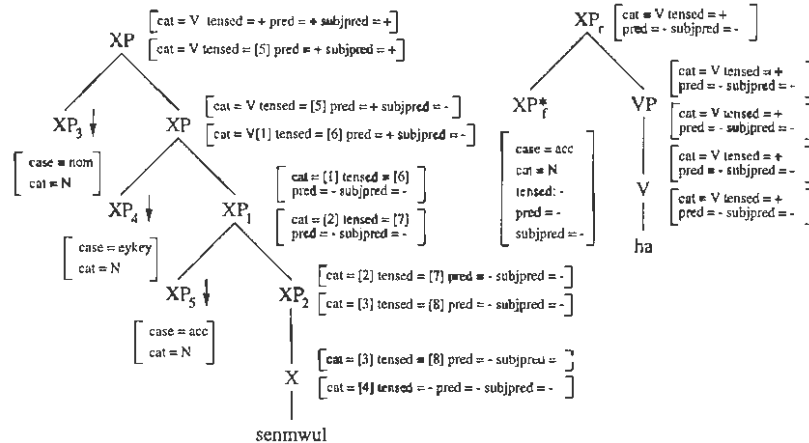

Figure 4: Base lexeme *senmwul* 'gift' (left) and light verb *ha* (right), not using node labels

We now show how our new way of representing node labels accounts for the LVC data by assigning the new features to the nodes in our example (Figure 1). Clearly, none of the nodes labeled XP in Figure 1 form predication structures, so all feature structures associated with them are [PRED:-, SUBJPRED:-]. The subtree anchored at the VP node represents the predicate, so the bottom feature structure of the VP node is [PRED:+, SUBJPRED:-]. Since no further adjunction at the VP node can alter the fact that a predication structure exists, the top feature structure is also [PRED:+, SUBJPRED:-]. Finally, the subtree rooted at the S node (even if adjuncts are adjoined to it later) is the saturated predication structure, so both bottom and top feature structures get [PRED:+, SUBJPRED:+]. The new tree for *semnwul* is shown in Figure 4 on the left.

We now turn to light *ha*. Since it is a light verb, it contributes the [CAT:V] information, but does not create a syntactic predication structure on its own (since it is semantically vacuous). Furthermore, it cannot be adjoined into an already existing predication structure, because predication structures are necessarily verbal (by assumption) and *ha*'s footnode is labeled [CAT:N]. Therefore, the root and foot nodes of light *ha* have top and bottom features labeled [PRED:-, SUBJPRED:-]. The new tree for light *ha* is shown in Figure 4 on the right. It is clear that our previous analysis of the light verb construction facts carries over essentially unchanged to the new representation.

Thus, we have shown that we can represent the information contained in node labels as features in a motivated manner. We can now define an LAF (i.e., a syntactically neutral representation of lexical argument structure) as a tree projected from a lexeme with substitution nodes for all its arguments, in which all syntactic features (CAT, PRED, SUBJPRED, TENSED) are undefined. Specifically, if we take the representation in Figure 4 on the left and set all syntactic features to undefined, then we obtain the LAF for *semnwul*. This LAF is the starting point for lexical derivations. Not all assignments of values for the four syntactic features are valid. In fact, as mentioned above, in Korean, only verbal structures (with [CAT:V]) can create nodes with [PRED:+], [SUBJPRED:+], or [TENSED:+] — the features simply don't make sense for [CAT:N]. Thus, if the choice of projecting to a verbal predication structure is made, then the analysis presented in §3 follows. If instead we choose [CAT:N] at the root node, then we do not get a predication structure, light *ha* cannot be adjoined, and all arguments are realized in the genitive, as desired.

6. Conclusion

We have shown how we can derive the Sino Korean LVC by assuming that the base lexemes have a single entry in the lexicon and a single light verb *ha* is adjoined into them to obtain the LVC. We have suggested that our analysis extends to English light verb constructions as well. Finally, we have shown how this analysis points to a TAG-based representation of lexical argument structure independent of syntactic categories such as lexical class. In future work, we intend to investigate more cross-linguistic data on LVCs in order to verify that our approach carries over to different types of LVCs, and we intend to verify our TAG-based representation of lexical argument structure by investigating nominalization in different languages.

References

CHOMSKY N. (1970). Remarks on nominalization. In R. JACOBS & P. ROSENBAUM, Eds., *Readings in English Transformational Grammar*. Waltham, Mass.: Ginn.

GRIMSHAW J. & MESTER A. (1988). Light verbs and theta-marking. *Linguistic Inquiry*, 19, 205–232.

HEYCOCK C. & LEE Y.-S. (1989). Subjects and predication in Korean and Japanese. In H. HOJI, Ed., *Japanese/Korean Linguistics*, volume 2, p. 239–254. Stanford: CSLI.

LARSON R. (1988). On the double object construction. *Linguistic Inquiry*, 19 (3 p.), 335–392.

PARK K. (1992). *Light Verb Constructions in Korean and Japanese*. PhD thesis, University of North Carolina at Chapel Hill.

THE XTAG-GROUP (1998). *A Lexicalized Tree Adjoining Grammar for English*. Technical Report IRCS 98-18, UPenn.

YOON J. (1991). Theta operation and the syntax of multiple complement constructions in Korean. In S. KUNO, I.-H. LEE, J. WHITMAN, J. MALING, Y.-S. KANG & Y. JOO KIM, Eds., *Harvard Studies in Korean Linguistics*, volume 4. Dept. of Linguistics, Harvard University.