

# On the Psycholinguistic Assessment of Relative Language Strength in Bilinguals: Evidence from Korean Heritage Speakers

Amy J. SCHAFER, William O'GRADY, On-Soon LEE,  
Jawee PERLA, Julia WIETING and Katherine PERDUE

Department of Linguistics, University of Hawaii  
1890 East-West Rd, Honolulu HI, USA

E-mail: {aschafer, ogrady, onsoon, perla, wieting, perdue}@hawaii.edu

**Abstract** Language loss is a pressing concern for many of the world's languages. A major obstacle to the early diagnosis of language loss and to the assessment of language maintenance efforts is the absence of an easy-to-use psycholinguistic measure of language strength. We have developed a series of simple psycholinguistic tasks that exploit the fact that the relative strength of two languages is reflected in the speed with which their words and structure-building routines are accessed in the course of speech. Test results can therefore be used to assess language attrition and revitalization in bilingual communities of many different types.

**Key words** Bilingualism, Heritage Language, Korean, Language Attrition, Picture Naming, Psycholinguistics

## 1. Introduction

Concern over the perilous plight of a large portion of the world's languages has far outstripped the development of the methods needed to assess the state of their decline and to monitor the success of revitalization efforts. To date, researchers have been forced to rely on indirect and informal measures of language loss and endangerment—*anecdotal reports by field workers and self-assessments by the speakers themselves in response to questionnaires, census surveys, and the like.* In many cases, this information simply confirms a decline that is already firmly entrenched, perhaps irreversibly, for lack of earlier indications of endangerment.

It is difficult to overestimate the challenges associated with assessing language endangerment, especially in its early stages. Information must be gathered from representative cross-sections of the community, it must track subtle changes in linguistic practices that signal imminent language shift, and it must be updated frequently in response to the ongoing social and economic forces

that contribute to language change in the first place.

The development of assessment techniques capable of providing information related to language shift encounters a second set of challenges. The tests must be economical, they must be easy to administer even in field situations, and their results must be easy to interpret. And perhaps most challenging of all, they must be adaptable for use with a very wide range of languages, including those about which very little is known.

The Hawai'i Assessment of Language Access (HALA) project (O'Grady et al. 2009) offers a simple and effective way to assess language strength in bilinguals. Based on independently established psycholinguistic principles, the HALA tests exploit the fact that the relative strength of two (or more) languages is reflected in the speed with which their words and structure-building routines are accessed in the course of speech. Test results can therefore be used to assess language attrition and revitalization in bilingual communities of many different types. Indeed, the HALA tests are designed to be language-neutral; they can be used to investigate any combination of languages.

## 2. Background

When a community shifts to a new language, it is invariably in response to external economic, social, and political pressures. Yet, language loss is ultimately a neurological phenomenon. Of necessity, it involves changes to the words, structure-building operations, and other resources that are implemented in the brain as ‘language’ and employed in the course of communication.

The profile for language shift and language loss is a familiar one—the delicate balance of bilingualism, sometimes stable for generations, tilts in favor of one language, whose greater and more varied use gradually marginalizes the other language, devaluing its status in the community and removing much of the incentive for its maintenance among younger speakers.

The maintenance of two language systems at comparable levels of activation—the sort of bilingual state that staves off language loss—is no easy task. As De Bot (2004: 234) notes, languages ‘need maintenance and advanced use ... learning another language does not remove older languages from memory, but does push them more to the background and makes it accordingly more difficult to access them.’

The factor that contributes most directly to the maintenance of a linguistic system is the frequency with which it is used. Put simply, the more often the vocabulary items and structure-building routines of a particular language are activated, the more accessible they are. And of course, the more accessible a linguistic system is, the more likely speakers are to feel comfortable using it. There is a natural cycle here—as a language becomes less accessible through infrequent use, its speakers become reluctant to use it, further decreasing its accessibility and creating the downward spiral that ultimately leads to language loss.

A widely acknowledged psycholinguistic reflex of accessibility is speed—a more highly activated lexical item or structure-building routine is accessed more quickly than a less highly activated counterpart. Thus, frequency of use translates into higher activation or strength, which makes possible quicker access.

It is well-established for monolingual speakers that the relative frequency with which lexical items are used is an excellent predictor of their accessibility—more frequently used words are

routinely produced more quickly on naming tasks than their less frequent counterparts (e.g., Bates et al. 2003).

Comparative measures of naming times have long been used to diagnose and confirm language strength in bilinguals, beginning with work by Ervin (1961) on Italian and English and by Májiste (1979) on Swedish and German; see also more recent work by Gollan et al. (2008), among many others. This work clearly establishes two generalizations: (i) naming latencies are shorter in a speaker’s dominant language, and (ii) they are shorter (within any language) for more frequent items than for less frequent ones. Just as importantly, other factors of potential relevance to language access do not appear to nullify those effects. For example, Bates et al. (2003) report strong cross-language correlations in frequency and naming times of equivalent lexical items for the seven languages that they examined, also noting that factors such as word length, syllable structure and morphological composition are less stable and less important than frequency and conceptual familiarity in predicting naming times.

Considerably less work has been done on phrase-building, especially in the context of bilingualism. However, there is very substantial evidence that phrase building requires time commitments above and beyond those associated with the selection of the initial lexical item (e.g., Allum & Wheeldon 2007). There is thus an opportunity to test the hypothesis that when two languages have phrase-building routines of comparable complexity, it takes longer to access those routines in the weaker language.

## 3. The HALA Project

The HALA project seeks to exploit the diagnostic potential of speech production latencies to develop a suite of simple tests that can be used to assess the relative strength of the languages in bilingual speakers and communities. In its current form, the HALA instrument consists of three tasks:

*A body-part naming task*, consisting of 43 test items, arranged into three strata. The first stratum elicits very basic body part terms (*face, foot, ear, head*, etc.); the subsequent two strata elicit less frequent and more specialized terms.

*A nature-image naming task*, consisting of 48 test items, arranged into three strata. The first stratum elicits the most basic

nature terms (*sun, moon, lightning, mountain, etc.*); the subsequent two strata elicit less frequent and more specialized terms.

A *phrase-building task*, consisting of 120 test items, arranged into three strata. The first stratum elicits relatively simple phrases typically consisting of just two content words. For example:

<i>small cat</i>	size adjective + noun
<i>green leaf</i>	color adjective
<i>pig's ear</i>	possessor + noun
<i>shoe and fish</i>	noun + conjunction + noun

The subsequent two strata elicit progressively more complex phrase types, involving extensions and combinations of the above phrase types and, in some cases, the use of numerals.

In each task, subjects are presented with a series of pictures—simple scenes in the case of the body part and nature image tasks, and contrasting sets of objects in the case of the phrase-building task. The pictures are presented at a predetermined rate in a sequence that has been incorporated into a video file, which can be downloaded and played on either a PC or Macintosh computer. Participants simply name/describe each picture as it appears on the screen before them. Their responses are digitally recorded for subsequent analysis of accuracy and, more importantly, of response latencies. Each task takes only a few minutes to run in each language – about five minutes each for the two naming tasks, and fifteen minutes for the phrase-building task – making it possible to collect all of the necessary data from a participant in just one or two short sessions.

## 4. Preliminary Results: Korean Heritage Speakers

### 4.1 Body Part Naming Task

O'Grady et al. (2009) developed and used a naming task to estimate the relative strength of any pair of languages. The task focuses on body part names, for several reasons: (i) common body part terms can be expected to have counterparts in all languages, (ii) they are relatively easy to elicit through pictures, (iii) at least some body part terms are basic enough to have been acquired by all users of a language, and (iv) basic body part terms are relatively resistant to replacement by borrowing, so that their elicitation is more likely to yield words from the native stratum of a language's vocabulary than is (say) the names of electronic devices.

The relevant terms were elicited in response to a series of photographs, in which the key body part was marked with a red circle superimposed on a black and white photo. Naming times were measured in milliseconds from the onset of the photo to the onset of the response, and these times were then compared for the two participants' two languages.

Eleven Korean-English bilinguals participated in our first test of the body part naming task. These speakers were chosen for the following reasons: (i) they were highly fluent in both languages, (ii) independent assessments had determined that English was their stronger language, and (iii) they were similar to speakers of endangered languages in having been exposed to a family language (Korean) at home and to an overwhelmingly dominant competitor language (English) outside the home.

The participants exhibited a very high level of accuracy in both languages, with no significant effect of language on accuracy overall or in any of the subsets of words in our test, despite a small numerical advantage for English (see O'Grady et al for details). This confirms that our participants were in fact highly bilingual.

The calculation of naming times, the key measure in our task, was conducted for just those test items in which the stimulus picture had been correctly identified. It revealed significantly faster naming times for all three strata of vocabulary items in English, compared to Korean.

In addition to confirming that English is the dominant language for the participants, this finding underlines one of the principal advantages of the HALA approach to the assessment of language strength: English emerged as the stronger language for all three vocabulary strata even though the participants were all highly fluent in Korean and even though no reliable difference was evident on accuracy measures. A similar effect held when the analysis was restricted to the five most accurate participants (O'Grady et al 2009).

In some language situations, it may be difficult for a researcher to collect native or near-native speaker judgments on accuracy. Yet if naming time is the strong indicator of language strength that we believe it is, the appropriate contrasts could well emerge even if it is not practical to rely on accuracy measures. To evaluate the

plausibility of this outcome, we also calculated naming times for this dataset without removing inaccurate responses. In this data treatment all responses were included, except for a handful of naming times that were either less than 100 ms or greater than 8000 ms. The results again showed the expected significant effect of language. That is, even in a data analysis that excludes only extreme naming times, we found significantly shorter naming times in the dominant language.

These findings support three key conclusions. First, although accuracy declined with decreasing frequency, it did not distinguish between the two languages and is thus a less sensitive measure than naming times, which manifested significant effects tied to both language strength and relative frequency within each language. Second, the language strength effect remains significant even with highly accurate speakers, pointing to the viability of its use to measure even subtle differences in language strength—a prerequisite for the early diagnosis of language decline. Third, the RT effect is robust enough to emerge even without coding for accuracy or complex analyses of potential errors or outliers, at least for this initial test population.

#### 4.2 Nature Term Naming Task

A nature term naming task followed the same approach as the body part task, except that 48 color photographs depicting nature terms were used as stimuli, and just seven English-dominant Korean Heritage participants have been tested to date. The current results show lower accuracy than desired for strata 2 and 3 in both languages, and marginally to significantly lower accuracy in Korean than in English across strata (overall Korean accuracy: 70% correct; English: 83%). This accuracy difference supports the dominance finding from the body part task, and demonstrates that as the naming task becomes more challenging, accuracy differences emerge. Limiting the RT analysis to a subset of 25 items that produced high accuracy in both languages (Korean: 87%, English: 98%), we replicated the effect of dominance on picture naming times: these items resulted in significantly shorter naming times in English than in Korean.

#### 4.3 Phrase-Building Task

Preliminary data for the phrase-building task comes from five

English-dominant Korean Heritage participants. As in the other tasks, participants described images presented on a computer screen. In this task images consisted of color displays of two or more objects, such as a small cat versus a large cat. Part of the image (e.g., the small cat) was circled, and participants described this portion of the image (e.g., "small cat"). The objects in this task are limited to animals and artifacts, selected to be common across a range of cultures, to provide different semantic fields than those in the other two tasks. All of the elicited phrases were limited to noun phrases, to minimize cross-linguistic differences at the clausal level (e.g. verbal agreement, tense and aspect contrasts) that might introduce confounds that mask dominance effects. The naming time from the onset of the image to the onset of speech was analyzed. The results generated high accuracy across all three strata (overall Korean accuracy: 88% correct; English: 90%). The RT results again showed a significant language dominance effect (Figure 1), with shorter naming times for English than for Korean.

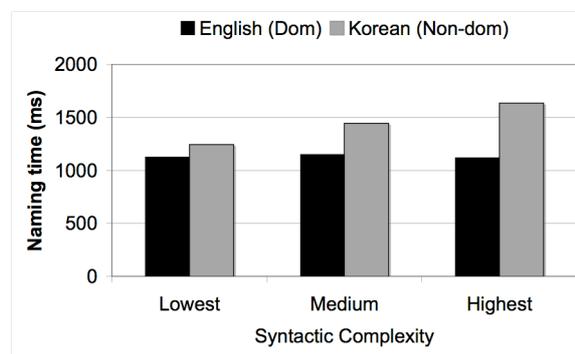


Figure 1: Naming times in the phrase-building task

In this task we were especially interested in whether reaction times would be affected by syntactic complexity. To analyze its effects, we divided the stimuli into three sets, based on the complexity of the first noun phrase in the elicited phrase: unmodified nouns (Set 1), nouns with a single modifier (Set 2), or nouns with more than one modifier (Set 3). We predicted that naming times would increase as complexity increased, primarily because of an increase in production difficulty. We further predicted that the effect of production difficulty across sets would be greater in the non-dominant language. The results from our five pilot participants show little variation in naming times within the

dominant language, but a significant effect of increasing complexity in the non-dominant language. These findings thus suggest that even in highly bilingual speakers, syntactic formulation is more difficult for the non-dominant language, and the non-dominant language is affected more strongly than the dominant language by increases in complexity. Our results therefore indicate that we can easily assess not just lexical access, as in the body part and nature term naming tasks, but the syntactic operations involved in the construction of simple phrases.

## **5. Discussion**

Overall the findings from our three pilot studies establish that small differences in language dominance can be verified with simple psycholinguistic tests that can readily be run in remote field situations. All of the tasks are simple enough to be conducted with minimal instruction and equipment, and can be used with a wide range of speakers, including children, young adults, and older adults. Our pilot data present converging evidence from three different tasks. Together they demonstrate dominance differences for multiple semantic sets of objects, and for simple object naming as well as phrase construction. Further, the pilot results establish that significant language dominance effects can emerge with as few as five speakers, making the tasks powerful enough for testing small cohorts of speakers.

## **6. Continuing research**

The HALA tasks are designed to detect statistically significant differences in language access, across a range of languages. Our next step will be to verify that we can detect dominance differences across more pairs of languages, under test conditions typical of small and endangered languages. Further, we assume that it is quite possible for a population of speakers to show significant differences in strength between two languages, without there being imminent risk to the weaker language. The results of our ongoing research will thus help us to establish expected ranges of RT differences for languages at different degrees of relative dominance. The critical question of whether a language has become or might soon become dangerously weak will ultimately depend on a range of data sources and detailed community knowledge. Nevertheless, we believe the HALA tasks can provide a useful tool to aid

researchers and communities who are concerned with relative language strength.

## **6. Conclusion**

The HALA tests comprise a new mode of language documentation that can be used to supplement and support more traditional techniques by creating a psycholinguistic record of a language's strength in the individuals and communities who speak it. This in turn opens the door for a new set of research initiatives, creating opportunities that can be expected to contribute in substantial ways to the field of language documentation and conservation. Our ongoing projects also extend psycholinguistic research to languages and cultural settings that have rarely been included in studies to date, providing a significantly broader and more general empirical base for that field.

Language endangerment is a pressing issue, for communities and speakers at risk for losing their native tongue and its intricate encoding of their culture and environment, and for scientists who stand to lose access to vital data relevant to the properties of linguistic phenomena, cognitive perspectives, and systems of specialized knowledge. Yet tragically, it is all too easy for communities to miss early signs of language loss. If the HALA project is successful, it will provide a simple method for communities to detect early shifts in language dominance, so that they can make informed choices about language use and maintenance.

## 7. References

- Allum, P. & Wheeldon, L. (2007). Planning scope in spoken sentence production: The role of grammatical units. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 33, 791-810.
- Bates, E., D'Amico, S., Jacobsen, T., Székely, A., Andonova, E., Devescovi, A., Herron, D., Lu, C.C., Pechmann, T., Pléh, C., Wicha, N., Federmeier, K., Gerdjikova, I., Gutierrez, G., Hung, D., Hsu, J., Iyer, G., Kohnert, K., Mehotcheva, T., Orozco-Figueroa, A., Tzeng, A. & Tzeng, O. (2003). Timed picture naming in seven languages. *Psychonomic Bulletin & Review* 10, 344-380.
- De Bot, K. (2004). Introduction: Special issue on language attrition. *International Journal of Bilingualism* 8, 233-237.
- Ervin, S. (1961). Learning and recall in bilinguals. *American Journal of Psychology* 74, 446-51.
- Gollan, T., Montoya, R., Cera, C. & Sandoval, T. (2008). More use almost always means a smaller frequency effect: Aging, bilingualism, and the Weaker Links Hypothesis. *Journal of Memory and Language* 58, 787-814.
- Mägiste, E. (1979). The competing language systems of the multilingual: A developmental study of decoding and encoding processes. *Journal of Verbal Learning and Verbal Behavior* 18, 79-89.
- O'Grady, W., Schafer, A.J., Perla, J., Lee, O.-S., & Wieting, J. (2009). A psycholinguistic tool for the assessment of language loss: The HALA project. *Language Documentation and Conservation*, 3, 100-112.