Problematizing the Native Speaker in Psycholinguistics: Replacing vague and harmful terminology with inclusive and accurate measures

Natasha Vernooij with Danielle Burgess, Lauretta S.P. Cheng, Ashley McDermott, Cecilia Solís-Barroso, and Savithry Namboodiripad

Psychology Methods Hour, 3/19/21
Where this came from, where it’s going (hopefully!)

Fall 2020 course

Problematizing the Native Speaker in Linguistics Research

Frontiers Special Issue

The Notion of the Native Speaker Put to the Test: Recent Research Advances

Thanks Julie!
Roadmap

- The Problem
  - Harmful
  - Vague
- Relevance to Psychology
- Recommendations
Why and how is “native speaker/signer” problematic?
“Native speaker/signer” and “nativeness” has been associated with:

- Hegemonic monolingualism (Sebba 2013)/monoglot standard (Silverstein 1996), and deficit models towards multilingualism
- Whiteness (Ramjattan 2019)
- Ideology within ELT about who is qualified to teach English (Holliday 2006)
- Ideologies of Languagelessness (Rosa 2016)
- Purism (Rhodes 2020, Namboodiripad *in press*, cf. work pushing against boundedness)
How would you define native speaker?
<table>
<thead>
<tr>
<th>Source</th>
<th>Definition of “native speaker”</th>
</tr>
</thead>
<tbody>
<tr>
<td>Benmamoun et al. (2013)</td>
<td>“A prototypical (educated) native speaker <em>lives in a monolingual environment</em>, or in a bilingual environment in which his/her original native language has not undergone attrition. Such a prototypical speaker is expected to have “native” pronunciation and a sizable, comprehensive vocabulary (about 20,000 words)” (130)</td>
</tr>
<tr>
<td>Rothman &amp; Treffers-Daller (2014)</td>
<td>“A native language is one that is acquired from naturalistic exposure, in <em>early childhood</em> and in an authentic social context/speech community” (95)</td>
</tr>
<tr>
<td>Boltova (2017)</td>
<td>“The term ‘native speaker’ has evolved to denote membership, not fluency” (22)</td>
</tr>
<tr>
<td>Abrahamsson &amp; Hyltenstam (2009)</td>
<td>A <em>native speaker</em> “(a) has <em>spoken only Swedish at home during childhood</em>; (b) has had <em>Swedish as the only language of instruction at school</em>; and (c) has <em>lived his or her whole life in a context in which Swedish has been the majority language</em>” (264)</td>
</tr>
</tbody>
</table>
“Native speaker/signer” and “nativeness” has been defined by

- Identity or allegiance
- Exposure or use of language (age of acquisition, order of acquisition, dominance, amount of exposure/use)
- Behavior (vocabulary, reaction time, signal similarity in production or processing strategies)
- Perceived nativeness (yes, you read that right)
<table>
<thead>
<tr>
<th>Operationalizations</th>
<th>Some potential problems</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Identity/Allegiance</strong> (e.g., are you a native English speaker?)</td>
<td>How is identity intertwined with the other factors that we might care about? Additionally, interactions with mother tongue, other locally relevant ideologies of how language and identity are intertwined (also, Rampton 1990, Lowe &amp; Pinner 2016). This is also where you might see participants labeling <em>themselves</em> as speaking no language at all (Rosa 2016, Namboodiripad <em>in press</em>).</td>
</tr>
<tr>
<td><strong>Exposure/use</strong> (e.g., Age of acquisition, order of acquisition, dominance, amount of exposure/use)</td>
<td>What are cutoffs based on? Do certain norms exclude entire communities (see Boltakova 2017, Costello 2008)? How does this reinforce deficit perspectives which are endemic in communities? How do measures compare (Solís-Barroso &amp; Stefanich 2019)?</td>
</tr>
<tr>
<td><strong>Behavior</strong> (e.g., vocab, reaction time, signal similarity in production/processing strategies)</td>
<td>How much homogeneity is assumed in those groups (see Abrahamsson &amp; Hyltenstam 2009)? Is variation taken into account? What types? How is are differences framed (are different processing paths to same outcome written as being a deficit)?</td>
</tr>
<tr>
<td><strong>Perceived nativeness</strong> (e.g., ratings)</td>
<td>Tied to raciolinguistic ideologies in many cases (Rosa &amp; Flores 2017), those who are categorized by researchers as native might not “pass” (e.g., Abrahamsson &amp; Hyltenstam 2009). But, there is agreement -- there seems to be face validity (But is there construct validity? What is the construct?).</td>
</tr>
</tbody>
</table>
Inconsistencies could result in problems for replicability

Recruitment: No idea how participants interpret ‘native speaker’

Variable Operationalization: Inconsistently defined (if at all) but used as a proxy for other measures

Data Analysis: No homogeneity among monolinguals anyways (Laturnus 2018, Chang 2012)
Nativeness is an ill-defined essentialist construct
In our sibling fields, there is a robust tradition and practice of theorizing and problematizing “native speaker”

- Linguistic Anthropology
- English Language Teaching
- Educational Linguistics
- Second Language Acquisition
How is this construct used in psychological methods?
<table>
<thead>
<tr>
<th>Recruitment</th>
<th>Research questions</th>
<th>Analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Shortcut to getting participants who fit a desired profile</td>
<td>Compare native to non-native speakers</td>
<td>Group native and non-native speakers, between-groups analyses</td>
</tr>
<tr>
<td></td>
<td>How do different types of language exposure affect learning or attainment? How are non-native speakers perceived? How is non-native speech processed?</td>
<td>Assumption of within-group homogeneity</td>
</tr>
</tbody>
</table>
### Recruitment

“Are you a native speaker of English?”

Who is systematically excluded? How is this term understood by participants? Does it target what researchers really care about?

### Research questions

Compare native to non-native speakers

Risks supporting deficit perspectives, can be applied without properly theorizing perceived nativeness (cf. White Listening Subject, Rosa & Flores 2017)

### Analysis

Group native and non-native speakers

Most common statistical analyses (and research Qs) assume a common “center” from which there are deviations in the group, as opposed to emergent patterns
So... what can we do?
Recommendations

Psychologists who collect language info (but don’t *study* it)

- Think about whether you need this information
- Don’t use the term, instead identify the relevant aspect of language experience
- Don’t assume homogeneity

Psychologists who study language
Overarching recommendations: Be explicit and specific

Be explicit about the types of language use/experience/allegiance which are relevant for your research questions -- and which are relevant for the social context of participants.

Operationalize your variables within the context of your formal or verbal theory.
Recommendations:

Reflection - Design - Recruitment - Analysis

- Reflect on our assumptions about ‘native speakers’
  - What does 'native speaker' mean to you?
  - What theories regarding language experience do you support?
  - How would including speakers with different language profiles affect your predictions and why?
  - What aspects of language use/experience/allegiance are important for your research questions and why?
Recommendations:
Reflection - Design - Recruitment - Analysis

- Be explicit about the types of language use/experience/allegiance which are relevant for your research questions and local context.

- Choose and define comparison groups carefully
  - avoid deficit perspectives
  - consider whether you need to divide participants into groups at all
Questions to help operationalize variables

Reflection - Design - Recruitment - Analysis

1. What aspect of language experience is important for this research question?
   a. Why is this aspect important?
   b. What predictions (if any) does it make?
   c. Would a speaker with language profile Y affect my data? How?

2. Who is excluded from my research?
   a. What acquisition and use contexts am I including and excluding?
   b. What would happen if I included them?

3. What assumptions and theories do my design explicitly and implicitly espouse?
   a. Are differences presented as deficits? Could they be interpreted as such? Do I actively stand against deficit perspectives, especially towards minoritized communities/individuals?

End Goal: What aspect of language experience is important for this experiment and what does my design imply?
Recommendations:
Reflection - Design - Recruitment - Analysis

Whether the goal is to collect data from a specific population of interest or to approximate relative homogeneity within your sample, it is important to recruit and characterize participants carefully and explicitly.
<table>
<thead>
<tr>
<th>Factor of Interest</th>
<th>Categorical Questions</th>
<th>Open-Ended/Gradient Questions</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Age of Acquisition</strong></td>
<td>Did you start learning [Language X] before [Age Y]?</td>
<td>At what age did you begin learning [Language X]?</td>
</tr>
<tr>
<td><strong>Order of Acquisition</strong></td>
<td>Is [Language X] your first language?</td>
<td>Consider probing</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- How long [Language X] was the only language used</td>
</tr>
<tr>
<td><strong>Context of Acquisition</strong></td>
<td>Did you grow up speaking [Language X] in [Region Z]?</td>
<td>May consider probing</td>
</tr>
<tr>
<td></td>
<td>Did you grow up speaking [Language X] at home/in school?</td>
<td>- presence of and/or interaction with [Language X]-speaking community networks</td>
</tr>
<tr>
<td></td>
<td>Were you exposed to [Language X] at home/in school?</td>
<td>- interest in [Language X] media</td>
</tr>
<tr>
<td><strong>Language Usage Practices</strong></td>
<td>Are you a (fluent) [Language X] speaker?</td>
<td>Self-reported or objective measures of proficiency.</td>
</tr>
<tr>
<td></td>
<td>Can you speak/read [Language X]?</td>
<td>May consider probing</td>
</tr>
<tr>
<td></td>
<td>Do you mainly use [Language X] (at home/at work/in daily life)?</td>
<td>- language use in different spheres (e.g., home, work, school)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- literacy</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- which language varieties they use</td>
</tr>
<tr>
<td><strong>Language Identity / Allegiance</strong></td>
<td>Are you a [Language X] speaker?</td>
<td>May consider probing</td>
</tr>
<tr>
<td></td>
<td>Do you consider yourself a [Language X] speaker?</td>
<td>- feelings or perceptions of pride, value, community, nationality related to using or learning [Language X]</td>
</tr>
</tbody>
</table>
Recommendations:
Reflection - Design - Recruitment - Analysis

- Collect continuous measures of language experience
- Preserve continuous data in your analysis
  - Linear models with continuous predictors
  - Individual differences analyses
  - Principal components analyses to infer what aspects of language background/experience are relevant
  - Multidimensional scaling to see groups within data
- More suggestions?

There are statistical and interpretive consequences to treating continuous variables as categorical in analyses (Young 2016) so if you are unfamiliar with regression-based statistics consider looking into it!
Nativeness is better thought of as an ideology, as opposed to an idealization:

**idealization**: assumes there is a Thing to be measured

**ideology**: when we ask about nativeness, we’re asking about ideologies of nativeness (researchers’ and participants’)

Nativeness is better thought of as an **ideology**, as opposed to an **idealization**
“Nativeness” isn’t worth it

Because of its connection to essentialist ideas about language and race/ethnicity (see deGraff 2005, Ramjattan 2019, Rosa & Flores 2017, also various discussions of the shared origin of the emergence of modern nation-states, colonialism, race science, anthropology, and linguistics), this construct has the potential to do harm if it is not consistently interrogated.

Delineations of who gets to count as a speaker of a language and who gets to have intuitions and give judgments is a (racialized) site of exclusion and harm (Namboodiripad, Hou, & Occhino 2019)

This doesn’t seem like what we’re interested in most of the time as psychologists -- so let’s treat it as an ideology that is part of speakers’ linguistic knowledge that also can influence their language use
Let’s be more ethical and conscientious researchers and move our field forward!
‘Native speaker’

1. What relevance does this term have for your research?
2. What issues do you see with the proposed solutions; what alternatives might there be?
Readings

1. Key concepts
   a. Otheguy, García, & Reid (2015): Clarifying translanguageing and deconstructing named languages: A perspective from linguistics
   b. Rosa 2016: Standardization, Racialization, Languagelessness: Raciolinguistic Ideologies across Communicative Contexts
   c. Holliday 2006: Key concepts in ELT: Native speakerism

2. Raciolinguistics
   a. Rosa & Flores 2017: Unsettling race and language: Toward a raciolinguistic perspective
   b. Ramjattan 2019: The white native speaker and inequality regimes in the private English language school
   c. Lowe & Pinner 2016: Finding the connections between native-speakerism and authenticity
Readings

3. Maturation and acquisition
   b. Costello, Fernández & Landa 2008: The non-(existent) native signer: sign language research in a small deaf population
   c. Quer & Steinbach 2019: Handling sign language data: The impact of Modality
   d. Balari & Lorenzo 2015: Should it stay or should it go? A critical reflection on the critical period for language
   e. Morford & Carlson 2011: Sign perception and recognition in Non-native signers of ASL
Readings

4. Switch dominance
   a. Benmamoun, Montrul, & Maria Polinsky 2013: Heritage languages and their speakers: Opportunities and challenges for linguistics
   b. Rothman & Treffers-Daller 2014: A Prolegomenon to the Construct of the Native Speaker: Heritage Speaker Bilinguals are Natives Too!
   c. Sorace 2004: Native language attrition and developmental instability at the syntax-discourse interface: Data, interpretations and methods
   d. Au et al. 2002: Overhearing a Language During Childhood
   e. Solís-Barroso & Stefanich 2019: Measuring Language Dominance in Early Spanish-English Bilinguals

5. Settler colonialism
   a. Meek 2016: Shrinking Indigenous Language in the Yukon
Readings

5. Settler colonialism cont.
   a. Braithwaite 2020: Ideologies of linguistic research on small sign languages in the global South: A Caribbean perspective

6. High-contact languages
   a. DeGraff 2008: Linguists' most dangerous myth: The fallacy of Creole Exceptionalism
   b. Nero 1997: English is my native language...or so I believe
   c. Namboodiripad 2020: Non-Dravidian elements and (non)diasystematic change in Malayalam

7. Theoretical approaches
   a. Thraínsson 2012: Ideal Speakers and Other Speakers: The Case of Dative and Some Other Cases
   b. Höder 2012: Multilingual constructions: A diасystematic approach to common structures
   c. Backus 2013: A usage-based approach to borrowability