

REVIEW

Reviewer A
2016-07-31 12:58 PM

For author and editor

Subject: CONSONANT SEQUENCE REDUCTION IN CHILD PHONOLOGY

A review of article entitled Consonant sequence reduction in child phonology.

The paper investigates how children acquiring Yoruba and English deal with consonant sequences, both tauto- and hetero-syllabic. The authors provide data from four children, two of whom have English as their primary language and the remaining two have English as the language they speak more often at home. Part of the conclusions in the paper are based on a comparison of these two sets of subjects.

Some of the strengths: The authors' conclusions are based on primary data. One of the subjects had a longitudinal data. The authors also compare their results with claims in the literature that are based on studies of dominantly Indo-European languages.

Some of the weaknesses

- Poor data presentation, making it difficult to link various claims and generalisations with supporting data.
- Lack of relevant data to support some generalisations, making such generalisations tentative, at best.
- Weak/unclear argumentation in some places.
- Lack of detail and clarity on data collection methodology.
- Generalisations/conclusions not supported by data in the paper.

My Recommendation:

Revise and resubmit.

Detailed page-by-page comments are shown below.

NEY is a tone language operating three tone levels and two contour tones just as in Yoruba.

This needs further discussion and motivation. Are the authors claiming that pitch is semantically contrastive in NEY just as it is in Yoruba? I notice that the authors mark tone on English words, which makes this question more important. What purpose does it serve, especially since prosodic analyses are not part of what the authors discuss in the paper?

Page 2

On this and other pages, it will be good to put sound inventories (vowels and consonants) in tables for easy viewing.

Page 3:

As for consonant clusters, it is widely reported that NEY does not permit this since Yoruba itself does not (Jowitt 1991:81-82)

Only one reference is not enough for something that is widely reported

This however does not mean that speakers are as comfortable with consonant clusters as they are with single consonants.

It is not clear what this means. Is it based on an empirical psycholinguistic study? Are speakers (especially those without linguistic training) conscious of the differences

between words with consonant clusters and those without them? Did they expressly state that they don't like words with clusters? If there is a study on this, it should be quoted

Page 4

In this article, "consonant sequence" has two possible interpretations; (i) in Yoruba, it refers to homorganic nasals followed immediately by the onset consonant of the following syllable, and (ii) in NEY it refers to clusters of voiceless obstruents and sonorants within single syllables.

What is the motivation for defining two meanings of consonant sequences based on language? Why should it mean one thing in one language and another thing in another language, especially when both languages are spoken by the same set of speakers.

I think the authors simply say that the only consonant sequence in Yoruba is a homorganic NC sequence, whereas in English, there are more possibilities, including NC sequences, tautosyllabic and heterosyllabic obstruent sequences.

While TU and JS were 3;8 and 2;4 years old respectively at the time of data collection, IY and AR were followed longitudinally. For IY, data were collected at ages 2;3-2;5, 2;10-3;00, and 3;4, whereas AR's data were collected first at age 2;1, and then at 2;5. Specifically, IY and AR were acquiring Yoruba primarily, with English being secondary.

At this point, the authors should refer the reader to Table 3 for ease of reference to the summary of the data.

"Also, at the time of data collection, IY and AR spoke English as if it were Yoruba" "...Speakers JS and TU, on the other hand, were acquiring English primarily and at the time of data collection their Yoruba was as if it were still English"

It is not clear what the authors mean here. What does it mean to speak one like language as if it were another?

Page 5

"The collection of data from Speakers JS and TU were in single sessions during which they were asked questions from a wordlist and their responses were transcribed phonetically in a notebook. In instances of uncertainty in the qualities of sounds, the researchers usually debated these and further repetitions were required before choices were made. This was done with all the four children involved, and it was partly because recordings were not made due to unstable attention spans of children (Braine 1976)"

This paragraph is loaded. It seems to me that what is contained in it needs to be explained in detail, in over one or two pages. The relevant questions that need to be answered in clarifying the content of the paragraph are:

1. 1. "...they were asked questions from a wordlist": *What is the nature of the wordlist? Where is the wordlist? How did they do it, given that we are dealing with children and two different languages? Were the questions asked in one language and answers required in another? How much data did it yield?*
2. 2. "...In instances of uncertainty in the qualities of sounds, the researchers usually debated these and further repetitions were required before choices were made": *I'm not sure if I understand the procedure given that the authors said the recordings were done in one session. Does it mean that the initial and repeated recordings were all done in one session?*
3. 3. "... and it was partly because recordings were not made due to unstable attention spans of children (Braine 1976)" *It is not clear how the unstable attention spans of children influenced the data collection methodology they adopted. Further explanation is needed here.*

"This appears to be the strategy employed by speaker JS in (4c, e-h,l,q) where he simply deletes the nasal..."

Provide evidence, data.

More data from children acquiring Yoruba however show that omitting one of the adjacent consonants is not the primary strategy.

Are the authors referring to data from their own study or from other studies? In any case, they should show the data.

“There is ample evidence that children coalesce the consonants in each cluster, but this is only apparent when the post-nasal consonant is voiceless.”

Where is the evidence? Provide examples, data.

“This is because the voiceless post-nasal consonant then becomes voiced after the coalescence.”

Why should this be the reason/explanation for the observation the authors have just made? Is this empirical?

It seems that my provide-data query applies to every sentence in the remainder of this paragraph

Page 6

After my long quest for data, I see a table at the beginning of page 6, with a long list of data meant to illustrate observation and generalisation in the preceding paragraph. This is difficult to interpret and a bit confusing. The authors need to keep their reader, who may not understand Yoruba, in mind. A better alternative method of presenting the data, which the authors should adopt is to present the relevant data after every observation, generalisation or analysis made. In that way, anything that is not relevant for that generalisation will be left out. This will be very useful for the reader in evaluating the merit of the claims and conclusions they draw, and whether they are supported by the data. As it stands now, trying to match the generalisations to the data is more like looking for a needle in a huge pile of hay. This comment applies to all the tables with data, so it will not be repeated.

Page 9

“Notice that for IY/2 (8b) /sl/ is realised as [dʒ], /str/ as [dʒ] in (8c), and /ns/ as [dʒ] in (8d). In all of these, the voicing in the output is always contributed by the deleted sonorants”

A good evidence the authors need to support their claim that obstruent voicing is due to the voicing of the deleted sonorant is sequences of two voiceless obstruents such sk, sp, st (as in the English words skip, spit, and steal). If the authors' claim is right, we expect the consonant that is left behind to remain voiceless. If the consonant left behind in each of these sequences is voiced, that will be against their claim. Unfortunately, we have not even a single word with any of these sequences, casting doubt on the conclusion drawn by the authors. The authors should add such sequences if their claim is to be maintained.

Besides supporting this conclusion, the authors need such sequences for the sake of comprehensiveness in data and analyses. All examples of words with tautosyllabic clusters are sequences of an obstruent and a liquid. There are no sm, sk, st, sp sequences, (except in the word strong, which also has a liquid in the cluster. Without these sequences, the conclusions drawn here only represent about half of the possible consonant sequences, making the conclusions very weak and tentative.

Page 10-11 (Conclusion)

The authors are arguing that consonant coalescence and consonant harmony are either rare or unattested in adult speak, and that the observation of the latter is a significant finding. I think this the claim of non-existence or rarity is an exaggeration. Is it fair to say that segmental harmony and coalescence are MORE COMMON with vowels, but they are well attested with consonants in many languages across the world. See for instance Hansson (2001, 2007) for details.

Page 11.

The authors claim about their findings: "Our findings confirm the fact that evolving systems are pressured by widely varied factors, including, but not limited to production, perception, prosodic, grammatical, as well as statistical exposure (Rose 2009, Rose and Inkelas 2011)"

This is followed by one statement in support of their claim: "Specifically, it was argued that it is statistical exposure owing to the primary language being acquired that accounted for the different strategies employed by the children in dealing with the complexity of consonant sequences and consonant clusters."

This statement thus supports their claim that "statistical exposure" is a factor in shaping an "evolving system". My question is: about "production, perception, prosodic and grammatical" factors?

My point is, this statement is exaggerated. The authors are making a claim regarding the value of their finding that goes far beyond what is supported in the paper.

References

Hansson, Gunnar O. (2001). Theoretical and typological issues in consonant harmony. PhD

dissertation, University of California, Berkeley.

Hansson, Gunnar O. (2007). On the evolution of consonant harmony: the case of secondary

articulation agreement. *Phonology* 24 77–120

Subject: CONSONANT SEQUENCE REDUCTION IN CHILD PHONOLOGY

A review of article entitled Consonant sequence reduction in child phonology.

The paper investigates how children acquiring Yoruba and English deal with consonant sequences, both tauto- and hetero-syllabic. The authors provide data from four children, two of whom have English as their primary language and the remaining two have English as the language they speak more often at home. Part of the conclusions in the paper are based on a comparison of these two sets of subjects.

Some of the strengths: The authors' conclusions are based on primary data. One of the subjects had a longitudinal data. The authors also compare their results with claims in the literature that are based on studies of dominantly Indo-European languages.

Some of the weaknesses

- Poor data presentation, making it difficult to link various claims and generalisations with supporting data.
- Lack of relevant data to support some generalisations, making such generalisations tentative, at best.
- Weak/unclear argumentation in some places.
- Lack of detail and clarity on data collection methodology.
- Generalisations/conclusions not supported by data in the paper.

My Recommendation:

Revise and resubmit.

Details page-by-page comments are shown below.

NEY is a tone language operating three tone levels and two contour tones just as in Yoruba.

Reviewer A
2016-07-31 12:58 PM

For editor

This needs further discussion and motivation. Are the authors claiming that pitch is semantically contrastive in NEY just as it is in Yoruba? I notice that the authors mark tone on English words, which makes this question more important. What purpose does it serve, especially since prosodic analyses are not part of what the authors discuss in the paper?

Page 2

On this and other pages, it will be good to put sound inventories (vowels and consonants) in tables for easy viewing.

Page 3:

As for consonant clusters, it is widely reported that NEY does not permit this since Yoruba itself does not (Jowitt 1991:81-82)

Only one reference is not enough for something that is widely reported

This however does not mean that speakers are as comfortable with consonant clusters as they are with single consonants.

It is not clear what this means. Is it based on an empirical psycholinguistic study? Are speakers (especially those without linguistic training) conscious of the differences between words with consonant clusters and those without them? Did they expressly state that they don't like words with clusters? If there is a study on this, it should be quoted

Page 4

In this article, "consonant sequence" has two possible interpretations; (i) in Yoruba, it refers to homorganic nasals followed immediately by the onset consonant of the following syllable, and (ii) in NEY it refers to clusters of voiceless obstruents and sonorants within single syllables.

What is the motivation for defining two meanings of consonant sequences based on language? Why should it mean one thing in one language and another thing in another language, especially when both languages are spoken by the same set of speakers.

I think the authors simply say that the only consonant sequence in Yoruba is a homorganic NC sequence, whereas in English, there are more possibilities, including NC sequences, tautosyllabic and heterosyllabic obstruent sequences.

While TU and JS were 3;8 and 2;4 years old respectively at the time of data collection, IY and AR were followed longitudinally. For IY, data were collected at ages 2;3-2;5, 2;10-3;00, and 3;4, whereas AR's data were collected first at age 2;1, and then at 2;5. Specifically, IY and AR were acquiring Yoruba primarily, with English being secondary.

At this point, the authors should refer the reader to Table 3 for ease of reference to the summary of the data.

"Also, at the time of data collection, IY and AR spoke English as if it were Yoruba"
 "...Speakers JS and TU, on the other hand, were acquiring English primarily and at the time of data collection their Yoruba was as if it were still English"

It is not clear what the authors mean here. What does it mean to speak one like language as if it were another?

Page 5

"The collection of data from Speakers JS and TU were in single sessions during which they were asked questions from a wordlist and their responses were transcribed phonetically in a notebook. In instances of uncertainty in the qualities of sounds, the researchers usually debated these and further repetitions were required before choices were made. This was done with all the four children involved, and it was partly because recordings were not made due to unstable attention spans of children (Braine 1976)"

This paragraph is loaded. It seems to me that what is contained in it needs to be explained in detail, in over one or two pages. The relevant questions that need to be answered in clarifying the content of the paragraph are:.

1. 1. "...they were asked questions from a wordlist": *What is the nature of the wordlist? Where is the wordlist? How did they do it, given that we are dealing with children and two different languages? Were the questions asked in one language and answers required in another? How much data did it yield?*
2. 2. "...In instances of uncertainty in the qualities of sounds, the researchers usually debated these and further repetitions were required before choices were made": *I'm not sure if I understand the procedure given that the authors said the recordings were done in one session. Does it mean that the initial and repeated recordings were all done in one session?*
3. 3. "... and it was partly because recordings were not made due to unstable attention spans of children (Braine 1976)" *It is not clear how the unstable attention spans of children influenced the data collection methodology they adopted. Further explanation is needed here.*

"This appears to be the strategy employed by speaker JS in (4c, e-h,l,q) where he simply deletes the nasal..."

Provide evidence, data.

More data from children acquiring Yoruba however show that omitting one of the adjacent consonants is not the primary strategy.

Are the authors referring to data from their own study or from other studies? In any case, they should show the data.

"There is ample evidence that children coalesce the consonants in each cluster, but this is only apparent when the post-nasal consonant is voiceless."

Where is the evidence? Provide examples, data.

"This is because the voiceless post-nasal consonant then becomes voiced after the coalescence."

Why should this be the reason/explanation for the observation the authors have just made? Is this empirical?

It seems that my provide-data query applies to every sentence in the remainder of this paragraph

Page 6

After my long quest for data, I see a table at the beginning of page 6, with a long list of data meant to illustrate observation and generalisation in the preceding paragraph. This is difficult to interpret and a bit confusing. The authors need to keep their reader, who may not understand Yoruba, in mind. A better alternative method of presenting the data, which the authors should adopt is to present the relevant data after every observation, generalisation or analysis made. In that way, anything that is not relevant for that generalisation will be left out. This will be very useful for the reader in evaluating the merit of the claims and conclusions they draw, and whether they are supported by the data. As it stands now, trying to match the generalisations to the data is more like looking for a needle in a huge pile of hay. This comment applies to all the tables with data, so it will not be repeated.

Page 9

"Notice that for IY/2 (8b) /sl/ is realised as [dʒ], /str/ as [dʒ] in (8c), and /ns/ as [dʒ] in (8d). In all of these, the voicing in the output is always contributed by the deleted sonorants"

A good evidence the authors need to support their claim that obstruent voicing is due to the voicing of the deleted sonorant is sequences of two voiceless obstruents such sk,

sp, st (as in the English words skip, spit, and steal). If the authors' claim is right, we expect the consonant that is left behind to remain voiceless. If the consonant left behind in each of these sequences is voiced, that will be against their claim. Unfortunately, we have not even a single word with any of these sequences, casting doubt on the conclusion drawn by the authors. The authors should add such sequences if their claim is to be maintained.

Besides supporting this conclusion, the authors need such sequences for the sake of comprehensiveness in data and analyses. All examples of words with tautosyllabic clusters are sequences of an obstruent and a liquid. There are no sm, sk, st, sp sequences, (except in the word strong, which also has a liquid in the cluster. Without these sequences, the conclusions drawn here only represent about half of the possible consonant sequences, making the conclusions very weak and tentative.

Page 10-11 (Conclusion)

The authors are arguing that consonant coalescence and consonant harmony are either rare or unattested in adult speak, and that the observation of the latter is a significant finding. I think this the claim of non-existence or rarity is an exaggeration. Is it fair to say that segmental harmony and coalescence are MORE COMMON with vowels, but they are well attested with consonants in many languages across the world. See for instance Hansson (2001, 2007) for details.

Page 11.

The authors claim about their findings: "Our findings confirm the fact that evolving systems are pressured by widely varied factors, including, but not limited to production, perception, prosodic, grammatical, as well as statistical exposure (Rose 2009, Rose and Inkelas 2011)"

This is followed by one statement in support of their claim: "Specifically, it was argued that it is statistical exposure owing to the primary language being acquired that accounted for the different strategies employed by the children in dealing with the complexity of consonant sequences and consonant clusters."

This statement thus supports their claim that "statistical exposure" is a factor in shaping an "evolving system". My question is: about "production, perception, prosodic and grammatical" factors?

My point is, this statement is exaggerated. The authors are making a claim regarding the value of their finding that goes far beyond what is supported in the paper.

References

Hansson, Gunnar O. (2001). Theoretical and typological issues in consonant harmony. PhD

dissertation, University of California, Berkeley.

Hansson, Gunnar O. (2007). On the evolution of consonant harmony: the case of secondary

articulation agreement. *Phonology* 24 77–120

Close