Language and speech are old
– going back half a million years or so, and the Neandertals spoke too –

Neandertal man

Signal of genetic admixture

Dan Dediu

Labex EFL International Chair (4th edition)
February, 2016
Paris, France

Language and Genetics
Genetic Biases in Language and Speech
Max Planck Institute for Psycholinguistics
Nijmegen, The Netherlands
Overview

On the antiquity of language: the reinterpretation of Neandertal linguistic capacities and its consequences

Dan Dediu1,2* and Stephen C. Levinson2,3†
Background

H. erectus

H. heidelbergensis

Neandertals

Denisovans

PNG, Australia...

Eurasia...

Outside Africa

Modern humans

~0.5

~0.25

0 mya

~0.03


Background

Background

- more complex patterns
- modern humans with recent Neandertal admixture: Ust’-Ishim (45kya), Peștera cu Oase (40kya)

Background

- more complex patterns
- modern humans with recent Neandertal admixture: Ust’-Ishim (45kya), Peștera cu Oase (40kya)
- average ~23 Mb Neandertal DNA per individual, total ~15 Gb ≈ 20% of Neandertal genome
- apparently at least two episodes of admixture
Background

- more complex patterns
- modern humans with recent Neandertal admixture: Ust’-Ishim (45kya), Peștera cu Oase (40kya)
- average ~23 Mb Neandertal DNA per individual, total ~15 Gb ≈ 20% of Neandertal genome
- apparently at least two episodes of admixture
- negative selection against introgressed Neandertal DNA (X chr., less around genes)?
  → smaller effective populations, high inbreeding in Neanderals

Background

- more complex patterns
- modern humans with recent Neandertal admixture: Ust’-Ishim (45kya), Peštera cu Oase (40kya)
- average ~ 23 Mb Neandertal DNA per individual, total ~15 Gb ≈ 20% of Neandertal genome
- apparently at least two episodes of admixture
- negative selection against introgressed Neandertal DNA (X chr., less around genes)?
  → smaller effective populations, high inbreeding in Neanderals

- introgressed genes (Neandertal & Denisovan):
  ◦ positive selection on skin & hair alleles
  ◦ immune system (Toll-like receptors)
  ◦ “negative” (autoimmune, diabetes, addictions)
  ◦ adaptation to low oxygen (Tibet)
  ◦ adaptation to cold (Greenland Inuit)
  …

Broad models for language origins

2   1   ~0.5   0 mya

H. erectus

“Proto-language”

Neandertals

Denisovans

H. heidelbergensis

Modern language

Modern humans

The “standard” view: language emerged recently and abruptly


Broad models for language origins

The “standard” view:
language emerged recently and abruptly

H. erectus
“Proto-language”

Neandertals

Denisovans

Modern language

Modern humans

“Modern package”/“Modern human revolution”


Broad models for language origins

The “alternative” view: language emerged early and gradually

Compare the proposals, but...

there is a heated debate raging
Some strands of evidence

Ancient and modern DNA

Admixture → Same biological species

- hybridization
- multiple definitions of “species”

→ BSC is what people think of
→ on the verge of speciation


Some strands of evidence

Ancient and modern DNA

- Abrigo do Lagar Vehlo (P)
- Peștera cu Oase (RO)
- Peștera Muierii (RO)
- Mladeč (CZ)
- Riparo Mezzena (IT)...

Admixture

Same biological species

Contested hybrid fossils


Some strands of evidence

Some strands of evidence

Ancient and modern DNA

Admixture

Same biological species

Contested hybrid fossils

Some strands of evidence

Ancient and modern DNA

Admixture

Same biological species

Contested hybrid fossils


Some strands of evidence

Ancient and modern DNA

Admixture

Same biological species

Contested hybrid fossils

Some strands of evidence

Developmental Verbal Dyspraxia (DVD); OMIM 602081
Some strands of evidence

Multiple phenotypes:

Ancient and modern DNA
- Same biological species
- Contested hybrid fossils


Some strands of evidence

Ancient and modern DNA

Admixture

Same biological species

Contested hybrid fossils

Multiple phenotypes:


Some strands of evidence

Ancient and modern DNA

Admixture

Same biological species

Contested hybrid fossils

Multiple phenotypes:


Some strands of evidence

Multiple phenotypes:

Ancient and modern DNA

Admixture

Same biological species

Contested hybrid fossils


Some strands of evidence

Multiple phenotypes:

Ancient and modern DNA


*FOXP2* is not “the” gene “for” language and speech
Some strands of evidence

Ancient and modern DNA
Admixture
Same biological species
Contested hybrid fossils


Some strands of evidence

Figure 2. Silent and replacement nucleotide substitutions mapped on a phylogeny of primates. Bars represent nucleotide changes. Grey bars indicate amino-acid changes.

Some strands of evidence

Ancient and modern DNA

Admixture

→ Same biological species

Contested hybrid fossils

+ two “human-specific” AAs on same haplotype

→ intron 8 (POU3F2 binding)

→ ancestral allele: Africa

→ molecular activity unclear

Some strands of evidence

- downregulated \(\downarrow\) by *FOXP2*
- neurexin family
- involved in autism, SLI, normal language variation, language development...

Ancient and modern DNA

Admixture

→ Same biological species

Contested hybrid fossils

+ two “human-specific” AAs on same haplotype
  - intron 8 (POU3F2 binding)
  → ancestral allele: Africa
  → molecular activity unclear

*CNTNAP2*

Some strands of evidence

Ancient and modern DNA

Admixture

Same biological species

Contested hybrid fossils

**FOXP2**

- + two “human-specific” AAs on same haplotype
  - intron 8 (POU3F2 binding)
  - → ancestral allele: Africa
  - → molecular activity unclear
- coding change (Ile → Val)
  - → functional?

**CNTNAP2**


Some strands of evidence

Ancient and modern DNA

Admixture

– Same biological species
– Contested hybrid fossils

FOXP2
– + two “human-specific” AAs on same haplotype
– intron 8 (POU3F2 binding)
– ancestral allele: Africa
– molecular activity unclear

CNTNAP2
– coding change (Ile → Val)
– functional?

MEF2A
– developmental plasticity?

Some strands of evidence

Some strands of evidence

Anatomy & development

Partly due to lifestyle differences

Some strands of evidence

Anatomy & development

Partly due to lifestyle differences

Birth canal

Dental eruption

MEF2A ~0.5mya

Prolonged childhood


Some strands of evidence

Anatomy & development

Partly due to lifestyle differences

Birth canal

MEF2A ~0.5mya

Dental eruption

Prolonged childhood

comparable brain size

Some strands of evidence

Anatomy & development

Partly due to lifestyle differences

Birth canal

Dental eruption

MEF2A ~0.5mya

Prolonged childhood

comparable brain size

larger eyes

occipital bun

Cognition?

Neural organization?

Some strands of evidence

Some strands of evidence

Vocal production & perception

Tuning production ↔ perception

Some strands of evidence

Vocal production & perception

Tuning production ↔ perception

Audiograms


Some strands of evidence

Vocal production & perception

Tuning production ↔ perception

Audiograms

Ear ossicles

Incus:

Modern humans

Neandertal

Chimp

P. robustus

Early modern humans

Some strands of evidence

Vocal production & perception

Tuning production ↔ perception

Audiograms

Ear ossicles

Hyoid bone
Some strands of evidence

Vocal production & perception

Tuning production ↔ perception

Audiograms

Ear ossicles

Hyoid bone

Some strands of evidence

Vocal production & perception

Tuning production ↔ perception
Audiograms
Ear ossicles
Hyoid bone

Some strands of evidence

Vocal production & perception

Tuning production ↔ perception

Audiograms

Ear ossicles

Hyoid bone


Some strands of evidence

Vocal production & perception

Tuning production ↔ perception

Audiograms

Ear ossicles

Hyoid bone

Breathing control

Some strands of evidence

Vocal production & perception

Tuning production ↔ perception

Audiograms

Ear ossicles

Hyoid bone

Breathing control

All for singing?

Some strands of evidence

Some strands of evidence

Symbolic behaviour

Extremely difficult and controversial

“Modern human revolution”

“a rewiring of the brain took place in some individual, call him Prometheus, yielding the operation of unbounded Merge, applying to concepts with intricate (and little understood) properties…” (Chomsky, 2010)

Some strands of evidence

Symbolic behaviour

Extremely difficult and controversial

“Modern human revolution”

Some strands of evidence

Symbolic behaviour

Extremely difficult and controversial

“Modern human revolution”


Some strands of evidence

Symbolic behaviour

Extremely difficult and controversial

“Modern human revolution”

Complex toolkit (Mousterian): bone, wood, strings...


Some strands of evidence

Symbolic behaviour

Extremely difficult and controversial

“Modern human revolution”

Complex toolkit (Mousterian): bone, wood, strings...

Intentional burial, grave offerings

Sick & infirm, medicinal plants


Some strands of evidence

Symbolic behaviour

Extremely difficult and controversial

“Modern human revolution”

Complex toolkit (Mousterian): bone, wood, strings...

Intentional burial, grave offerings

Sick & infirm, medicinal plants

Body ornaments: ocher, beads, art


Some strands of evidence

Symbolic behaviour

Extremely difficult and controversial

“Modern human revolution”

Complex toolkit (Mousterian): bone, wood, strings...

Intentional burial, grave offerings

Sick & infirm, medicinal plants

Body ornaments: ocher, beads, art

Archaeological absence ≠ absence of capacity!


Some strands of evidence

Symbolic behaviour

Extremely difficult and controversial

“Modern human revolution”

Complex toolkit (Mousterian): bone, wood, strings...

Intentional burial, grave offerings

Sick & infirm, medicinal plants

Body ornaments: ocher, beads, art

Archaeological absence ≠ absence of capacity!

Handedness


Some strands of evidence

Some strands of evidence


If language & speech are old...

Recognizably modern speech and language (not “proto-language”)
~ 0.5mya

If language & speech are old...

Recognizably modern speech and language (not “proto-language”) ~ 0.5mya

Limited genetic contact:


Cultural contact:
... what can we say about them?

Language:
- small communities $\rightarrow$ **complex** languages?

... what can we say about them?

Language:

- small communities → complex languages?
- ASPM, MCPH1

D. Robert Ladd


... what can we say about them?

Language:
- small communities → complex languages?
- ASPM, MCPH1
... what can we say about them?

Language:

- small communities → complex languages?
- ASPM, MCPH1 → tone languages?

... what can we say about them?

Language:

- small communities → **complex** languages?
- *ASPM, MCPHI* → **tone** languages?
- **linguistic contact** (borrowing)?


Seán Roberts  
Balthasar Bickel
Language contact?

Possible scenarios:

Language contact?

Possible scenarios:
1. Language extinction


Language contact?

Possible scenarios:
1. Language extinction
2. Language shift
**Language contact?**

Possible scenarios:
1. Language extinction
2. Language shift
3. Pidginization


Possible scenarios:

1. Language extinction
2. Language shift
3. Pidginization
4. Borrowing


Language contact?

Linguistic borrowing → **structural** differences between African/non-African (and possibly Australian+Papuan/non-AuPNG) languages

---

Why would all this matter?

It isn't anymore of question of *if*,
but a question of what were the (small) *differences*
and of what do they *mean* for:

- *language evolution*, and
- *present-day linguistic diversity*.

Only *one experiment* in getting and having language: us $\rightarrow$ very hard to make inferences...

But if a relatively independent *second experiment* (Neandertals) $\rightarrow$ now that's exciting!
Why would all this matter?

Language sciences:

- recent sudden origins model is dominant
- language as monolithic and special with
  Merge (recursion) as the “secret ingredient”
→ reinforced by and reinforcing a speciation model for modern humans

Why would all this matter?

Language sciences:
- recent sudden origins model is dominant
- language as monolithic and special with Merge (recursion) as the “secret ingredient”
→ reinforced by and reinforcing a speciation model for modern humans
- but this seems implausible given evolutionary theory and genetics of language

Why would all this matter?

Linguistic diversity:
- exploration of the design space

Why would all this matter?

Linguistic diversity:

- exploration of the design space
- was there enough time? → rates of language evolution (structural stability)
  - most features seem pretty labile
  - but some seem very stable (e.g., parts of basic vocabulary, structural features)

Why would all this matter?

Linguistic diversity:

- exploration of the design space
- was there enough time? → rates of language evolution (structural stability)
  - most features seem pretty labile
  - but some seem very stable (e.g., parts of basic vocabulary, structural features)


Why would all this matter?

Linguistic diversity:
- exploration of the design space
- was there enough time? → rates of language evolution (structural stability)
  - most features seem pretty labile
  - but some seem very stable (e.g., parts of basic vocabulary, structural features)
→ Neandertal languages – independent exploration of the (same?) design space
  – injected diversity into modern human languages?

Language contact?

Linguistic borrowing → structural differences between African/non-African (and possibly Australian+Papuan/non-AuPNG) languages

Potential issues:

0. Neandertals did not have language/it was too different from ours/there was no language contact

Language contact?

Linguistic borrowing → structural differences between African/non-African (and possibly Australian+Papuan/non-AuPNG) languages

Potential issues:

0. Neandertals did not have language/it was too different from ours/there was no language contact

1. Too faint/too much time has passed for the signal to survive to the present day


Language contact?

Linguistic borrowing → structural differences between African/non-African (and possibly Australian+Papuan/non-AuPNG) languages

Potential issues:

0. Neandertals did not have language/it was too different from ours/there was no language contact

1. Too faint/too much time has passed for the signal to survive to the present day

2. Too many large-scale demographic and linguistic events → erasure of the signal

Language contact?

Linguistic borrowing → **structural differences** between African/non-African (and possibly Australian+Papuan/non-AuPNG) languages

Potential issues:

0. Neandertals did not have language/it was too different from ours/there was no language contact

1. **Too faint/too much time** has passed for the signal to survive to the present day

2. Too many large-scale demographic and linguistic events → **erasure of the signal**

3. The signal we detect is **not the correct one** (i.e., much too recent, different process)


Linguistic borrowing $\rightarrow$ structural differences between African/non-African (and possibly Australian+Papuan/non-AuPNG) languages

Potential issues:

0. Neandertals did not have language/it was too different from ours/there was no language contact

1. Too faint/too much time has passed for the signal to survive to the present day

2. Too many large-scale demographic and linguistic events $\rightarrow$ erasure of the signal

3. The signal we detect is not the correct one (i.e., much too recent, different process)

4. The signal might be there, but we don't have the methods to extract it
Language contact?

Linguistic borrowing → **structural differences** between African/non-African (and possibly Australian+Papuan/non-AuPNG) languages

Potential issues:

0. Neandertals did not have language/it was too different from ours/there was no language contact
1. Too faint/too much time has passed for the signal to survive to the present day
2. Too many large-scale demographic and linguistic events → erasure of the signal
3. The signal we detect is not the correct one (i.e., much too recent, different process)
4. The signal might be there, but we don't have the methods to extract it
5. The signal might be there, but we don't have the data
Language contact?

Linguistic borrowing → **structural differences** between African/non-African (and possibly Australian+Papuan/non-AuPNG) languages

Potential issues:

0. Neandertals did not have language/it was too different from ours/there was no language contact
1. Too faint/too much time has passed for the signal to survive to the present day
2. Too many large-scale demographic and linguistic events → **erasure** of the signal
3. The signal we detect is **not** the correct one (i.e., much too recent, different process)
4. The signal might be there, but we **don't** have the methods to extract it
5. The signal might be there, but we **don't** have the data

So... do we just give up?
Language contact?

Linguistic borrowing → structural differences between African/non-African (and possibly Australian+Papuan/non-AuPNG) languages

Principles:

- we don't expect the vocabulary (i.e., cognacy) to survive that long...
- … but structural (typological) properties might
- we don't expect individual features to carry enough signal...
- … but multivariate patterns might (like in human population genetics)
- we expect the signal to be extremely faint, swamped by competing signals/noise
- thus we need to use as much data as possible (even of unclear quality)...
- … and very powerful statistical methods
- given the weakness of the signal, a false negative is costlier than a false positive

So, to work!

Language contact?

Linguistic borrowing → **structural differences** between **African/non-African** (and possibly **Australian+Papuan/non-AuPNG**) languages


Language contact?

Linguistic borrowing → **structural differences** between African/non-African (and possibly Australian+Papuan/non-AuPNG) languages


Language contact?

Linguistic borrowing → **structural differences** between African/non-African (and possibly Australian+Papuan/non-AuPNG) languages

**K-means** on feature values
- languages weighted by relatedness
- correlated features removed
→ % correctly classified
→ F-score

**Permutation tests:**
- full (cell-wise)
- languages in sets
- families in sets
- families, equal size

**Strength of classification**

**Distinctive feature patterns**
→ % correctly classified
→ F-score
+ proportion better than permuted


Language contact?

Linguistic borrowing → **structural differences between** African/non-African (and possibly Australian+Papuan/non-AuPNG) languages

- MDS on feature values → 2 dimensions → distances b/w languages
- Permutation tests: - language-wise
- Strength of classification
- Distinctive feature patterns → distinction b/w continents


Language contact?

Linguistic borrowing → **structural differences** between African/non-African (and possibly Australian+Papuan/non-AuPNG) languages


Linguistic borrowing → **structural differences** between African/non-African (and possibly Australian+Papuan/non-AuPNG) languages

**STRUCTURE** on binarized features
- ± geographic priors (AUTOTYP)
- word order features

Strength of classification
Distinctive feature patterns → distinction b/w continents


Language contact?

Linguistic borrowing → **structural differences between African/non-African** (and possibly **Australian+Papuan/non-AuPNG**) languages

Conclusions

Thus,

- Language and speech are **old**, ~0.5 mya
- **Two experiments** in having language might help better understanding “our” language
- But it is **extremely hard** to infer something from the present-day linguistic diversity
  - A lot of **noise**/overlapping processes
  - A lot of **information loss**
  - A lot of **missing data**

Nevertheless worth trying:

- **Method development/testing**
- Identification of **patterns of linguistic diversity**

Thank you!

**Acknowledgments**: Steve Levinson, Bob Ladd, Simon Fisher, Didier Demolin, Bart de Boer, Sean Roberts, Asifa Majid, Mark Dingemanse, Sonja Vernes, Harald Hammarström, Kenny Smith, Heriberto Avelino, Scott Moisik, Rick Janssen & the Language and Genetics department

**Funded by Netherlands Organisation for Scientific Research (NWO) VIDI grant 276-70-022**