

Empirisch bewijs voor universals

Milan Lopuhaä

27 februari 2014

Universals

- ▶ *(Absolute) universal*: eigenschap die alle natuurlijke talen hebben.

Universals

- ▶ *(Absolute) universal*: eigenschap die alle natuurlijke talen hebben.
- ▶ Voorbeeld: “Heeft medeklinkerfonemen”

Universals

- ▶ (*Absolute*) *universal*: eigenschap die alle natuurlijke talen hebben.
- ▶ Voorbeeld: “Heeft medeklinkerfonemen”
- ▶ Maar: mensen hebben tegenvoorbeelden geopperd voor “heeft klinkerfonemen” ! (Kabardisch)

Hoe vinden we bewijs voor universals?

- ▶ Het is onmogelijk alle levende talen te bekijken, laat staan alle dode talen!

Hoe vinden we bewijs voor universals?

- ▶ Het is onmogelijk alle levende talen te bekijken, laat staan alle dode talen!
- ▶ Daarom wordt vaak gebruik gemaakt van *surveys*: je selecteert een verzameling talen en kijkt of al deze talen voldoen aan je universal.

Hoe vinden we bewijs voor universals?

- ▶ Het is onmogelijk alle levende talen te bekijken, laat staan alle dode talen!
- ▶ Daarom wordt vaak gebruik gemaakt van *surveys*: je selecteert een verzameling talen en kijkt of al deze talen voldoen aan je universal.
- ▶ Zo'n survey moet veel talen bevatten en zowel genetisch als geografisch veel diversiteit bevatten.

Hoofdvraag

Hoeveel talen zijn er nodig om een universal aan te tonen?

Hoofdvraag

Hoeveel talen zijn er nodig om een universal aan te tonen?

- ▶ Absolute zekerheid krijgen we niet.

Hoofdvraag

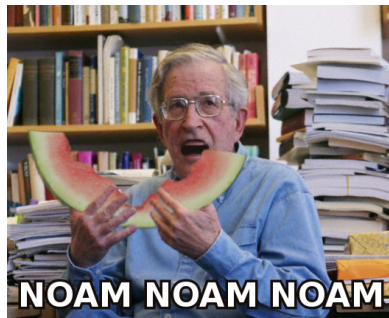
Hoeveel talen zijn er nodig om een universal aan te tonen?

- ▶ Absolute zekerheid krijgen we niet.
- ▶ Met statistiek kunnen we wel een schatting maken van de onzekerheid.

Structuuruniversals vs oppervlakteuniversals

Structuuruniversals

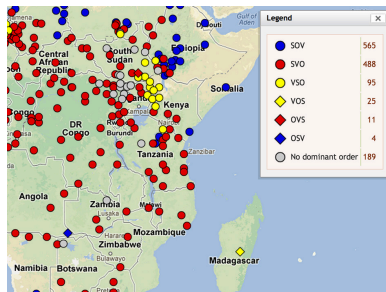
- ▶ Uitspraken over de diepe structuur van een taal
- ▶ Voorbeeld: lijdend voorwerp is *onderliggend* ondergeschikt aan werkwoord (in UG)
- ▶ Nadeel: uitkomst afhankelijk van analyse
- ▶ Nadeel: diepgaande analyse is maar voor relatief weinig talen gedaan



Structuuruniversals vs oppervlakteuniversals

Oppervlakteuniversals

- ▶ Uitspraken over de oppervlaktestructuur van een taal
- ▶ Voorbeeld: dominante zinsvolgorde is SOV
- ▶ Nadeel: gaat voorbij aan wat er echt aan de hand is
- ▶ Voordeel: veel data



Hoeveel talen zijn genoeg?

- ▶ *World Atlas of Language Structures*: surveys van 192 features in 2679 talen, maar lang niet allemaal compleet.

Hoeveel talen zijn genoeg?

- ▶ *World Atlas of Language Structures*: surveys van 192 features in 2679 talen, maar lang niet allemaal compleet.
- ▶ Voorbeeld: voor de eigenschap “fixed stress locations” zijn er 502 talen onderzocht, met de volgende uitslag:

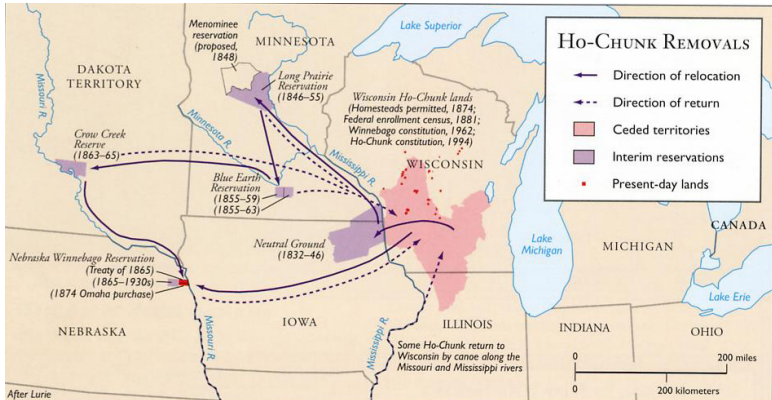
Hoeveel talen zijn genoeg?

- ▶ *World Atlas of Language Structures*: surveys van 192 features in 2679 talen, maar lang niet allemaal compleet.
- ▶ Voorbeeld: voor de eigenschap “fixed stress locations” zijn er 502 talen onderzocht, met de volgende uitslag:

geen vaste klemtoon	220
eerste	92
tweede	16
derde	1
voorvoorlaatste	12
voorlaatste	110
laatste	51

Winnebago (Hocąk)

- ▶ Enige taal met vaste klemtoon op 3^e lettergreep
- ▶ Centraal Sioux, Wisconsin/Minnesota



Assiniboine (Nakóda)

- ▶ Fonemisch accent
- ▶ Centraal Siouxs, Saskatchewan



Hoeveel talen zijn genoeg?

- ▶ Winnebago en Assiniboine verschillen in hun klemtoon

Hoeveel talen zijn genoeg?

- ▶ Winnebago en Assiniboine verschillen in hun klemtoon
- ▶ Als het survey was uitgevoerd met Assiniboine in plaats van Winnebago, was het geografisch en genetisch even divers

Hoeveel talen zijn genoeg?

- ▶ Winnebago en Assiniboine verschillen in hun klemtoon
- ▶ Als het survey was uitgevoerd met Assiniboine in plaats van Winnebago, was het geografisch en genetisch even divers
- ▶ Maar er zou uitkomen dat “Heeft geen vaste klemtoon op de derde lettergreep” een universal is!

Hoeveel talen zijn genoeg?

- ▶ Winnebago en Assiniboine verschillen in hun klemtoon
- ▶ Als het survey was uitgevoerd met Assiniboine in plaats van Winnebago, was het geografisch en genetisch even divers
- ▶ Maar er zou uitkomen dat “Heeft geen vaste klemtoon op de derde lettergreep” een universal is!
- ▶ Zijn 502 talen dan wel voldoende?

Maak je eigen universal

Implicatieve universal: uitspraak van de vorm “Elke taal met X heeft Y ”.

Maak je eigen universal

Implicatieve universal: uitspraak van de vorm “Elke taal met X heeft Y ”.

Neem nu X heel zeldzaam, en Y heel frequent \Rightarrow ‘Universal’!

- ▶ “Elke taal met een vaste klemtoon op de 3e lettergreep heeft geen velaire nasaal” (174 talen)
- ▶ “Elke taal met VOS heeft bilabialen” (381 talen)
- ▶ “Elke taal waarin meervoud met toon gemarkeerd wordt heeft alleen (C)V” (297 talen)

Maak je eigen universal

Implicationale universal: uitspraak van de vorm “Elke taal met X heeft Y ”.

Neem nu X heel zeldzaam, en Y heel frequent \Rightarrow ‘Universal’!

- ▶ “Elke taal met een vaste klemtoon op de 3e lettergreep heeft geen velaire nasaal” (174 talen)
- ▶ “Elke taal met VOS heeft bilabialen” (381 talen)
- ▶ “Elke taal waarin meervoud met toon gemarkeerd wordt heeft alleen (C)V” (297 talen)

Waarom: tegenvoorbeeld heeft wel X , niet Y , en *zeldzaam* \times *zeldzaam* = *heel zeldzaam*.

Statistiek

Situatie: een eigenschap P waarvan we willen weten of hij universeel is. Verder hebben we een survey van N talen.

Statistiek

Situatie: een eigenschap P waarvan we willen weten of hij universeel is. Verder hebben we een survey van N talen.

Twee mogelijke hypotheses:

- ▶ H_0 : P is niet universeel;
- ▶ H_1 : P is wel universeel.

Statistiek

Situatie: een eigenschap P waarvan we willen weten of hij universeel is. Verder hebben we een survey van N talen.

Twee mogelijke hypotheses:

- ▶ H_0 : P is niet universeel;
- ▶ H_1 : P is wel universeel.

We nemen H_0 aan als we minstens 1 taal in onze survey hebben zonder eigenschap P , anders nemen we H_1 aan.

Statistiek

	H_0 aannemen	H_0 verwerpen
H_0 waar	☺	type II
H_1 onwaar	(type I)	☺

We willen de kans op een type-II-fout (p -waarde) zo klein mogelijk hebben.

Statistiek

	H_0 aannemen	H_0 verwerpen
H_0 waar	☺	type II
H_1 onwaar	(type I)	☺

We willen de kans op een type-II-fout (p -waarde) zo klein mogelijk hebben.

Vaak gebruikte grens: $p = 0,05$.

Model voor niet-universals

- ▶ De p -waarde is de kans dat we elke taal in onze survey eigenschap P heeft, terwijl P niet universeel is.

Model voor niet-universals

- ▶ De p -waarde is de kans dat we elke taal in onze survey eigenschap P heeft, terwijl P niet universeel is.
- ▶ Om deze kans te berekenen moeten we een model hebben voor de kansverdeling van P over de survey.

Model voor niet-universals

- ▶ De p -waarde is de kans dat we elke taal in onze survey eigenschap P heeft, terwijl P niet universeel is.
- ▶ Om deze kans te berekenen moeten we een model hebben voor de kansverdeling van P over de survey.
- ▶ In de literatuur wordt hier nauwelijks aandacht aan besteed!

Makkelijkste model: *binomiaal model*

- ▶ Elke taal in de survey heeft kans ϑ om eigenschap P te hebben;

Makkelijkste model: *binomiaal model*

- ▶ Elke taal in de survey heeft kans ϑ om eigenschap P te hebben;
- ▶ Alle talen gedragen zich onafhankelijk;

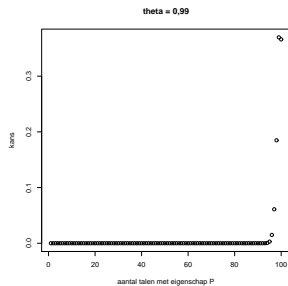
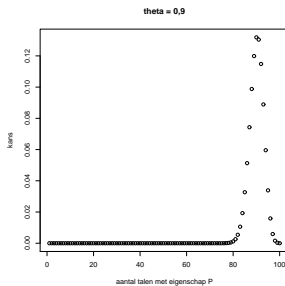
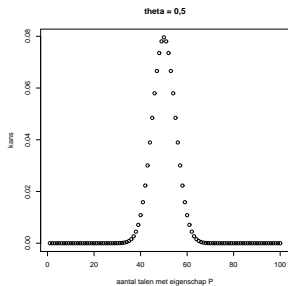
Makkelijkste model: *binomiaal model*

- ▶ Elke taal in de survey heeft kans ϑ om eigenschap P te hebben;
- ▶ Alle talen gedragen zich onafhankelijk;
- ▶ Geen rekening gehouden met taalcontact of genetische relaties.

Makkelijkste model: *binomiaal model*

- ▶ Elke taal in de survey heeft kans ϑ om eigenschap P te hebben;
- ▶ Alle talen gedragen zich onafhankelijk;
- ▶ Geen rekening gehouden met taalcontact of genetische relaties.
- ▶ Maar het is wel makkelijk rekenen!

Voorbeelden



Fisher's exact test

Kan gebruikt worden om te kijken of twee eigenschappen, P en Q , dezelfde ϑ hebben.

Idee: kijk naar de tabel:

	P	Q
Met eigenschap	10	18
Zonder eigenschap	80	55

Als de verhoudingen $\frac{10}{80}$ en $\frac{18}{55}$ 'te veel verschillen' zijn de ϑ verschillend (de p -waarde van de aanname dat ze gelijk zijn is lager).

Fisher's exact test

Kan gebruikt worden om te kijken of twee eigenschappen, P en Q , dezelfde ϑ hebben.

Idee: kijk naar de tabel:

	P	Q
Met eigenschap	10	18
Zonder eigenschap	80	55

Als de verhoudingen $\frac{10}{80}$ en $\frac{18}{55}$ 'te veel verschillen' zijn de ϑ verschillend (de p -waarde van de aanname dat ze gelijk zijn is lager).

Hier: $p = 0,355$, dus de aanname dat de verhoudingen gelijk zijn verwerpen.

Fisher's exact test

In ons (hypothetische) geval hebben we

	P	Q
Met eigenschap	N	?
Zonder eigenschap	0	?

Fisher's exact test

In ons (hypothetische) geval hebben we

	P	Q
Met eigenschap	N	?
Zonder eigenschap	0	?

Voor Q gebruiken we een combinatie van een aantal veel voorkomende eigenschappen in de WALS.

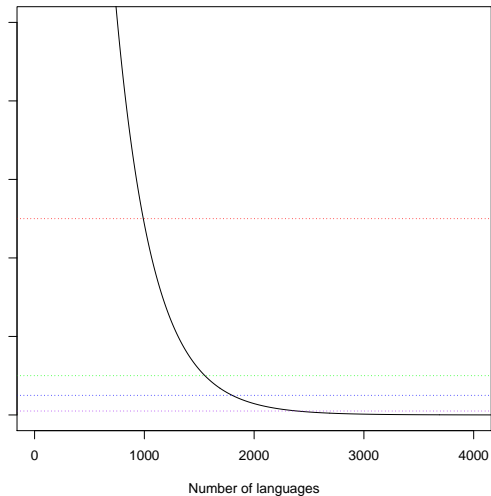
Fisher's exact test

In ons (hypothetische) geval hebben we

	P	Q
Met eigenschap	N	13703
Zonder eigenschap	0	43

Voor Q gebruiken we een combinatie van een aantal veel voorkomende eigenschappen in de WALS.

p -waarde



p -waarde	Minimale N
0,05	991
0,01	1552
0,005	1801
0,001	2393

Beperkingen

Dus een survey van 991 talen is voldoende! Maar...

Beperkingen

Dus een survey van 991 talen is voldoende! Maar...

- ▶ We hebben niet laten zien dat $\vartheta = 1$, alleen dat $\vartheta > \frac{13703}{13748} \approx 0,997$. Zonder formules: met duizend talen kun je niet een tegenvoorbeeld dat maar een op de miljoen keer voorkomt uitsluiten.

Beperkingen

Dus een survey van 991 talen is voldoende! Maar...

- ▶ We hebben niet laten zien dat $\vartheta = 1$, alleen dat $\vartheta > \frac{13703}{13748} \approx 0,997$. Zonder formules: met duizend talen kun je niet een tegenvoorbeeld dat maar een op de miljoen keer voorkomt uitsluiten.
- ▶ We gaan uit van 991 ongerelateerde talen, met gerelateerde talen heb je er waarschijnlijk meer nodig!

Beperkingen

Dus een survey van 991 talen is voldoende! Maar...

- ▶ We hebben niet laten zien dat $\vartheta = 1$, alleen dat $\vartheta > \frac{13703}{13748} \approx 0,997$. Zonder formules: met duizend talen kun je niet een tegenvoorbeeld dat maar een op de miljoen keer voorkomt uitsluiten.
- ▶ We gaan uit van 991 ongerelateerde talen, met gerelateerde talen heb je er waarschijnlijk meer nodig!
- ▶ Volgens Ethnologue zijn er 143 taalfamilies en 83 isolaten.

Welke p is significant?

- ▶ We gebruiken een grens van $p = 0,05$.

Welke p is significant?

- ▶ We gebruiken een grens van $p = 0,05$.
- ▶ Per definitie: 1 op de 20 bijna-universals classificeren we dus als universal!

Welke p is significant?

- ▶ We gebruiken een grens van $p = 0,05$.
- ▶ Per definitie: 1 op de 20 bijna-universals classificeren we dus als universal!
- ▶ In *The Universals Archive* heeft 2029 potentiële universals; zelfs als ze allemaal niet universeel zouden zijn, zouden we alsnog ± 400 universals 'vinden'.

Welke p is significant?

- ▶ We gebruiken een grens van $p = 0,05$.
- ▶ Per definitie: 1 op de 20 bijna-universals classificeren we dus als universal!
- ▶ In *The Universals Archive* heeft 2029 potentiële universals; zelfs als ze allemaal niet universeel zouden zijn, zouden we alsnog ± 400 universals 'vinden'.
- ▶ Maar voor 'strengere' p -waarden zijn de tot nu toe gedane surveys veel te klein (grootste in WALS heeft 1519 talen, meeste maar een paar honderd)

Conclusie

- ▶ Er is geen adequaat empirisch bewijs voor het bestaan van universals;

Conclusie

- ▶ Er is geen adequaat empirisch bewijs voor het bestaan van universals;
- ▶ Al helemaal niet voor implicationele universals;

Conclusie

- ▶ Er is geen adequaat empirisch bewijs voor het bestaan van universals;
- ▶ Al helemaal niet voor implicationele universals;
- ▶ 'Universals' uit de literatuur zijn dus niet dwingend!

Conclusie

- ▶ Er is geen adequaat empirisch bewijs voor het bestaan van universals;
- ▶ Al helemaal niet voor implicationele universals;
- ▶ 'Universals' uit de literatuur zijn dus niet dwingend!
- ▶ Meer onderzoek is nodig naar kansmodellen voor taaleigenschappen;

Conclusie

- ▶ Er is geen adequaat empirisch bewijs voor het bestaan van universals;
- ▶ Al helemaal niet voor implicationele universals;
- ▶ 'Universals' uit de literatuur zijn dus niet dwingend!
- ▶ Meer onderzoek is nodig naar kansmodellen voor taaleigenschappen;
- ▶ niet-absolute *tendencies* zijn interessanter dan universals!