

*Markos Kramer*

## **Ĉu tamen eblas pri-Esperanta lingvoscienco?**

... aŭ

### **Kiel difini kiuj estas la kompetentaj parolantoj kiuj determinas la lingvan regularon?**

#### **Resumo**

En sia artikolo “La neebleco de priesperanta lingvoscienco” Ken Miner argumentas ke ĉar Esperanto ne havas indiĝenajn parolantojn, kiuj povus servi kiel fidinda datumfonto pri la gramatikeco kaj signifo de Esperantlingvaj frazoj, ne eblas priskriba lingvoscienco pri Esperanto. En ĉi tiu artikolo mi ekzamenas lian argumenton, atentigante pri tio ke ĝi iugrade ankaŭ koncernas la lingvosciencan analizon de aliaj lingvoj, kaj proponante solvon al la de li elstarigita problemo. La konkludo estas, ke pri-Esperanta lingvoscienco ja eblas, sed ke necesas relative multaj rimedoj por kolekti fidindajn datumojn pri gramatikeco kaj signifo de Esperantlingvaj frazoj.

#### **Summary**

**Is an Esperanto linguistics none the less possible? or How to define who are the competent speakers to determine the linguistic rules.**

In his article on the impossibility of an Esperanto linguistics, Ken Miner argues that because Esperanto does not have indigenous speakers who can serve as a reliable source of data on the grammaticality and meaning of Esperanto sentences, it is not possible to describe a linguistics of Esperanto. In this article I examine his argument, pointing out in that connection that to some degree it

concerns the linguistic analysis also of other languages, and proposing a solution to the problem he has highlighted. The conclusion is that a linguistics of Esperanto is indeed possible, but that relatively large numbers of procedures are necessary to collect reliable data on the grammaticality and meaning of Esperanto sentences.

## 1. La argumento de Ken Miner

En la festlibro por Humphrey Tonkin publikigita en 2010 aperis artikolo de Ken Miner titolita “La neebleco de priesperanta lingvoscienco” (Miner, 2010). La artikolo estis republikigita iom reviziite kaj kun paralela angla traduko en la planlingvosciencia revuo InKoj (Miner, 2011).

Komence de la artikolo, li enkondukas la terminon *indiĝena parolanto*, por kiu li en la anglalingva versio uzas la kutiman terminon *native speaker*, kiu kutime estas tradukita per *denaska parolanto*. Li pravigas tiun nekutiman termin-elekton per tio, ke tiuj, kiujn oni kutime nomas denaskaj parolantoj de Esperanto (kaj por kiuj li anglalingve uzas la terminon *speakers-from-birth*), ne kreskas en Esperantlingva ĉirkaŭaĵo, infanaĝe lernas la ĉirkaŭaĵan lingvon pli bone ol Esperanton, kaj sekve ne estas konsiderataj la plej bonaj parolantoj de Esperanto, kiuj povus servi kiel datumfonto por lingvosciencaj esploroj. Do ili ne havas la ecojn, kiujn oni kutime asocias kun la termino *native speaker* kaj alilingvaj egalajoj en la lingvosciencia diskurso. Kvankam mi ne povas plene aprobi ĉi tiun devojiĝon de la kutima terminaro, mi por ĉi tiu reago al la artikolo de Miner transprenos lian termin-elekton.

La priskriba lingvoscienco priskribas la lingvajn regulojn surbaze de la lingvaj kapabloj de indiĝenaj parolantoj. Praktike tio realigeblas en du manieroj:

- Oni povas studi la lingvajon efektive uzatan de indiĝenaj parolantoj. Por tio oni kolektas multajn specimenojn de ilia lingvouzo en tiel nomata *tekstaro* (ankaŭ nomata *korpuso*) kaj poste studas, kiuj lingvaj fenomenoj troveblas kaj kiuj ne troveblas en ĉi tiu tekstaro.

- Oni povas demandi indiĝenajn parolantojn pri la gramatikeco (do gramatika akceptebleco) kaj pri la signifo de unuopaj frazoj (aŭ plurfrazaj lingvaj specimenoj).

La unua maniero studi la lingvon estas aplikata de la *tekstara lingvoscienco*. Ofte, kvankam certe ne ĉiam, la tekstara lingvoscienco laboras per tekstaroj konsistantaj el profesie produktitaj tekstoj (ekzemple la anglalingva tekstaro *BLLIP 1987-89 WSJ Corpus* kun tekstoj el la gazeto *Wall Street Journal*, kaj la traduktekstaro *Europarl Corpus* bazita sur aktoj de la Eŭropa Parlamento). Ĉe tiaj tekstaroj la koncepto de indiĝena parolanto ne grave rolas, ĉar la fakto ke la kolektitaj tekstoj estas profesie produktitaj kaj redaktitaj kutime jam certigas ke ili ne entenas esprimojn aŭ dirmanierojn neakcepteblajn por indiĝenaj parolantoj. Io komparebla por Esperanto estas la *Tekstaro de Esperanto* kompilita de Bertilo Wennergren el diversĝenraj tekstoj el diversaj tempoj ekde la estiĝo de Esperanto (sed superregas tekstoj de la frua tempo antaŭ la Unua Mondmilito kaj de la cifereca tempo ekde 1990). La tekstoj de ĉi tiu tekstaro estas libroj verkita de aŭtoroj ĝenerale konsiderataj bonaj aŭtoroj, kaj artikoloj aperigitaj en alte respektataj gazetoj, tiel ke la lingva nivelo de la aŭtoroj kune kun la lingva redakto de la publikigistoj certigas lingvan nivelon ĝenerale konsideratan modela en Esperantujo. Miner atentigas pri tio, ke la studado de tekstaroj ne alportas negativan evidenton, do evidenton pri tio, kio ne eblas en lingvo, kaj sekve ne povas esti elĉerpa metodo por la empiria lingvoscienco. Pro tio, la dua ĉi-supre priskribita maniero studi la lingvajn kapablojn de indiĝenaj parolantoj, do la gramatikec-demandoj al indiĝenaj parolantoj, estas ne-ellasebla. La lingvosciencistoj, kiuj faras tiajn studojn, kutime tre atentigas pri la efektiva indiĝeneco de la demanditoj, ĉar aliokaze iliaj datumoj estus poluitaj. Do kiel oni faru tiajn studojn pri Esperanto? Krom tio, ke Miner atentigas pri la diferenco inter indiĝenaj kaj denaskaj parolantoj, li ankaŭ mencias, ke ĉiuokaze jam estas vaste akceptite ke la Esperantaj denaskuloj ne havas la statuson de ideala reprezentanto de la lingvo uzebla por studi la lingvan regularon.

Ĉu oni anstataŭe demandu la elitajn Esperanto-parolantojn, ekzemple la Akademianojn? Responde al ĉi tiu propono Miner atentigas

pri tio, ke pri multaj tiklaj lingvaj demandoj oni ricevas sufiĉe variajn respondojn de elitaj Esperanto-parolantoj. Tio ŝuldiĝas al la fakto, ke ili rezonas pri lingvaj demandoj per tiaj kategorioj kiel logikeco, Zamenhofeco, Fundamenteco, nacilingva influo kaj simile anstataŭ fidi sian lingvosenton. Do ili respondas al tiaj lingvaj demandoj en tute alia maniero ol farus averaĝa indiĝena parolanto de iu ajn nacia lingvo.

Ĉar sekve ne eblas kolekti negativan evidenton pri la lingva regularo de Esperanto, Miner konkludas ke ne eblas pri-Esperanta lingvoscienco. Li koncedas, ke tamen eblas esperantologio, ĉar tradicie la esperantologio ja funkcias per tiaj kategorioj kiel Zamenhofeco, Fundamenteco, kohereco de la lingva sistemo, ekvilibro inter diversaj nacilingvaj influoj kaj simile, kiuj tamen estas tute fremdaj al la metodaro de la moderna priskriba lingvoscienco.

Miner prezentas la jenan ekzemplon por ilustrati sian argumenton: Supozu, ke ni ŝatus prezenti lingvosciencan analizon de la signifo de solaj *des*-esprimoj, do *des*-esprimoj sen partneraj *ju*-esprimoj, kiel en la sekva frazo:

(1) Se Johano ne venos, des pli bone.

Li atentigas ke la klarigoj pri tiaj solaj *des*-esprimoj prezentitaj en la *Plena Manlibro de Esperanta Gramatiko* de Bertilo Wennergren (2005) ne estas kontentigaj el lingvosciencia vidpunkto, kvankam ili ja taŭgas por praktikaj celoj. Li prezentas la jenan lingvosciencan hipotezon pri la signifo de solaj *des*-esprimoj:

Sola 'des'-esprimo estas ekstercentra konstruo (bezonas ion alian en la kunteksto por havi signifon), signifanta, ke io en la lingva kunteksto, ni nomu ĝin X, kreas pli altan atendeblecon (sed ne certecon) de Y, ol Y havus sen X. Y estas ĉiam kvante interpretebla.

Por testi tian hipotezon, oni devas kontroli ne nur, ĉu ĝi aplikeblas al reale troveblaj uzoj de sola *des*, sed ankaŭ, ĉu en la frazoj, al kiuj la hipotezo ne aplikeblas, ne estus akceptebla sola *des*. Oni do devus kontroli ekzemple ke (2), kiu ne plenumas la hipotezon, estas ne-akceptebla en Esperanto, dum (3), kiu plenumas ĝin, estas akceptebla:

(2) Johano havas 4 infanojn; des pli li havas 3 infanojn.

(3) Li estis simpatia; des pli la aliaj ŝatis lin.

Miner prezentas ankoraŭ pliajn frazojn kun dubindaj uzoj de sola *des-esprimo*. Fine li konkludas, ke ne eblas “eltiri el lingvo, kiun ĉiuj parolas kiel fremdan lingvon, la subtilajn intuiciojn, kiuj nepras por lingvosciencia analizo”.

## 2. La origino de la problemo

En ĉi tiu sekcio mi prezentos miajn ideojn pri tio, kio estas la kaŭzo de la problemo elstarigita de Miner. Aldone mi atentigos pri tio, ke ĝi ne nur rilatas al Esperanto, sed diversgrade ankaŭ al aliaj lingvoj kaj lingvo-variaĵoj, kiujn la priskriba lingvoscienco havas aŭ povus havi kiel studobjekton.

Por esplori la regularon de iu lingvo (aŭ lingva variaĵo) oni devas konsulti kompetentajn parolantojn de tiu lingvo. Sed kiuj estas la kompetentaj parolantoj de difinita lingvo? Nu, kompetentaj estas tiuj, kiuj sekvas la regulojn de la lingvo en sia lingvouzo (kaj en siaj juĝoj pri gramatikeco).

Sed nun ni havas ciklan difinon de la lingva regularo kaj de la koncepto de kompetenta parolanto. Kutime la lingvosciencistoj evitas ĉi tiun ciklecon per tio, ke ili uzas kiel kriterion por lingva kompetenteco ion sendependan de la reguloj de la lingvo: Plej ofte, ili uzas kiel kriterion por lingva kompetenteco, ke la koncerna persono estu lerninta la lingvon denaske kaj de siaj gepatroj kaj de sia ĉirkaŭaĵo, do ke temu pri indiĝena parolanto de la lingvo. Kutime estas tute ignorata la problemo, ke ekzistas homoj, kiuj ne havas indiĝenan lingvon.

Ĉi tiu solvo funkcias relative bone por lingvoj, kiuj estas grandskale uzataj de homoj, kiuj lernis la lingvon kiel ĉefan lingvon infanaĝe. Sed ja ekzistas vivantaj lingvoj, kiuj ne estas tiaj, ekzemple Esperanto. Aliaj ekzemploj de tiaj lingvoj estas la moderna hebrea de la frua 20a jarcento kaj la sangoa lingvo de la malfrua 20a jarcento. La moderna hebrea komence de la 20a jarcento jam estis vivanta lingvo, kaj jam

estis evoluinta lingvan regularon, kiu principe devus esti studebla de lingvoscienco. Sed la aliro surbaze de indiĝenaj parolantoj tiutempe ankoraŭ ne estus funkciinta.

La sangoa lingvo estas nun la nacia lingvo de la Centra Afrika Respubliko. Ĝi estiĝis kiel komerca interlingvo surbaze de certa dialekto de la ngbanda lingvo. La sangoa lingva regularo certe diferencas de la ngbanda lingva regularo, kaj certe estiĝis jam antaŭ ol estis signifa kvanto de indiĝenaj parolantoj de la sangoa. Ankaŭ ĉi tie, la kutima maniero difini lingvan regularon surbaze de indiĝenaj parolantoj ne funkcias.

La kutima solvo ankaŭ malbone funkcias por studi la regulojn de neprestiĝaj lingvovariaĵoj. Se oni ekzemple volas studi la regulojn de la parola germana lingvo, oni frontas la problemon, ke demandante germanlingvanojn pri frazo kun ĉefpropozicia vortordo post *weil* ('ĉar'), oni konstatos, ke la plej multaj el ili verŝajne malakceptos ĝin. Kaj tamen ili ĉiutage uzas tian vortordon post *weil*. La problemo devenas de tio, ke cerbumante pri gramatika ĝusteco de frazoj, ili aplikas la regulojn de la norma skriblingvo, kiu en ĉi tiu detalo diferencas de la reguloj de la parola lingvo.

Do lingvosciencie esplori la lingvan regularon de neprestiĝaj lingvotavoloj ankaŭ alportas malfacilaĵojn pro nefidindeco de la asertoj de la lingvouzantoj, kiuj pro siaj prilingvaj asertoj aplikas aliajn regulojn ol por sia efektiva lingvouzo.

Krome la kutima solvo alfrontas aldonan problemon ĉe dialektoj aŭ sociolektoj, kies parolantaro ne estas klare apartigebla de la parolanto de norma lingvo parenca al la dialekto/sociolekto. Konsideru ekzemple la afrik-usonan sociolekton de la angla. Kiel oni povus lingvosciencie esplori ĝian lingvan regularon? Eĉ se ni momente supozu ke ekzistas solvo al la problemo menciita en la antaŭa alineo, ni havas ankoraŭ la aldonan problemon, ke ni devas difini, kiu estas kompetenta uzanto de tiu sociolekto kaj kiu ne. Ni ne povas simple konsideri ĉiujn afrikdevenajn usonanojn, ĉar multaj el ili tute ne parolas tiun sociolekton. Aliaj eble foje uzas esprimojn el tiu sociolekto, sed ne kreskis en ĉirkaŭaĵo kie tiu sociolekto estas kutima, kaj sekve ne bone regas ĝin.

Aliaj eble ja kreskis en tia ĉirkaŭaĵo, sed translokiĝis kiam ili komencis siajn universitatajn studojn, kaj ekde tiam ne plu estas en tia ĉirkaŭaĵo; kiam ili tamen foje uzas tiun sociolekton, ili uzas ĝin en maniero tre influita de la lingvovariaĵoj, kiujn ili nun uzas multe pli.

Do por studi la lingvan regularon de la sociolekto, ni devas ignori ĉiujn ĉi homojn ne vere kompetentajn pri tiu sociolekto. Alivorte, por difini kiu estos kompetenta parolanto akceptebla por provizi al ni fidindan pozitivan kaj negativan evidenton, ni devas konsideri la kompetentecan de la parolantoj. Sed kio estas kompetenteco? Simpla respondo estus: Uzanto de iu lingva variaĵo estas kompetenta, se ĝi<sup>1</sup> kapablas uzi la variaĵon laŭ la lingva regularo de tiu variaĵo. Sed nun ni denove havas la probleman ciklan difinon.

### 3. Rekursia fakto-trovado kiel solvo al la problemo

Eblas eviti la ciklecon de la difinoj de lingva regularo kaj kompetenta parolanto en alia maniero, nome anstataŭigante la ciklan difinon per plurŝtupa rekursia difino.

En la unua ŝtupo de ĉi tiu plurŝtupa difino, oni formas unuan hipotezon pri la reguloj de la lingvo surbaze de la ĝenerala lingvouzo de la uzantoj de la koncerna lingvo, sen atento je ilia kompetenteco, kaj poste formas unuan hipotezon pri la lingva kompetenteco de la parolantoj surbaze de tio, kiom ilia lingvouzo similas al la hipotezita regularo. En la dua ŝtupo de la difino, oni unue formas plibonigitan hipotezon pri la reguloj de la lingvo surbaze de la lingvouzo de la kompetentaj parolantoj (do la parolantoj hipotezitaj kompetentaj en la unua ŝtupo), kaj poste formas plibonigitan hipotezon pri la lingva kompetenteco de la parolantoj surbaze de la plibonigita regularo-

---

1 Mi uzas la pronomon *ĝi* kiel sekse neŭtralan pronomon por paroli pri persono, konforme kun la Lingva Respondo 91 "Pri pronomo por 'homo'" de Zamenhof (La Revuo, 1901, Aŭgusto), en kiu li diris interalie: "Kiam ni parolas pri homo, ne montrante la sekson, tiam estus regule uzi la pronomon "ĝi" (kiel ni faras ekzemple kun la vorto "infano"), kaj se vi tiel agos, vi estos gramatike tute prava." Vidu L.L. Zamenhof, *Lingvaj respondoj*, red. G. Waringhien (Kyoto: Eldonejo Ludovikito, 1990), pĝ. 85. Pli da informoj pri la temo troveblas en mia artikolo Esperanto kaj sekso publikigita interrete ĉe <http://lingvakritiko.com/2014/10/16/esperanto-kaj-sekso>

hipotezo. Ĉi tiun procedon eblas daŭrigi ĝis kiam la rezultoj konverĝas, do ĝis kiam ne plu estas signifa ŝanĝo de unu ŝtupo al la sekva.

Matematike la problemo estas simila al problemo studata sub la nomo *fakto-trovado* (angle *fact finding* aŭ *truth discovery*). Supozu ke diversaj homoj prijuĝas la veron aŭ malveron de iuj asertoj, pri kies vervaloro vi scias nenion. Por fari juĝon pri la vervaloro de la asertoj surbaze de la prijuĝoj de ĉi tiuj juĝintoj, vi povus simple kalkuli la plimultan opinion pri ĉiu unuopa aserto, kaj supozi ke verŝajne pravas la plimulto. Sed supozu ekzemple ke dum la plej multaj juĝintoj havis inter si tre similajn juĝojn, estas unu juĝinto, kiu ĝenerale juĝas malsame ol la aliaj. Tiam estas racie supozi ke tiu lasta juĝinto ne estas kompetenta. Se do estas iu aserto, ĉe kiu la plimulto dependas de ĝia juĝo, estas preferinde kalkuli la plimultan opinion de la kompetentaj juĝintoj, ignorante la opinion de la nekompetenta juĝinto.

Ĉi tie ni do havas situacion similan al la situacio pri la lingva kompetenteco: Ni volas prijuĝi, kiuj asertoj estas veraj surbaze de la opinio de la kompetentaj juĝintoj, kaj samtempe prijuĝi, kiu estas kompetenta surbaze de tio, kiugrade ĝiaj juĝoj estas konformaj kun la vero. Tio similas al la supre menciita cikla difino, laŭ kiu ni ŝatus trovi la regularon de lingvo surbaze de la lingvaj juĝoj de kompetentaj lingvo-uzantoj, kaj ŝatus samtempe prijuĝi, kiu estas lingve kompetenta surbaze de tio, kiugrade ĝiaj lingvaj juĝoj konformas kun la lingva regularo.

Ekzistas diversaj algoritmoj por fakto-trovado, sed ĉiuj baziĝas sur la ideo de plurŝtupa kalkulado de la verŝajna vero kaj la verŝajna kompetenteco de la juĝintoj.

Fakto-trovado estas aplikata por multaj celoj, interalie por ricevi fidindajn rezultojn surbaze de *amas-eksterigo* (angle *crowdsourcing*). Amas-eksterigo estas la praktiko interrete doni iujn taskojn relative facile plenumebajn de homoj, sed ne facile aŭtomatigeblajn per komputiloj, al multaj homoj samtempe. Unu ekzemplo por tio estas la primarkado de lingvosciencaj tekstaroj, ekzemple marki kiuj vortoj en teksto estas propraj nomoj aŭ marki la gramatikan rolon de diversaj vortgrupoj. Se oni amas-eksterigas la primarkadon de tekstaro, oni petas multajn homojn samtempe tiel primarki tekstaron. Se ne ĉiuj

primarkintoj primarkis iun teksteron en la sama maniero, oni tradicie simple transprenis la primarkadon de la plimulto. Sed eblas plibonigi la fidindecon de la rezulto se oni uzas taŭgan fakto-trovan algoritmon, kiu paralele kalkulas la verŝajne ĝustan primarkadon kaj la verŝajnan kompetentecon de la primarkintoj.

## 4. La matematiko de fakto-trovado

En ĉi tiu sekcio mi iom skizas kiel fakto-trovado estas traktata matematike. Ne-matematikemaj legantoj povas senprobleme transiri al la sekva sekcio.

La plej multaj algoritmoj por fakto-trovado baziĝas sur la ideo, ke la fidindeco-nivelo donata en ĉiu ŝtupo al ĉiu juĝinto estas reala nombro inter 0 kaj 1. Fidindeco-nivelo de 0 indikas, ke la juĝinto estas tute ne-fidinda, dum fidindeco-nivelo de 1 indikas, ke la juĝinto estas tute fidinda. Se la algoritmo baziĝas sur probablo-teoriaj principoj, fidindeco inter 0 kaj 1 povas esti interpretata kiel la probablo, ke la koncerna juĝinto ĝuste scias, ĉu aserto estas vera aŭ malvera.  $F(j)$  estu la fidindeco-nivelon de juĝinto  $j$ . La fidindeco-nivelon hipotezitan je la  $n$ -a ŝtupo de la algoritmo ni nomas  $F n(j)$ .

Ĉiu aserto estas aŭ vera aŭ malvera, sed ni ne anticipu scias ĝian vervaloron. En ĉiu ŝtupo de la algoritmo, ni determinas la verŝajnecon de tio, ke ĝi estas vera, kaj tiu verŝajneco ankaŭ estas mezurata per valoro inter 0 kaj 1. La verŝajneco, kiun ni atribuas al la vereco de aserto  $a$  je la  $n$ -a ŝtupo de la algoritmo estas nomata  $V n(a)$ .

Notu ke mi uzis la vorton *probablo* por la fidindeco de la juĝintoj, sed la vorton *verŝajneco* por la valoro atribuita al la asertoj. Tio estas grava termina distingo en la probabloteorio.<sup>2</sup> La unuan mi nomas *probablo*, ĉar ĝi estas unu el la parametroj de nia modelo: Ni supozas, ke la fidindeco de ĉiu juĝanto povas esti mezurata kiel la probablo, ke ĝi faras ĝustan aserton. La duan ni nomas *verŝajneco*, ĉar ĝi ne estas probablo fiksita en la modelo. Ja por la asertoj la modelo antaŭsupozas, ke ili estas ĉiuj

---

2 En la angla lingvo la koncernaj terminoj estas *probability* kaj *likelihood*, en la hispana *probabilidad* kaj *verosimilitud*.

aŭ veraj aŭ malveraj, kio estas duuma valoro kaj ne kontinua valoro kiel ĉe probablo. La verŝajneco de aserto do ne estas identigebla kun iu probablo fiksita en la modelo, sed estas la nivelo de certeco, kiun ni havas pri tio, ke iu aserto estas vera.

La algoritmo devas difini iujn komencajn valorojn  $F_0(j)$  kaj/ aŭ  $V_0(a)$  por ĉiu juĝanto  $j$  kaj ĉiu aserto  $a$ , kaj aldone iun metodon por kalkuli  $F_{n+1}(j)$  kaj  $V_{n+1}(a)$  surbaze de  $F_n(j)$  kaj  $V_n(a)$  por ĉiu natura nombro  $n$ . Unu simpla ekzemplo de tia algoritmo, kiu estas iomete simpligita versio de algoritmo origine difinita de Yin k.a. (2008), estas la jena: La komenca valoro  $V_0(a)$  estas 0.5 por ĉiu aserto  $a$  (por  $F_0(j)$  oni en ĉi tiu algoritmo ne bezonas komencan valoron), kaj  $F_{n+1}(j)$  kaj  $V_{n+1}(a)$  estas kalkulataj jene, kie  $A$  estas la aro de prijuĝataj asertoj kaj  $J_a$  estas la aro de juĝintoj, kiuj juĝis  $a$  vera:

$$F^{n+1}(j) = \frac{\sum_{a \in A_j} V^n(a)}{|A|}$$

$$V^{n+1}(a) = 1 - \prod_{j \in J_a} (1 - F^{n+1}(j))$$

$$V(a) = \lim_{n \rightarrow \infty} V^n(a)$$

La tria formulo indikas, ke por ekscii la finan verŝajnecon  $V(a)$  de aserto  $a$ , oni devas kalkuli la limvaloron, al kiuj konverĝas la valoroj  $V_n(a)$ . Kompreneble eblas pruvi, ke tiuj valoroj ĉiam konverĝas.

Ĉi tiu algoritmo de Yin k.a. ja baziĝas sur probabloteoriaj ideoj, sed fakte ne tute rigore aplikas la probabloteoriajn metodojn por kalkuli verŝajnecon de la asertoj surbaze de la ĉi-supre priskribita probabloteoria modelo. Se oni faras tiel, oni havas iom pli malsimplan algoritmon, kiu estis priskribita de Wang k.a. (2014).

Unu el la novaj evoluoj en la pristudado de fakto-trovado estas, ke oni celas per logikaj metodoj sistemece prijuĝi, kiu el la diversaj

proponitaj algoritmoj estas preferinda surbaze de difinitaj racieco-kriterioj (Kruger k.a., 2014). Ankoraŭ ne estas kunigitaj la novaj probabloteoriaj ideoj de Wang k.a. (2014) kaj la novaj logikaj ideoj de Kruger k.a. (2014), sed tio certe estos la sekva evoluo en la studado de la matematiko de fakto-trovado.

## 5. Apliko de fakto-trovado al lingvosciencaj studoj

Uzante la metodojn de fakto-trovado, oni povas fari lingvosciencajn studojn sen havi kriterion, kiu ebligas anticipi klasigi lingvo-uzantojn kiel kompetentajn, kiel ekzemple la kriterio esti indiĝena parolanto de lingvo. Anstataŭe, oni kolektas datumojn pri la lingvouzo kaj gramatikaj juĝoj de lingvouzantoj sendepende de ilia kompetenteco, kaj poste uzas fakto-trovadon por paralele konkludi, kiuj datumoj estas fidindaj kaj kiuj lingvouzantoj estas kompetentaj.

La sola problemo kun ĉi tiu aliro estas, ke ne sufiĉas simple demandi kelkajn lingvouzantojn pri iliaj juĝoj pri la gramatikeco kaj signifo de kelkaj frazoj, sed necesas demandi relative multajn lingvouzantojn pri iliaj juĝoj pri relative multaj frazoj. La granda kvanto de lingvaj juĝoj faritaj de ĉiu lingvouzanto partoprenanta en la studo certigas ke la mezuroj de kompetenteco de la lingvouzantoj kalkulitaj per la fakto-trovado estas fidindaj. Kaj la granda kvanto de lingvouzantoj partoprenantaj la studon certigas, ke la verŝajnecoj atribuitaj de la fakto-trova algoritmo al la gramatikeco de la studataj frazoj aŭ al iliaj eblaj signifoj estas fidindaj.

Por fari rigoran lingvosciencan studon pri Esperanto surbaze de ĉi tiuj ideoj, oni do bezonas relative multajn rimedojn, pro kio estus bone se tia studo estus farita en kunordigita maniero fare de pluraj esperantologoj kaj Esperantaj lingvosciencistoj. Unue necesus kolekti liston de lingvosciencaj demandoj pri Esperanto, pri kiuj oni volas havi datumojn por priskriba lingvosciencan analizon. Poste oni devus fari liston da frazoj, kies gramatikecon aŭ signifon la demandotaj Esperanto-parolantoj devos prijuĝi. Tiu listo devus enhavi almenaŭ unu frazon (prefere plurajn) por ĉiu pristudata lingvosciencan demando.

Aldone la listo povus enhavi kelkajn frazojn, kies (ne-)gramatikeco aŭ signifo ne estas pridubata, sed kiuj povus helpi pli bone juĝi la kompetentecon de la demandotaj Esperanto-parolantoj. Krome oni per tiuj aldonaj frazoj povus testi, ĉu la metodo donas la atendatajn rezultojn ĉe ne-pridubataj detaloj de la Esperanta gramatiko kaj vortprovizo. Mi hipotezas, ke ĝi ja donas la atendatajn rezultojn ĉe ĉiuj detaloj, pri kiuj regas plena konsento inter esperantologoj. Se tia studo havus rezulton kontraŭan al ĉi tiu hipotezo, tio ja estus tre interesa rezulto, kiu neprigus ian teorion klarigon.

Fine oni devus peti relative multajn Esperanto-parolantojn partopreni tiun studon per siaj juĝoj pri la gramatikeco kaj signifo de ĉiuj frazoj kolektitaj por la studo. Se oni ebligas al Esperanto-parolantoj interrete partopreni tiun studon, oni devus igi tre klare, ke ili estas petataj respondi nur surbaze de sia lingva kapablo, sen rigardi en vortaroj aŭ gramatikoj. Aliokaze ili simple reproduktus la lingvajn juĝojn de tiuj referencverkoj, tiel ke la studo ne studus la efektivan lingvosenton de la Esperanto-parolantoj, sed nur ian kombinon de ilia lingvosento kaj la rekomendoj de tiaj referencverkoj.

Al la demandoj pri gramatikeco aŭ signifo de diversaj frazoj oni eble volas permesi ne nur duumajn respondojn (ekzemple “gramatike ĝusta” / “gramatike erara”, aŭ “havas signifon A” / “havas signifon B”), sed pli da alternativoj (ekzemple “gramatike kaj stile bona” / “gramatike ĝusta sed stile malbona” / “gramatike erara”, aŭ “havas signifon A” / “havas signifon B” / “havas signifon C”). La fakto-trovado kiel priskribita en la antaŭa sekcio estas aplikebla nur al duumaj respondoj, ĉar la vervaloroj de la asertoj estas duuma. Sed fakte ekzistas ankaŭ fakto-trovaj algoritmoj aplikeblaj en situacioj, kiam estas pli ol du alternativoj: la tri cititaj artikoloj de X. Yin, J. Han kaj P. S. Yu (2008), Wang kaj aliaj (2014) kaj Kruger kaj aliaj. (2014) fakte ĉiuj difinas fakto-trovajn algoritmojn aplikeblaj al situacioj, en kiuj estas pli ol du alternativoj (mia simpligo en la prezento de la algoritmo de Yin k.a. konsistis ĝuste en tio, ke mi limigis ĝin al la duuma kazo).

## 6. Konkludo

Ken Miner atentigis pri grava problemo de pri-Esperanta lingvoscienco, nome ke pro la manko de indiĝenaj parolantoj de Esperanto ne eblas fari lingvosciencajn studojn por Esperanto en la maniero, kiu estas nuntempe kutima en la priskriba lingvoscienco, nome per tio, ke oni petas al indiĝenaj parolantoj de la lingvo juĝi la gramatikecon kaj signifojn de diversaj frazoj. Kiel mi klarigis en ĉi tiu artikolo, la problemo iugrade ankaŭ aperas ĉe studado de aliaj lingvoj.

La problemo venas de tio, ke la teorie dezirata difino-paro por difini la lingvan regularon kaj la lingvan kompetentecon estas cikla: La difino de la unua dependas de la dua, kaj la difino de la dua dependas de la unua. Kutime lingvosciencistoj evitas ĉi tiun problemon per tio, ke ili alprenas eksterlingvan kriterion por lingva kompetenteco, nome la indiĝenecon de la demandata lingvo-uzanto. Sed tiu solvo ne estas aplikebla al Esperanto.

Mi proponas solvi la problemon anstataŭigante la ciklan difinon per plurŝtupa rekursia difino. Tion eblas precizigi per la metodoj de la *fakto-trovado*, matematika metodo jam aplikata por diversaj aliaj problemoj, ĉe kiuj necesas surbaze de la juĝoj de iuj homoj samtempe determini la faktojn kaj la fidindecon de la juĝintoj.

La solvo kiun mi proponas ebligas lingvosciencajn studojn pri Esperanto. Sed por ke la solvo bone funkcii, la studoj devas esti grandaj. Pro tio estus bona ideo se pluraj esperantologoj kaj Esperantaj lingvosciencistoj kunordigus unu tian studon, en kiu oni studas multajn lingvosciencajn demandojn samtempe.

## Cititaj verkoj

- Kruger, J., Endriss, U., Fernandez, R., & Qing, C. (2014). Axiomatic analysis of aggregation methods for collective annotation. *Proceedings of the 2014 International Conference on Autonomous Agents and Multiagent Systems* (pp. 1185– 1192). Richland, SC: International Foundation for Autonomous Agents and Multiagent Systems.
- Miner, K. (2010). La neebleco de priesperanta lingvoscienco. In D. Blanke & U.Lins (Eds.), *La arto labori kune. Festlibro por Humphrey Tonkin* (pp. 239–270). Rotterdam: Universala Esperanto-Asocio
- Miner, K. (2011). The impossibility of an Esperanto linguistics / La neebleco de priesperanta lingvoscienco. *InKoj. Philosophy & Artificial Languages*, 2, 1, 26–51.
- Wang, D., Kaplan, L., & Abdelzaher, T. F. (2014). Maximum likelihood analysis of conflicting observations in social sensing. *ACM Transactions on Sensor Networks*, 10, 2 [2530289], <https://doi.org/10.1145/2530289>.
- Wennergren, B. (2005). *Plena manlibro de Esperanta gramatiko*. Emeryville, CA: Esperanto-Ligo por Norda Ameriko.
- Yin, X., Yu, P. S., & Jiawei H. (2008). Truth discovery with multiple conflicting information providers on the web. *IEEE Transactions on Knowledge and Data Engineering*, 20, 6, 796–808.

**Markos Kramer** (nacilingva ortografio: Marcos Cramer) estas scienca kunlaboranto en la komputoscienca fakultato de la Universitato de Luksemburgo. De 2003 ĝis 2007, Kramer bakalaŭre kaj magistre studis matematikon kaj filozofion en Oksfordo, Britujo. De 2008 ĝis 2013 li doktoriĝis en Bonno (Germanujo) pri temo, kiu ligas matematikan logikon kaj komputan lingvosciencan, kaj ekde 2013 li esploras komputosciencajn logikojn en Luksemburgo. Inter 2006 kaj 2009 Kramer estis la redaktisto de *TEJO Tutmonde*. De 2010, li membras en la Akademio de Esperanto, kaj ekde 2013 li estas la direktoro de ĝia Sekcio pri Ĝenerala Vortaro.