Խոսույթի վերլուծում

Սա [Jur2024] Խոսույթի Կապակցվածություն *հատվածի կարճ շարադրումն է։*

ԲՈՎԱՆԴԱԿՈՒԹՅՈՒՆ

Տեսություն 1

Տեղական եւ գլոբալ կապակցվածություն 1

Կապակցվածության հարաբերությունը 3

Հոետորական կառուցվածքի տեսություն (RST) 3

Փենսիլվանիայի համալսարանի խոսույթային ծառադարան (PDTB) 4

Խոսույթի կառուցվածքի վերլուծություն 4

EDU զատումը RST վերլուծության համար 5

RST վերլուծում 5

Խոսույթի PDTB վերլուծում 6

Ամփոփում 7

Տեսություն

Մենք խոսույթ ենք անվանում կապակցված կառուցվածքով նախադասություն-ների խումբ, եւ «կապակցվածություն» բառն օգտագործում ենք նախադասությունների միջեւ այն փոխհարաբերությանը անդրադառնալու համար, որը իրական խոսույթները տարբերում է նախադասությունների պատահական համակցություններից։

Տեղական եւ գլոբալ կապակցվածություն

Իրական խոսույթները ցուցադրում են ինչպես տեղական, *այնպես էլ* գլոբալ *կա-պակցվածություն։ Եկեք քննարկենք երեք եղանակ, որոնցով իրական խոսույթներում ապահովվում է «տեղական կապակցությունը».*

Նախ, իրական խոսույթներում անկախ կամ ստորադաս նախադասությունները համակարգված ձեւով են կապված մոտակայքի նախադասությունների հետ։ Դիտար-կենք Հոբսի (1979) այս օրինակը.

(24.1) Ջոնը գնացքով Փարիզից գնաց Ստամբուլ։ Նա սիրում է սպանախ։

Այս հաջորդականությունը անհամատեղելի է, քանի որ ընթերցողի համար *անհասկանալի է, թե ինչո՞ւ է երկրորդ նախադասությունը հաջորդում առաջինին. ի՞նչ կապ ունի սպանախ սիրելը գնացքով ուղեվորվելու հետ։*

Դրան հակառակ, հետեւյալ կապակցված օրինակում՝

(24.2) Ջեյնը գնացքով Փարիզից գնաց Ստամբուլ։ Նա պետք է մասնակցեր համաժողովի։

*Երկրորդ նախադասությունը տալիս է ՊԱՏՃԱՌԸ առաջինում բերված Ջեյնի արարքի։ Տեքստային միավորների միջեւ գոյություն ունեցող կառուցվածքային հարաբերությունը, ինչպիսին է օրինակ ՊԱՏՃԱՌԸ, կոչվում է **կապակցված հարաբերություն։** Կապակցված խոսույթները կառուցվում են բազմաթիվ նման հարաբերություններով։*

*Երկրորդ եղանակը, որով խոսույթը կարող է տեղայնորեն կապակցված լինել, դա ինչ-որ մեկի կամ ինչ-որ բանի «մասին» լինելու շնորհիվ է։ Կապակցված խոսույթում որոշ էություններ ակնառու կարեվոր են, եւ խոսույթը կենտրոնանում է դրանց վրա եւ չի տատանում ուրիշների միջեւ։ Սա կոչվում է **էության-հենքով** հետեւողականություն։ Դիտարկենք հետեւյալ անկապ հատվածը, որտեղ **ակնառու** էության դերը կտրուկ անցնում է Ջոնից Ջենին, հետո դաշնամուրների խանութին, հետո՝ հյուրասենյակին, հետո՝ նորից Ջենիին, ապա կրկին դաշնամուրին.*

(24.3) Ջոնը ուզում էր դաշնամուր գնել իր հյուրասենյակի համար։ Ջենին նույնպես ուզում էր դաշնամուր գնել։ Նա գնաց դաշնամուրի խանութ, որը մոտակայքում էր գտնվում։ Հյուրասենյակը երկրորդ հարկում էր։ Նա չգտավ իրեն դուր եկած որեւէ բան։ Նրա գնած դաշնամուրը դժվար էր երկրորդ հարկ բարձրացնել։

*Էության-հենքով մոդելները հետեւողականությունը գնահատում են խոսույթի երկայնքում առկա **ակնառու** էություններին հետեւելով։ Օրինակ՝ **կենտրոնացման տեսությունը*** (Grosz et al., 1995), *որը էության-հենքով կապակցվածության ամենաազդեցիկ տեսությունն է, հետեւում է, թե խոսույթի մոդելում որ էություններն են ակնառու ցանկացած պահին (ակնառու էությունները ավելի հավանական է, որ դերանվանականացվեն կամ հայտնվեն ակնառու շարահյուսական դիրքերում, ինչպիսիք են ենթական կամ խնդիրը)։ Կենտրոնացման տեսության մեջ անցումները նույն ակնառու էությունը ունեցող նախադասությունների միջեւ համարվում են ավելի կապակցված, քան նրանց միջեւ, որոնցում էությունները անընդհատ տեղաշարժվում են։ Լայնորեն օգտագործվող էության կապակցվածության քառակուսի-ցանցային* (grid) *մոդելը* (Barzilay and Lapata, 2008) *իրականացնում է կենտրոնացման տեսության գործիքակազմի ներըմբռնումները։*

*Վերջապես, խոսույթները կարող են լինել տեղային-կապակից, **թեմատիկ կապակցվածության** շնորհիվ։ Այդ դեպքում մերձակա նախադասությունները, որպես կանոն, նույն թեմայի մասին են։ Թեմաները քննարկվում են նույն կամ նմանատիպ բառապաշար օգտագործելով։ Քանի որ թեմատիկորեն կապակցված խոսույթները հիմնված են մեկ իմաստաբանական դաշտի կամ թեմայի վրա, դրանք հակված են ցուցադրել **բա-***

ռապաշարային կապակցվածություն *կոչվող մակերեսային հատկությունը* (Halliday and Hasan, 1976)*. մոտակա նախադասություններում նույնական կամ իմաստային առումով կապված բառեր են կիսում։ Օրինակ, այն փաստը, որ տուն, բուխարի, ձեղնահարկ, պահարան եւ պատուհան բառերը, որոնք պատկանում են նույն իմաստաբանական դաշտին, հանդիպում են (24.4)-ի երկու նախադասություններում, կամ որ դրանք կիսում են նույն կղմինդրապատ բառը, ցույց է տալիս, որ երկուսն էլ կապված են որպես խոսույթ.*

(24.4) Ձմռանից առաջ ես բուխարի կառուցեցի եւ կղմինդրով պատեցի իմ տան կողմերը... Այսպիսով, ես ունեմ ամուր կղմինդրով պատված եւ սվաղված տուն... ձեղնահարկով եւ պահարանով, մեծ պատուհանով յուրաքանչյուր կողմում...

Հարակից կամ մոտակա նախադասությունների միջեւ տեղական կապակցվածությունից բացի, խոսույթները նաեւ ցուցաբերում են գլոբալ կապակցվածություն։ Տեքստի շատ ժանրեր կապված են որոշակի ավանդական խոսույթային կառուցվածքների հետ։ Ակադեմիական հոդվածները կարող են ունենալ բաժիններ, որոնք նկարագրում են մեթոդաբանությունը կամ արդյունքները։ Պատմությունները կարող են հետեւել ավանդական սյուժետային գծերին կամ մոտիվներին։ Համոզիչ էսեներն ունեն որոշակի պնդում, որը փորձում են փաստարկել, եւ էսեն կարող է արտահայտել այդ պնդումը՝ կառուցվածքային նախադրյալների ամբողջության հետ միասին, որոնք հաստատում են փաստարկը եւ ապամոնտաժում հնարավոր հակափաստարկները։

Կապակցվածության հարաբերությունը

Խոսույթի տեքստային հատվածների միջեւ կապերը կարող են ներկայացվել որպես ***կապակցվածության հարաբերությունների*** *ամբողջություն։ Հաջորդ երկու բաժինները նկարագրում են կապակցվածության հարաբերությունների եւ դրանց հետ կապված կորպուսների երկու լայնորեն օգտագործվող մոդելներ՝ Հռետորական կառուցվածքի տեսությունը* (RST) *եւ Փենսիլվանիայի համալսարանի խոսույթային ծառադարանը* (PDTB)*։*

Հռետորական կառուցվածքի տեսություն (RST)

RST-*ում հարաբերությունները սահմանվում են տեքստի երկու հատվածների՝ սովորաբար* միջուկի *եւ* արբանյակի *միջեւ։ Միջուկը այն միավորն է, որն առանցքային է հեղինակի նպատակի համար եւ անկախ մեկնաբանելի է. արբանյակն առանցքային չէ եւ, որպես կանոն, մեկնաբանելի է միայն միջուկի նկատմամբ։*

Գլխավոր, երկրորդական նախադասություններով կամ բառակապակցություններով ծավալվող տեքստային հատվածները RST-*ում կոչվում են* ***խոսույթի տարրական***

միավորներ (elementary discourse units) *կան* EDU-*ներ։ Այդ միավորները կարող են նաեւ անվանվել* ***խոսույթի հատվածներ****։ Քանի որ դրանք կարող են համապատասխանել տեքստի կամայական հատվածների,* EDU-*ի սահմանների որոշումը կարեվոր խնդիր է դառնում կապակցվածության հարաբերություններ գտնելու համար։ Կոպիտ ասած, խոսույթի հատվածները կարելի է դիտարկել որպես նախադասության շարահյուսության բաղադրիչների նմանակներ, եւ իրականում, ինչպես կտեսնենք 24.2 բաժնում, մենք սովորաբար դիմում ենք վերլուծական ալգորիթմների խոսույթի կառուցվածքը արտածելու համար։*

Փենսիլվանիայի համալսարանի խոսույթային ծառադարան (PDTB)

PDTB *պիտակավորումը հիմնվում է բառապաշարի վրա։ Փոխանակ պիտակավորողներին խնդրել, որ ուղղակիորեն նշեն տեքստի կապակցվածությունը, նրանց խոսքային հարաբերություններ նշող շաղկապների (լայն իմաստով), ինչպիսին են օրինակ՝ որովհետեւ, չնայած, երբ, քանի որ կամ որպես արդյունք ցանկ է տրվում։ Տեքստի այն մասում, որտեղ այդ բառերը նշում են երկու տեքստային երկայնքների միջեւ կապ, կապակցող բառը եւ երկայնքները այնուհետեւ մեկնաբանվում են, ինչպես 24.13 օրինակում։ Այդտեղ արտահայտությունը որպես արդյունք ազդարարում է պատճառահետեւանքային կապ այն բանի միջեւ, ինչը* PDTB-*ն անվանում է* Arg1 *(առաջին երկու նախադասությունները, այստեղ՝ թեքատառով) եւ* Arg2 *(երրորդ նախադասությունը, այստեղ՝ թավատառով) միջեւ։*

(24.13) Համալիր խանութներում զարդերի ցուցափեղկերը հաճախ խառնաշփոթ էին եւ անոգեշնչող։ Եվ ապրանքները, իհարկե, կեղծ էին։ ***Արդյունքում, կեղծ թանկարժեք քարերի մարքեթինգով զբաղվողները աստիճանաբար կորցնում էին իրենց տեղը հանրախանութներում՝ հօգուտ ավելի նորաձեւ մրցակիցների՝ կոսմետիկայի արտադրողների։***

(24.14) Հուլիսին Շրջակա միջավայրի պաշտպանության գործակալությունը աստիճանաբար արգելք դրեց ասբեստի գրեթե բոլոր օգտագործման վրա։ (անուղղակի = որպես արդյունք) ***Մինչեւ 1997 թվականը քաղցկեղ առաջացնող ասբեստի գրեթե բոլոր մնացած կիրառումները ապօրինի կհամարվեն։***

Խոսույթի կառուցվածքի վերլուծություն

Տրված նախադասությունների հաջորդականությունը գննելով՝ ինչպե՞ս կարող ենք ավտոմատ կերպով որոշել դրանց միջեւ կապակցվածության հարաբերությունները։ Այս խնդիրը հաճախ անվանում են ***խոսույթի վերլուծություն*** *(չնայած* PDTB-*ի*

դեպքում մենք պիտակներ ենք վերագրում միայն տերեւների երկայնքով եւ չենք կառուցում ամբողջական վերլուծական ծառ, ինչպես անում ենք RST-ի դեպքում)։

EDU զատումը RST վերլուծության համար

RST վերլուծությունը սովորաբար կատարվում է երկու փուլով։ Առաջին փուլը՝ EDU հատվածումը, որոշվում է յուրաքանչյուր EDU-ի սկիզբը եւ վերջը։ Այս փուլի արդյունքը կլինի հետեւյալի նման պիտակավորում.

(24.16) [Պարոն Ռեմբոն ասում է]*e1* [nr 3.2 ակր տարածքի]*e2* [Սան Ֆեռնանդո հովտին նայող]*e3* [գինը 4 միլիոն դոլար է]*e4* [քանի որ այնտեղ ապրել է հանգուցյալ դերասան էրոլ Ֆլինը]*e5*

Քանի որ EDU-ները մոտավորապես համապատասխանում են նախադասություններին, EDU զատման վաղ մոդելները նախ կիրառում էին շարահյուսական վերլուծիչ, ապա հետմշակում արդյունքը։ Ժամանակակից համակարգերը սովորաբար օգտագործում են նեյրոնային հաջորդական մոդելներ, որոնք ուսանվում են EDU զատման վերահսկելի տվյալների ոսկե հավաքածուներով[1], ինչպիսին է, օրինակ՝ RST խոսույթային ծառադարանը։

RST վերլուծում

Խոսույթի համար RST կապակցված կառուցվածք ստանալու գործիքները մինչ օրս հիմնվել են շարահյուսական վերլուծության ալգորիթմների վրա, ինչպիսին է shift-reduce վերլուծումը (Marcu, 1999)։ Ջիից եւ Այզենշտեյնից (2014) ի վեր ժամանակակից RST վերլուծիչներ օգտագործում են 19-րդ գլխում բերված նեյրոնային շարահյուսական վերլուծիչները՝ օգտագործելով ներկայացուցչական ուսանում՝ խոսույթի յուրաքանչյուր հատվածի ներկայացումներ կառուցելու համար, եւ մարզելով վերլուծիչը ճիշտ shift (շեղում) եւ reduce (նվազեցում) գործողությունների ընտրություն անել, հիմնվելով մարզման ոսկե հավաքածուի վրա։ Մենք կնկարագրենք Յուի եւ այլոց (2018) shift-reduce վերլուծիչը։ Վերլուծիչի վիճակը բաղկացած է Դեզ [Հայ2022։։339] եւ Հերթ[2] տվյալների կառույցներից, որոնք ստեղծվում են վիճակների վրա մի շարք գործողություններ կատարելով։ Գործողությունները ներառում են՝

• shift. Հերթի առաջին EDU-ն տեղափոխում է Դեզ՝ ստեղծելով մեկ հանգույցից բաղկացած ենթածառ։

---

[1] Ոսկե տվյալներ են կոչվում ավտոմատ կամ ձեռքով պատրաստված բարձրորակ, 100% վստահելի տվյալներ, որոնք օգտագործվում են ՄՈ, տեստավորման, կամ համեմատման համար։

[2] Հերթը ավելի բնական եւ հասկանալի տվյալների տարրերի կառույց է։ Հերթից տվյալները հեռացվում են որեւէ գործողություն կատարելու (սպասարկման) համար «բնական», ինչպես ոչ հայկական խանութի հերթում, հաջորդականությամբ՝ առաջին եկածը առաջինն է սպասարկվում։

• reduce(l,d). *միավորում է Դեզի վերին երկու ենթածառերը, որտեղ* l*-ը կապակցվածության հարաբերության պիտակն է, իսկ* d*-ն՝ միջուկայնության ուղղությունը,* d ∈ {NN,NS,SN}[3]

Խոսույթի PDTB վերլուծում

Խոսույթի PDTB *վերլուծումը, այսինքն տեքստի երկայնքների միջեւ* PDTB *կապակցվածության հարաբերությունները հայտնաբերելու խնդիրը, երբեմն անվանում է* ***խոսույթի մակերեսային վերլուծում****, քանի որ ներառում է միայն տեքստային երկայնքների պարզ հարաբերությունները, այլ ոչ թե* RST *վերլուծության ամբողջական ծառերը։ Լինը եւ այլոք (2014) մշակվել են* PDTB *խոսույթի վերլուծության չորս ենթաքայլերը՝ առաջին ամբողջական համակարգը, ուղղակի (քայլեր 1-3) եւ անուղղակի (քայլ 4) կապակցող բառերը գտնելու համար.*

1. Գտեք խոսույթային կապակցող բառերը (տարբերակելով դրանք ոչ խոսույթայիններից)

2. Յուրաքանչյուր կապակցող բառի համար գտեք երկու երկայնքները

3. Պիտակավորեք այդ երկայնքների միջեւ եղած հարաբերությունը

4. Յուրաքանչյուր հարակից նախադասությունների զույգի հարաբերությունը հատկանշեցեք

4-րդ քայլի համար առաջարկվել են բազմաթիվ մոտեցումներ՝ մուտք ընդունելով հարակից նախադասությունների զույգը, ելքային արժեք նշանակել կապակցվածության հարաբերության իմաստի պիտակը։

Իրացումը հաճախ հետեւում է Լինի եւ այլոց (2009) աշխատանքին, ենթադրելով ոսկե նախադասության երկայնքի սահմանները եւ յուրաքանչյուր հարակից երկայնքին նշանակել 11 երկրորդ կարգի PDTB *պիտակներից մեկը կամ ոչ մեկը։*

Մյուս բոլոր քայլերը նույնպես դիտարկվել են։ 1-ն քայլը խոսույթային կապակցող բառերի տարանջատումն է ոչ խոսույթայիններից։ Օրինակ, ինչպես նշում են Փիթլերը եւ Նենկովան (2009), «եւ»-ը խոսույթային շաղկապ է, որը կապում է երկու նախադասությունները մանրամասնող/ընդլայնող հարաբերությամբ (24.24)-ում, մինչդեռ (24.25)-ում այն ոչ խոսույթային NP *շաղկապ է։*

(24.24) Վաճառքը արագացավ, քանի որ նախորդ գնորդները հրաժարվեցին իրենց դիրքերից, եւ ագրեսիվ կարճ վաճառողները՝ հետագա անկումը կանխատեսելով, մտան խաղի մեջ։

(24.25) Իմ սիրելի գույները կապույտն ու կանաչն են։

---

[3] *Օգտագործվում է ծածկագրիչ-վերծանիչ ճարտարապետությունը [Հայ2022։։325-329]; Որոշ սահմանումների բացատրություններ (մասնավորապես՝* միջուկայնության ուղղություն*) չգտա. Երեւի պետք է* Yu et al. (2018)*-ում փնտրել։*

Նմանապես, «once»*-ը (24.26)-ում խոսույթային շաղկապ է, որը ցույց է տալիս ժամանակային հարաբերություն, բայց (24.27)-ում պարզապես ոչ խոսույթային մակբայ է, որը նշանակում է «նախկինում» եւ փոփոխում է* «used»*-ի իմաստը:*

(24.26) The asbestos fiber, crocidolite, is unusually resilient once it enters the lungs, with even brief exposures to it causing symptoms that show up decades later, researchers said. {Ասբեստի մանրաթելը՝ կրոկիդոլիտը, անսովոր դիմացկուն է թոքեր մտնելուց հետո, եւ նույնիսկ դրա հետ կարճատեւ շփումը առաջացնում է ախտանիշներ, որոնք ի հայտ են գալիս տասնամյակներ անց, ասում են հետազոտողները:}

(24.27) A form of asbestos once used to make Kent cigarette filters has caused a high percentage of cancer deaths among a group of workers exposed to it more than 30 years ago, researchers reported. {Հետազոտողները հայտնել են, որ ասբեստի մի տեսակ, որը ժամանակին օգտագործվում էր Kent ծխախոտի ֆիլտրեր պատրաստելու համար, քաղցկեղից մահացության բարձր տոկոսի պատճառ է դարձել ավելի քան 30 տարի առաջ դրան ենթարկված աշխատողների խմբի շրջանում:}

Այսպիսով՝ խոսույթային կապակցող լինելու որոշումը, բառի երկիմաստության առանձնահատուկ դեպքի տարբերակում է: Երկիմաստության վերաբերյալ վաղ շրջանի աշխատանքները ցույց են տվել, որ PDTB*-ի 4 բարձր մակարդակի իմաստային դասերը կարող են տարբերակվել բարձր (94%) ճշգրտությամբ՝ օգտագործելով ոսկե վերլուծված ծառերի շարահյուսական առանձնահատկություններից* (Pitler and Nenkova, 2009):

Ամփոփում

Այս գլխում մենք ներկայացրեցինք խոսույթի կապակցվածության տեղական եւ գլոբալ մոդելները:

• *Խոսույթները նախադասությունների կամայական հավաքածուներ չեն, դրանք պետք է լինեն կապակցված: Խոսույթը կապակցված դարձնող գործոնների շարքում են նախադասությունների միջեւ կապակցվածության հարաբերությունները, էության-հենքով կապակցվածությունը եւ թեմատիկ կապակցվածությունը:*

• *Առաջարկվել են կապակցվածության եւ հռետորական հարաբերությունների զանազան բազմություններ: Հռետորական կառուցվածքի տեսության* (RST) *մեջ հարաբերությունները պահպանվում են տեքստի հատվածների միջեւ եւ ներկայացվում են ծառի տեսքով: Այդ պատճառով սովորաբար տեղաշարժ-կրճատում* (shift-reduce) *եւ այլ վերլուծական ալգորիթմներ են օգտագործվում այդ կառուցվածքները վերագրելու համար: Փենսիլվանիայի համալսարանի խոսույթային ծառադարանում* (The Penn Discourse Tree-

bank - PDTB) *պիտակավորվում են միայն հատվածների զույգերի միջեւ հարաբերությունները, եւ պիտակները սովորաբար վերագրվում են հաջորդականության մոդելներով*[4]*։*

• *Էության-հենքով կապակցվածությունը արտացոլում է այն միտումը, որ խոսույթները մի որեւէ էության մասին եւ շարունակում են հիշատակել այդ էությունը նախադասությունից նախադասություն։ Կենտրոնացման տեսությունը մոդելների ընտանիք է, որը նկարագրում է, թե ինչպես է մոդելավորվում խոսույթի էության կարեվորությունը, եւ, հետեւաբար, ինչպես է կապակցվածությունը ձեռք բերվում նույն էությունը խոսույթի ընթացքում ակնառու պահելու միջոցով։ Էության ցանցային մոդելը միջոց է վարից-վեր հաշվարկելու, թե որ էության իրականացման անցումներ են հանգեցնում կապակցվածությանը։*

• *Բազմաթիվ ժանրերում գլոբալ կապակցվածության յուրովի տեսակներ են գործածվում։ Համոզիչ հոդվածներն ունեն պնդումներ եւ սկզբունքներ, որոնք վերցվում են փաստարկների ոլորտից, գիտական հոդվածների կառուցվածքը պայմանավորվում է նպատակներով, մեթոդներով, արդյունքներով, եւ համեմատություններով։*

---

[4] *Նեյրոնային նման ցանցի ճարտարապետության եւ աշխատանքի մանրամասները տես [Հայ2022։ ։314-332]։*