

**ORIENTAL JOURNAL OF PHILOLOGY**

journal homepage:

<http://www.supportscience.uz/index.php/ojp/about>**EMPIRICAL VALIDATION OF LINGUISTIC ANNOTATION CRITERIA IN LITERARY TEXT MATERIALS: A MULTI-LEVEL PILOT STUDY****Mohinur Alikulova***Tashkent University of Applied Sciences**Teacher of Foreign Languages and Literature**Science ID: MSD-0426-0088**Tashkent, Uzbekistan***ABOUT ARTICLE**

**Key words:** linguistic annotation; literary text; criteria validation; inter-annotator agreement; multi-level analysis; Uzbek literature; digital humanities.

**Received:** 01.08.26**Accepted:** 03.08.26**Published:** 03.08.26

**Abstract:** The multifaceted nature of literary texts and their implicit meanings require a rigorous empirical verification of linguistic annotation criteria when converting them into a digital format. This article presents the results of a pilot empirical study aimed at verifying the practical viability of multi-level linguistic annotation criteria applied to literary materials. A pilot corpus consisting of approximately 52,000 word-forms, selected from 20th-century Uzbek prose and poetry, was analyzed by four specially trained annotators across seven levels: lexico-morphological, syntactic, semantic, pragmatic, stylistic, discursive, and conceptual-cognitive. The study was based on five core management criteria: explicitness, layer independence, cross-level flexibility, interpretive transparency, and scalability. The inter-annotator agreement rate was measured using Cohen's kappa, Krippendorff's alpha, and inter-segment F1 coefficients. The data obtained from the pilot study demonstrated a differentiated nature of agreement indicators, which serves as an objective basis for the targeted refinement of the initial criteria.

**BADIIY MATN MATERIALLARIDA LINGVISTIK ANNOTATSIYA MEZONLARINING EMPIRIK VALIDATSIYASI: KO'P SATHLI PILOT TADQIQOT****Mohinur Aliqulova***Toshkent Amaliy fanlar universiteti**Xorijiy til va adabiyoti o'qituvchisi**Ilmiy identifikatsiya raqami: MSD-0426-0088*

---

**MAQOLA HAQIDA**


---

**Kalit so'zlar:** lingvistik annotatsiya; badiiy matn; mezonlar validatsiyasi; annotatorlararo kelishuv; ko'p sathli tahlil; o'zbek adabiyoti; raqamli gumanitar fanlar.

**Annotatsiya:** Badiiy matnlarning ko'p qirrali tabiati va tarkibidagi yashirin ma'nolar ularni raqamli formatga o'tkazishda lingvistik annotatsiya mezonlarini qat'iy empirik tekshirishni talab etadi. Ushbu maqolada ko'p sathli lingvistik annotatsiya mezonlarining badiiy materiallarga tatbiq etilgandagi amaliy hayotiyligini tekshirishga qaratilgan pilot empirik tadqiqot natijalari bayon etilgan. XX asr o'zbek nasri va nazmidan saralab olingan, taxminan 52 ming so'zshakldan iborat pilot korpus maxsus tayyorgarlikdan o'tgan to'rtta annotator tomonidan yettita sath (leksik-morfologik, sintaktik, semantik, pragmatik, stilistik, diskursiv va konseptual-kognitiv) bo'yicha tahlil qilindi. Bunda beshta asosiy boshqaruv mezonlari: eksplitsitlik, qatlamlar mustaqilligi, sathlararo moslashuvchanlik, interpretiv shaffoflik va miqyoslanuvchanlik asos qilib olindi. Annotatorlararo kelishuv darajasi Koen kappasi, Krippendorff alfasi va segmentlararo F1 koeffitsiyentlari yordamida o'lchandi. Pilot tadqiqot natijasida olingan ma'lumotlar kelishuv ko'rsatkichlarining tabaqalashgan xarakterini ko'rsatdi va bu dastlabki mezonlarni maqsadli ravishda takomillashtirish uchun obyektiv asos bo'lib xizmat qiladi.

---

**ЭМПИРИЧЕСКАЯ ВАЛИДАЦИЯ КРИТЕРИЕВ ЛИНГВИСТИЧЕСКОЙ  
АННОТАЦИИ В МАТЕРИАЛАХ ХУДОЖЕСТВЕННЫХ ТЕКСТОВ:  
МНОГОУРОВНЕВОЕ ПИЛОТНОЕ ИССЛЕДОВАНИЕ**

**Мохинур Аликулова**

*Ташкентский университет прикладных наук*

*Преподаватель иностранного языка и литературы*

*Научный идентификационный номер: MSD-0426-0088*

*Ташкент, Узбекистан*

---



---

**О СТАТЬЕ**


---

**Ключевые слова:** лингвистическая аннотация; художественный текст; валидация критериев; межаннотаторское согласие; многоуровневый анализ; узбекская литература; цифровые гуманитарные науки.

**Аннотация:** Многогранная природа художественных текстов и их скрытые смыслы требуют строгой эмпирической проверки критериев лингвистической аннотации при их переводе в цифровой формат. В данной статье представлены результаты пилотного эмпирического исследования, направленного на проверку практической жизнеспособности критериев

---

многоуровневой лингвистической аннотации применительно к художественным материалам. Пилотный корпус объемом около 52 тысяч словоформ, отобранный из узбекской прозы и поэзии XX века, был проанализирован четырьмя специально подготовленными аннотаторами по семи уровням: лексико-морфологическому, синтаксическому, семантическому, прагматическому, стилистическому, дискурсивному и концептуально-когнитивному. При этом за основу были взяты пять основных критериев управления: эксплицитность, независимость уровней, межуровневая гибкость, интерпретационная прозрачность и масштабируемость. Уровень межаннотаторского согласия измерялся с помощью каппы Коэна, альфы Криппендорфа и межсегментных коэффициентов F1. Данные, полученные в результате пилотного исследования, продемонстрировали дифференцированный характер показателей согласия, что служит объективной основой для целенаправленного совершенствования исходных критериев.

**Kirish.** Nazariy jihatdan shakllantirilgan annotatsiya tizimidan uni amaliy joriy etish bosqichiga o'tish kompyuter va korpus tilshunosligida badiiy matnlarni tadqiq qilishning doimiy muammolaridan biri hisoblanadi. Har qanday nazariy model qanchalik puxta ishlab chiqilgan bo'lmasin, uning metodologik qonuniyligi faqat amaliy qo'llash sinovidan o'tgandagina tasdiqlanadi. Ya'ni, bir-biridan mustaqil bo'lgan annotatorlar ayni bir material ustida, yagona yo'riqnoma asosida ishlagan holda bir xil yoki o'xshash tahliliy xulosalarga kela olishi lozim [1;557-b.]. Ushbu talab badiiy matnlar kontekstida yanada ulkan ahamiyat kasb etadi; [2;102-b.]. chunki badiiy asarlarning ko'p qatlamli semiotik zichligi gazeta, parlament yoki texnik korpuslarda odatda qo'llaniladigan standartlashtirish qoidalariga bo'ysunmaydi [3;312-b.]. So'nggi yillardagi ilmiy izlanishlar annotatsiya metodologiyasining markaziga aynan empirik validatsiya jarayonini qo'ymoqda. Masalan, Metaforani aniqlash tartibi [5;3-b] va uning kengaytirilgan varianti bo'lgan MIPVU [6;25-b] kabi protseduralar interpretatsiya jihatidan murakkab bo'lgan toifalarni aniq qaror qabul qilish protokollari va annotatorlarni bosqichma-bosqich o'qitish orqali qanday qilib amaliy tizimga aylantirish mumkinligini yaqqol ko'rsatib berdi. Shu bilan bir qatorda,

kelishuv statistikasi bo'yicha olib borilgan metodologik ishlar — xususan,  $\backslash(\kappa\alpha\pi\alpha\backslash)$ -oilasi va  $\backslash(\alpha\backslash)$ -oilasi koeffitsiyentlarining tizimli qiyoslanishi — badiiy annotatsiyaga xos bo'lgan ko'p kodlovchili nominal sharoitlarda ishonchlilik darajasini qanday o'lchash kerakligiga aniqlik kiritdi [4;221-b].

Mavzuning dolzarbligi raqamli gumanitar fanlar rivoji bilan yanada ortmoqda. Biroq, ushbu metodologik yutuqlarga qaramay, ingliz tilidan tashqaridagi badiiy korpuslarda ko'p sathli mezonlarni empirik tekshirgan tadqiqotlar hali ham juda kam sonli hisoblanadi. Turkiy tillar oilasi, xususan, agglyutinativ morfologiya va o'ziga xos madaniy-stilistik qoliqlar tufayli tayyor g'arbiy sxemalarni to'g'ridan-to'g'ri qabul qilishni qiyinlashtiradigan o'zbek tili uchun bunday ishlar deyarli yo'q darajada [16;79-b]. Ushbu tadqiqot qayd etilgan ilmiy bo'shliqni to'ldirishga qaratilgan bo'lib, XX asr o'zbek badiiy adabiyoti materiallariga tatbiq etilgan ko'p sathli annotatsiya tizimining empirik tahlilini taqdim etadi.

**Tadqiqot metodologiyasi.** Mazkur tadqiqot doirasida raqamlashtirish fenomenining mohiyatini fundamental darajada ochib berish maqsadida, tizimli-ko'p bosqichli metodologik interpretatsiya strategiyasi qo'llanildi. Izlanishning konseptual poydevori filologiya, semiotik tahlil hamda raqamli gumanitar fanlar integratsiyalashuvi natijasida yuzaga kelgan fanlararo paradigmalar majmuasiga tayanadi. Taklif etilayotgan tizimning beshta asosiy mezoni (eksplitsitlik, qatlamlar mustaqilligi, sathlararo moslashuvchanlik, interpretiv shaffoflik va miqyoslanuvchanlik) real badiiy matnlarda qo'llanilganda yettita annotatsiya sathi bo'yicha qanchalik amaliy jihatdan hayotiy ekanligini aniqlashdan iborat.

Tadqiqotning vazifalari: 1. XX asr o'zbek nasri va nazmining eng sara namunalarini o'z ichiga olgan vakillik xususiyatiga ega pilot korpusni shakllantirish. 2. Yuqorida qayd etilgan beshta mezonni aniq amaliy annotatsiya yo'riqnomalari ko'rinishida shakllantirish. 3. Har bir sath bo'yicha annotatorlararo kelishuv ko'rsatkichlarini matematik-statistik baholash. 4. Metodologik jihatdan qayta ishlash va takomillashtirishni talab qiladigan mezon hamda sathlarni aniqlash.

Ushbu tadqiqot badiiy matnlarni annotatsiya qilishda lingvistik-morfologik va sintaktik sathlarda yuqori, kontseptual-kognitiv sathda esa metodologik jihatdan asoslangan kelishuvni qayd etadi. Tadqiqot uchta asosiy ilmiy savolga javob beradi: annotatsiya mezonlarining samaradorligi, sathlararo farqlar va empirik material asosida tizimni takomillashtirish [4;65-b]. Tadqiqot 20-asr o'zbek adabiyotidan 52 ming so'zshakldan iborat korpusda, 4 nafar mutaxassis va TEI-XML standarti yordamida o'tkazilgan [4;77-78-b].

**Tahlil va Natijalar.** Olib borilgan pilot annotatsiya jarayoni aniq tabaqalangan kelishuv modelini namoyon etdi (1-jadval). Leksik-morfologik sath barcha uchta metrika bo'yicha eng yuqori ko'rsatkichlarni qayd etdi va undan keyingi o'rinni biroz farq bilan sintaktik sath egalladi. Semantik va stilistik sathlar o'rtacha diapazonni qamrab oldi. Pragmatik, diskursiv va konseptual-

kognitiv sathlar esa, dastlabki ishchi gipotezaga muvofiq, pastroq, ammo metodologik jihatdan talqin qilinadigan qiymatlarni ko'rsatdi.

*1-jadval. Yettita annotatsiya sathi bo'yicha annotatorlararo kelishuv ko'rsatkichlari.*

| Annotatsiya sathi   | Koen $\kappa$ (Kappa) | Krippendorff $\alpha$ (Alfa) | Segmentlararo $F_1$ (Span $F_1$ ) |
|---------------------|-----------------------|------------------------------|-----------------------------------|
| Leksik-morfologik   | 0.91                  | 0.89                         | 0.93                              |
| Sintaktik           | 0.84                  | 0.81                         | 0.86                              |
| Semantik            | 0.72                  | 0.69                         | 0.74                              |
| Pragmatik           | 0.61                  | 0.58                         | 0.63                              |
| Stilistik           | 0.66                  | 0.63                         | 0.68                              |
| Diskursiv           | 0.58                  | 0.55                         | 0.60                              |
| Konseptual-kognitiv | 0.51                  | 0.48                         | 0.54                              |

Ushbu raqamlarni metodologik adabiyotlarda keltirilgan tayanch mezonlar bilan qiyoslash olingan natijalarni samarali baholash imkonini beradi [4;277-b]. Odatda, 0.80 dan yuqori koeffitsiyentlar sezilarli darajadan deyarli mukammalgacha bo'lgan kelishuvni, 0.60 va 0.80 oralig'idagilar barqaror kelishuvni, 0.40 va 0.60 gacha bo'lgan koeffitsiyentlar esa mo'tadil muvofiqlikni anglatadi [1;576-b].

Mazkur shkala bo'yicha yondashilsa, quyi sathlar yuqori ishonchlilik diapazoniga, semantik va stilistik sathlar barqaror kelishuv doirasiga, konseptual-kognitiv sath esa mo'tadil darajaga to'liq mos keladi. Eng muhimi, yuqori interpretiv sathlardagi bunday mo'tadil qiymatlar ko'chma ma'noli til birliklarini aniqlash bo'yicha amaldagi empirik jarayonlar natijalariga o'xshashdir; masalan, MIPVU doirasidagi tadqiqotlarda  $\alpha$  koeffitsiyenti muntazam ravishda 0.55–0.65 oralig'ida qayd etiladi [6;99-b].

Raqamli xulosalardan tashqari, annotatorlararo kelishmovchilik holatlarini sifat jihatidan tahlil qilish belgilangan mezonlarning ishlash mexanizmi haqida chuqurroq tasavvur beradi. Leksik-morfologik sathda deyarli barcha nomuvofiqliklar agglyutinatив so'z shakllaridagi chegara hodisalari, xususan, klitikalarining segmentatsiyasi va qo'shma affikslar ketma-ketligini tahlil qilish bilan bog'liq bo'ldi. Bunday holatlar morfologik jihatdan boy tillar uchun analitik tillar asosida ishlab chiqilgan tayyor kategoriyalar inventaridan tashqariga chiquvchi, maxsus ichki yo'riqnomalar zarurligi haqidagi ilmiy qarashlarni yana bir bor tasdiqlaydi [16;84-b].

Sintaktik sathda kelishmovchiliklarning asosiy o'chog'i badiiy uslubga, ayniqsa she'riyatga xos bo'lgan inversiyalar va ellipsislarni tahlil qilish jarayonida kuzatildi. Annotatorlar matndagi normadan og'ishlarni aniq belgilay olishgan bo'lsa-da, bu og'ishni sintaktik sathga yoki stilistik

sathga birlashtirish masalasida turlicha qaror qabul qilishdi. Ushbu konvergent noaniqlik sathlararo moslashuvchanlik tamoyili orqali hal qilinishi lozim bo'lgan samarali ziddiyatni yuzaga keltirdi: model matnni faqat bitta sathga bog'lashga majburlash o'rniga, har ikkala sathda parallel ravishda annotatsiya qilishni va ularni aniq o'zaro havola orqali birlashtirishni ta'minlaydi.

Semantik sathdagi ixtiloflar ko'p ma'nolilik va ko'chma ma'nolar atrofida guruhlandi. Masalan, A. Qodiriy nasrida tana a'zolari leksemalari (ko'ngil, ko'z) ishtirok etgan iboralarning qo'llanilishi, so'zlarni to'g'ri ma'noda belgilashga moyil annotatorlar bilan ularni an'anaviy metaforik ma'noda tushunadigan annotatorlar o'rtasida tizimli tafovutlarni keltirib chiqardi. Generativ leksika metodologiyasiga [13;47-b] murojaat qilish ma'no ko'chishlarini alohida lug'aviy birlik sifatida emas, balki kognitiv struktura tomonidan ruxsat etilgan model sifatida kodlash imkonini berdi va bu qiyinchilikni butunlay yo'qotmasa-da, sezilarli darajada yumshatdi.

Pragmatik sath dastlabki ko'rsatmalarining cheklangan jihatlarini ochib berdi. Cho'lponning hikoya qilish uslubida tez-tez uchraydigan nutq ko'chishi nutqiy akt kuchi muallifga, personajga yoki har ikkisiga tegishliligi borasida annotatorlar o'rtasida jiddiy kelishmovchiliklar tug'dirdi. Ishonchlilik koeffitsiyentining kiritilishi muammoni to'liq bartaraf etmagan bo'lsa-da, bu nomuvofiqliklarni interpretiv shaffoflik mezoniga muvofiq, ochiq va tahlil qilinadigan holatga keltirdi.

Stilistik sathda trop va figuralarning tizimli aniqlanishi korpus stilistikasi adabiyotlarida qayd etilgan modellarni takrorladi [8;4-b]. Annotatorlar an'anaviy figuralarni osonlikcha bir xil aniqlashdi [9;40-b], biroq semantik nomutanosiblik va sintaktik prayming kabi nozikroq tuzilmalarda fikrlar ayri bo'ldi. Metafora va metonimiya o'rtasidagi chegara holatlari [11;43-b], metaforani aniqlash metodologiyasida [5;17-b] ko'rsatilganidek, aniq yechimga eng qiyin bo'ysunadigan qismlar bo'lib chiqdi.

Diskursiv sathdagi kelishmovchiliklar, asosan, ritorik birliklarni segmentlarga ajratish va ular o'rtasidagi aloqalarni aniqlash bilan bog'liq bo'ldi; bu holat Ritorik struktura nazariyasi annotatsiyasining umumiy tajribasiga to'liq mos keladi [14;248-b]. Nihoyat, konseptual-kognitiv sath eng past kelishuv darajasini ko'rsatdi va undagi tafovutlar konseptual metaforalarni nomlash hamda mental fazolar konfiguratsiyasini chegaralashda [10;105-b] ko'proq namoyon bo'ldi. Shunday bo'lsa-da, ishonchlilik qiymatlaridan foydalanish talqin qilishdagi noaniqlik o'chog'ini [12;89-b] yaqqol ko'rsatib berdi va uni tadqiqot uchun to'siqdan tahliliy resursga aylantirdi. Pilot ma'lumotlar asosida belgilangan mezonlar uchun uchta muhim xulosa shakllantirildi. Birinchidan, eksplitsitlik tamoyili mutlaqo zarurligini isbotladi: o'quv bosqichidan so'ng eng yuqori o'sish ko'rsatkichini qayd etgan sathlar yo'riqnomalarida ko'plab "inkor misollar" keltirilgan sathlar bo'ldi; bu esa ma'lum bir toifaga nima kirmasligini aniq ko'rsatish, unga nimalar kirishini tavsiflash kabi muhim ahamiyatga ega ekanligini anglatadi [18;22-b]. Ikkinchidan, sathlararo

moslashuvchanlik tamoyilini kengaytirish talab etiladi: ikki tomonlama bog‘liqlik holatlari model faqat sathlararo havolalarga yo‘l qo‘yib qolmay, balki aniq bivalent hodisalar uchun avtomatik ravishda ko‘p sathli qo‘shma annotatsiyani qo‘llab-quvvatlashi kerakligini ko‘rsatdi. Uchinchi, ishonchlilik qiymati va sharhlar maydoni orqali amalga oshirilgan interpretiv shaffoflik tamoyili past kelishuvli annotatsiyalarning ilmiy qiymatini sezilarli darajada oshirdi; model bunday holatlarni chetga surib tashlash o‘rniga, kompyuter ishonchliliqi tadqiqotlaridagi eng yaxshi amaliyotlarga [4;286-b] muvofiq, ularni talqin qilish qiyin bo‘lgan nuqtalar haqidagi dalillarga aylantirdi.

**Xulosa.** Mavjud empirik jarayonlarda qayd etilgan kelishuv darajalari bilan solishtirilganda, ushbu tadqiqot natijalari metodologik jihatdan ishonchli mavqega ega. Faqat bitta interpretiv qatlamni juda aniq qoidalar asosida tekshiradigan MIPVU adabiyotlarida  $\alpha$  qiymatlari turli tillar va janrlar kesimida odatda 0.55–0.70 atrofida bo‘ladi [6;99-b]. Nutq, yozuv va fikr taqdimotining korpus-stilistik tadqiqotlari ham toifalar bo‘yicha shunga o‘xshash o‘zgaruvchanlikni ko‘rsatadi; bunda eng aniq belgilangan toifalar 0.80 dan yuqori, eng noaniqlari esa 0.55 ga yaqin qiymatlarda jamlanadi [8; 95-b]. Bu yerda sinalgan yetti sathli sxema ushbu tayanch ko‘rsatkichlar bilan to‘liq mos keladigan kelishuv darajalarini qayd etdi, shu bilan birga empirik miqyosni badiiy tilshunoslikda qo‘llaniladigan barcha sathlar kesimiga hamda xalqaro annotatsiya tadqiqotlarida kam o‘rganilgan turkiy tillar an‘anasiga kengaytirdi [17;37-b].

XX asr o‘zbek nasri va nazmiga asoslangan ko‘p sathli annotatsiya tizimining pilot tadqiqoti, ushbu metodologiya eksplikatsiya, qatlamlar mustaqilligi va miqyoslash mezonlariga muvofiq amaliyotda samarali ekanligini tasdiqladi. Yuqori, o‘rta va quyi sathlar bo‘yicha annotatorlararo kelishuv tahlili modelning lingvistik tavsif va badiiy talqin kesishmasida qo‘llanilishi mumkinligini ko‘rsatadi. Ushbu tizimni kelgusida korpus hajmini kengaytirish va yarim avtomatik qayta ishlash orqali takomillashtirish rejalashtirilgan.

#### Foydalanilgan adabiyotlar ro‘yxati:

1. Artstein, R., & Poesio, M. (2008). Inter-Coder Agreement for Computational Linguistics. *Computational Linguistics*, 34(4), 555–596.
2. McEnery, T., & Hardie, A. (2012). *Corpus Linguistics: Method, Theory and Practice*. Cambridge: Cambridge University Press. — 294 p.
3. Ide, N., & Pustejovsky, J. (Eds.). (2017). *Handbook of Linguistic Annotation*. Dordrecht: Springer. — 1295 p.
4. Krippendorff, K. (2018). *Content Analysis: An Introduction to Its Methodology* (4th ed.). Thousand Oaks: SAGE. — 472 p.
5. Pragglejaz Group. (2007). MIP: A Method for Identifying Metaphorically Used Words in Discourse. *Metaphor and Symbol*, 22(1), 1–39.

6. Steen, G. J., Dorst, A. G., Herrmann, J. B., Kaal, A. A., Krennmayr, T., & Pasma, T. (2010). *A Method for Linguistic Metaphor Identification: From MIP to MIPVU*. Amsterdam: John Benjamins. — 238 p.
7. Halliday, M. A. K., & Matthiessen, C. M. I. M. (2014). *Halliday's Introduction to Functional Grammar* (4th ed.). London: Routledge. — 808 p.
8. Semino, E., & Short, M. (2004). *Corpus Stylistics: Speech, Writing and Thought Presentation in a Corpus of English Writing*. London: Routledge. — 248 p.
9. Mahlberg, M. (2013). *Corpus Stylistics and Dickens's Fiction*. London: Routledge. — 232 p.
10. Stockwell, P. (2020). *Cognitive Poetics: An Introduction* (2nd ed.). London: Routledge. — 222 p.
11. Lakoff, G., & Johnson, M. (2003). *Metaphors We Live By* (with new afterword). Chicago: University of Chicago Press. — 276 p.
12. Fauconnier, G., & Turner, M. (2002). *The Way We Think: Conceptual Blending and the Mind's Hidden Complexities*. New York: Basic Books. — 464 p.
13. Pustejovsky, J. (1995). *The Generative Lexicon*. Cambridge, MA: MIT Press. — 312 p.
14. Mann, W. C., & Thompson, S. A. (1988). Rhetorical Structure Theory: Toward a Functional Theory of Text Organization. *Text*, 8(3), 243–281.
15. Burnard, L. (2014). *What is the Text Encoding Initiative? How to Add Intelligent Markup to Digital Resources*. Marseille: OpenEdition Press. — 142 p.
16. Hardie, A. (2014). Modest XML for Corpora: Not a Standard, but a Suggestion. *ICAME Journal*, 38, 73–103.
17. Hoover, D. L., Culpeper, J., & O'Halloran, K. (2014). *Digital Literary Studies: Corpus Approaches to Poetry, Prose, and Drama*. London: Routledge. — 272 p.
18. Leech, G. (2005). Adding Linguistic Annotation. In M. Wynne (Ed.), *Developing Linguistic Corpora: A Guide to Good Practice* (pp. 17–29). Oxford: Oxbow Books.