**Про тегсет LT**

Ось наші спільні побажання щодо тегсету:

1. Варто уніфікувати назви тегів. Наразі суміш англ. і укр. Найбільше муляє, коли обидві мови змішано: наприклад, замість insert\_sl можна insert, замість prysudk\_sl – predic.

Деякі перевів в англійську, все ж таки англ. назви виглядають читніше, щоправда з тегами відмінків тощо ще вагаюся

2. impers - далеко не до всіх дієслів подано безособові форми, а деякі подані неправильно. Як генерується impers? Приклади помилок:

зламано - adv

виписано виписано adv

відписано відписано adv

вписано вписано adv:v-u

уписано уписано adv:v-u

надписано надписано adv

перезаписано перезаписано adv

переписано переписано adv

пописано пописано adv

приписано приписано adv

розписано розписано adv

списано списано adv

Ці поправив, ці форми генеруються зі словника spell-uk з дієслів прапорцями W та 0, проблема, що багато таких форм дуже схожі на прислівники, які генеруються прапорцем W з прикметників. Тож тре пересовувати W з прикметників до дієслів. Підозрюю там ще є.

3. Варто додати теги unknown (нерозпізнане), abbr (скорочення), number (на арабські й римські числові позначення).

Додав unknown (хоч зараз він вживається всюди де є якась невизначеність), abbr, щодо number не зрозумів, в коді Java всі арабські числа автоматично отримують тег числівника (numr).

4. Для іменників так званого спільного роду варто вказати окремо всі можливі омонімічні реалізації. Наприклад, ледащо може бути всіх трьох родів (3 омоніми). Так уже зроблено з прізвищами на -ко (Авраменко) і -чук (Ковальчук). Завваж, що Іменники спільного (жіночого і середнього) роду мають різні парадигми відмінювання: жін.р, род.в. - бабищі, сер.р., род.в. - бабища. Більше про іменники спільного роду: uk.wikipedia.org/wiki/**Спільний**\_**рід**

Це велика справа, яку боюсь я сам не потягну, прізвища добряче впливали на правила, тож я пару випадків підігнав, але на більше мене не вистачило. Тут був би вдячний за допомогу.

5. Треба прописати рід в абревіатур типу НАТО - наразі: НАТО НАТО noun:nv

Це у файлі base-abbr.lst тре додавати. Я не ризикну швидко по всіх скороченнях проставити правильний рід. Не відмовлюся від допомоги.

6. Додати до списку тегів ist, bad, v-u, imperf, perf, tran і nontran – ці теги використано в списку словоформ, але їх бракує в тегсеті. Nontran ліпше назвати intran (intransitive).

Додав опис, поміняв на intran

7. Інколи tran застосовано неправильно, наприклад, до зворотних дієслів, які **всі** є насправді неперехідними: абеткуватися абеткуватися verb:inf:rev:imperf:tran

виправив

8. Дивно виглядає застосування взаємовиключних тегів: абоновано абонувати impers:imperf:perf:tran. Має бути дві омонімічні форми, внесені окремо

Здається, виправив

9. Дивний повтор perf: зробивши зробити dieprysl:perf:perf:nontran:tran

виправив

10. Бракує тега super в словах на щонай-: щонайактивніший щонайактивніший adj:m:v\_naz

зроблено

11. Треба додати список найчастотніших прикметників у короткій формі (вони не відмінюються) - ясен, дрібен, зелен, повен, славен, красен, винен, потрібен, певен, годен, ладен, рад тощо.

Якщо ти зможеш надати список, я додам (там тре буде подумати як його найкраще прикрутити)

12. Для всіх прикметників передбачити генерування нестягненої форми в жін. і сер. роді: *гарная, гарную; гарнеє; гарнії*. Це переважно фольклорний і поетичних вжиток або стилізація. Нестягнені форми трапляються й у дієприкметників (*розмальованії*), але рідко й, до того ж, можуть мати форму русизмів (*писаную*).

Залишу собі в TODO

13. Лемою дієприслівника має бути дієприслівник (так само як з дієприкметником).

Це тре обговорити, по-перше словники (ВТССУМ) не дають дієприслівники, також уліф завжди їх дає всередині статті про дієслово; для мене це проблематично, бо коли я пишу правило на «зняти/знімати квартиру», одразу ловиться «знявши/знімаючи крартиру», якщо я розіб’ю, кожного разу мені тре писати «зняти/знявши/знімаючи квартиру» тощо

14. Помилка imperf:imperf:perf

нарізаючи нарізати dieprysl:imperf:imperf:perf:nontran:tran

гойдаючись гойдати dieprysl:rev:imperf:imperf:nontran:tran

абортуючись абортуватися dieprysl:rev:imperf:imperf:perf:nontran:tran

виправив

15. Додати форми кличного відмінка для всіх іменників, а не тільки (з неістот):

конюшино конюшина noun:f:v\_kly

досадо досада noun:f:v\_kly

це можна (тре додати прапорець кличного до іменників в spell-uk/src/Dictionary/\*.lst, але боюсь що багато форм є дуже рідковживаними, тож вони будуть заважати перевірці орфографії, хіба що додати тег :rare і ігнорувати в спелчекері

16. Додати категорію особи до займенників.

приклади?

17. Розвести по окремих рядках (омонімах) perf i imperf у двовидових дієсловах:

адресував адресувати verb:past:m:imperf:perf:tran

заадресував заадресувати verb:past:m:perf:tran

переадресував переадресувати verb:past:m:perf:tran

Можеш докладніше, навіщо це?

18. Послідовно помаркувати як adj всі займенники, які відмінюються, як adj (бо наразі мішанина). Наприклад, такі:

сякий-такий сякий-такий pron:m:v\_naz

цей цей pron:m:v\_naz

той той pron:m:v\_naz

сякий-такий сякий-такий pron:m:v\_naz

вся весь pron:f:v\_naz

Ось тут не знаю, в мене так було, але я почав примусово ставити їм теги займенників (і вже в правилах це використовую); яка причина мати тег прикметника на займенникові?

19. Cполучники поділити (додати новий тег) на сурядні і підрядні

Якщо можеш надіслати перелік, додам

20. Бракує форм *чиїхось, чиїмось* та, можливо, інших зі вставним о.

В Уліфі їх нема, але є у Голоскевича, якщо дасиш форми, я закину в словник

21. Бракує прислівників найвищого ступеня на най-, щонай- та якнай-: *щонайдовше* і *якнайдовше* потрактовано лише як прикметники. Немає також прислівників вищого ступеня типу *довше, сильніше*

TODO

22. Бракує складених прийменників типу *з-попід, з-позад*, а також *вві, уві, понадо.*

*Перші два і понадо додав, вві і уві вже були*

23. помилка: пополудні пополудні dav - має бути adv

Виправив

24. Дієприкметники було б добре промаркувати окремим тегом. Це важливо для подальшого аналізу

Було б добре мати список, бо вручну їх вишукувати буде довго

25. Хотілося б зрозуміти, як саме із основ у словнику ти отримуєш словоформи. Ми розуміємо, що є коди, але як за ними додаються квазізакінчення? Коротше, допоможи розшифрувати позначення в myspell-uk (щоб порівняти з анотацією в граматичному словнику).

Інформація у файлах spell-uk/src/Affix, плюс багато логіки в spell-uk/bin/tag/mk\_pos\_dict.py

26. Не зуміли побачити в роботі програмку з додавання нових слів до словника. Ruby встановити, але забракло якогось перехідного елемента між Ruby і Qt - сайт, що його надає, не працює. Як би то побачити цю програмку в дії?

Під лінуксом тре ще встановити пакунок smoke-qt, але також тре мати встановлений aspell і словник aspell-uk (бажано останньої версії). Щоправда я зараз намагаюся позбутися залежності від aspell

27. Зі свого боку, хочемо запитати тебе, як ми можемо допомогти?

Це тре обговорити, в цілому: поповнення/вдосконалення словника, правила розомонімізації теж не завадило б додати, але як це робити і хто буде допомагати тре узгодити.