

# Sense Ranking in Dual-Language Online Dictionaries

**Mark Kit**, Ph.D., Language Interface (USA),  
**Elena Berg**, Ph.D., Ural State Law University  
(Russian Federation)



# Introduction

- Dual-language online dictionaries facilitate high performance in translation and can be easily updated
- Developers of these dictionaries face many challenges
- Among other aspects, the order of presentation of retrieved translations is of great importance because it affects the speed of finding the required translation
- Merging different source dictionaries in a single dataset may result in translations placed in non-systematic order

# Finding required translation in the dictionary output is often hard and time-consuming

lead [li:d] *сущ.* | Вебстер | фразы

Поиск Eng

*общ.* грузило; отвес; руководство; инициатива; пример; директива; вводная часть; первое место; поводок; привязь; разводье (во льдах); электропроводка; трубопровод; канал; глава; грифель; пломба; ведущее место в состязании; первый ход (в игре); краткое введение к газетной статье; золотиносный песок; проводка; шпоны; провод; подводный провод; свинец (металл, [l e d]); должик (*Chita*); передовица (в газете); дорожка; указания; графит (в карандаше); наводка (термин в продажах *HoldMyDrink*); зацепка (например, при проведении следствия *scherfas*); водило (*Gruzovik*); приводить (impf of привести *Gruzovik*); деловой контакт (*Дмитрий\_P*)

*общ.* электропровод

*Ebay.* пробник тестера

*SAP.tech.* руководитель

*авиа.* питающий провод

*авиа.мед.* ведущий самолёт; лётчик ведущего самолёта; отведение (электрофизиологическое)

*авт.* подъём резьбы; подъём; вывод (батареи)

*автом.* заборная часть (метчика); угол наклона; центрирующий пояс; заходная часть (метчика); ход (винта, резьбы)

*архит.* маяк (устанавливаемый при ведении каменной или кирпичной кладки); свинцовые листы для покрытия крыши; электрический провод

*бизн.* первенство; указание; менеджер (*Oleg Sollogub*); «лид» <не рек.> (*MichaelBurov*); вести дела (*Johnny Bravo*)

*бур.* подводный кабель; графит (led); уплотнительная смазка для трубных резьб (led)

*бухг.* ключ (к решению); управление

*воен.* упреждение (по движущейся цели); приведение огня; ведущая позиция (*Kisелев*); величина выноса точки прицеливания (*Kisелев*); вынос точки прицеливания (*Kisелев*); голова колонны (*Kisелев*); дистанция (между нами колонны *Kisелев*); промежуток времени (*Kisелев*); упреждение (при ведении огня по

# Example of translation of the Russian word *вести*

вести  Поиск Eng

**вести** *сущ. фразы*

*общ.* lead down; go (о дороге, пути); sail; tend (к чему-л.); keep a diary (дневник, счета, бухгалтерию); keep up a running commentary (радиопередачу); carry on (дело); pace; drive through; anchor (передачу *joyand*); take up the running (в гонке); go with (к чему-л. в речи (Where do I go with this?) *bellbird*); take; rumors (pl *Gruzovik*); rumours (pl *Gruzovik*); pass over/across (impf of провести *Gruzovik*); run over (impf of провести *Gruzovik*); lay (impf of провести *Gruzovik*); install (impf of провести *Gruzovik*); build (impf of провести *Gruzovik*); moderate (заседание, собрание, as in: Following this announcement, the Co-Chair invited Eva Muller, Director, FAO, to moderate the panel discussions on new developments in forest financing. *Игорь Муз*); host (передачу; заседание *Игорь Муз*)

*авт.* slip (о сцеплении *sega\_tarasov*)

*воен., тех.* steer (например, мотопонтон)

*каб.* conduct (переговоры); run (бизнес)

*Макаров* carry over; drive at (к чему-л.); drive up; take (занятия и т. п. *В.И. Макаров*)

*матем.* ran

*мор.* navigate (корабль); steer (судно)

*образн.* keep on track ("Political forces create an environment where it's difficult to keep technical discussions on track". "Политические силы порождают среду, в которой невозможно продолжать обмен техническими идеями".| "Fig. to make sure that someone or some process continues to progress properly". *inyazserg* (<http://idioms.thefreedictionary.com/keep+on+track>.)

*разг.* keep on at a person (записи, счета); tip off (You have a good memory but it's your emotions that are tipping you off, not your head. *VLZ\_58*)

*спорт.* be up (в счёте: we are up 2-0 *Юрий Гомон*)

**ВЕСТЬ** *сущ. в начало*

*общ.* news; message; gossip; rumor; word

*матем.* piece of news

**Вести** *сущ. в начало*

*общ.* Vestie (мужское имя)

*бизн.* Chronicle ("San Francisco Chronicle" = "Вести Сан-Франциско" *vatnik*)

**"Весть"** *сущ. в начало*

*рел.* Nahaa (78-я сура Священного Корана); Tidings (78-я сура Священного Корана)

**вести** *сущ. в начало*

*общ.* к чему-л. build to (напр., о сюжете *plushkina*)

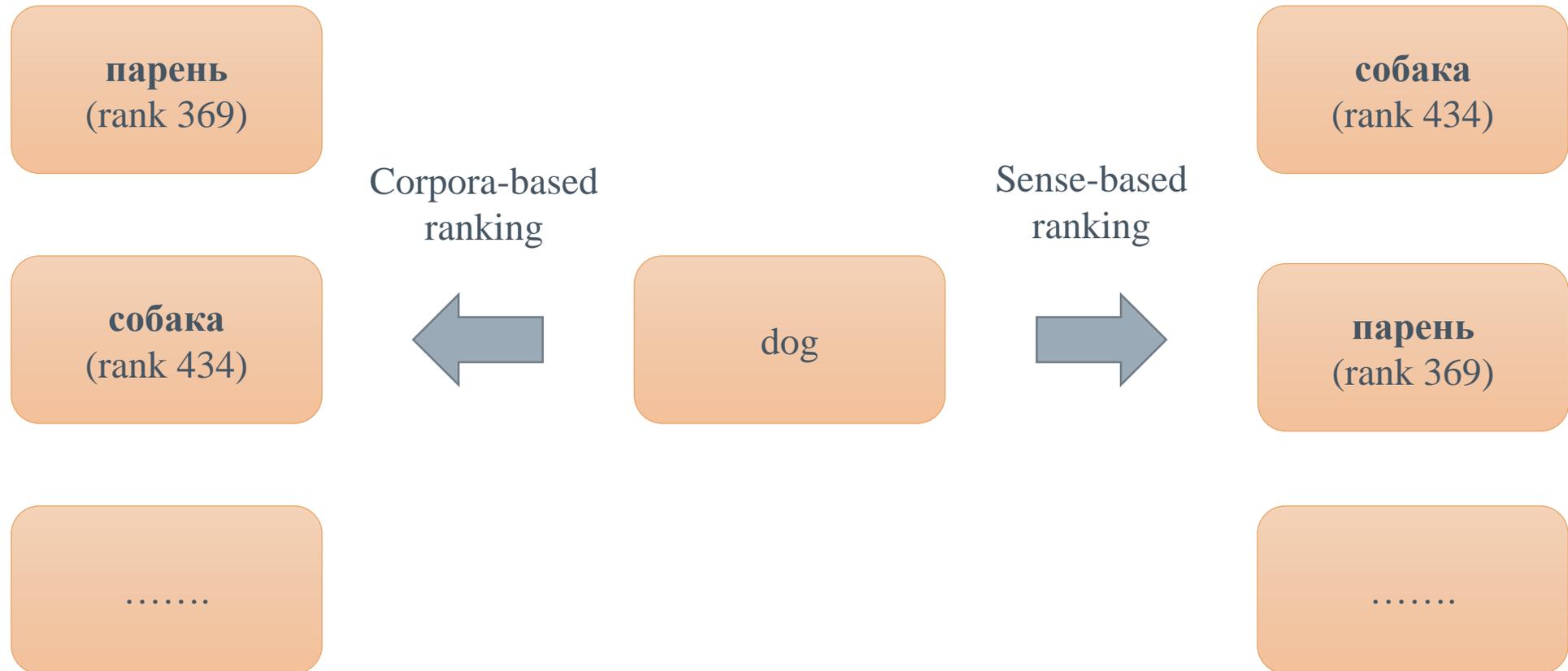
**вести хозяйственную или иную ) деятельность на грани нарушения закона** *сущ. в начало*

*общ.* operate on the fringes of the law (*Игорь Муз*)

**вести** *сущ. в начало*

*общ.* образ жизни follow (The 75-year-old rocker - who has been a longtime animal activist - has followed a vegetarian lifestyle for years - 75-летний рокер - который является давним защитником животных - ведёт вегетарианский образ жизни на протяжении многих лет (о Поле МакКартни) *Lily Snape*)

# Sense-based vs. corpora-based ranking



# Ranking with the LexSite Platform

## lexsite-dictionary.com

The screenshot displays the LexSite 5.0 interface. At the top, there is a navigation bar with the LexSite logo (LS) and the text 'LexSite 5.0'. Below this, a search bar contains the word 'load' and a search icon. To the right of the search bar are buttons for 'Browse', 'New Node', 'Aliases', and 'Rank List'. Below the search bar, there are two buttons: 'Noun' (selected) and 'Verb'. The main content area shows the search results for 'load' as a noun. The results are listed in a table with a scroll bar on the right. Each row contains a small brown square icon, the Russian translation, and a blue circular arrow icon pointing up.

load <i>noun</i>
■ нагрузка
■ загрузка
■ бремя
■ ноша
■ заряд
■ партия груза на вагон
■ артиллерийский выстрел
■ избыток
■ обилие
■ «товар», запас нелегально приобретенных наркотиков
■ патрон
■ снижение способности к выживанию средней особи в популяции из-за повреждений генома
■ шихта
■ колоша

# Results of Manual Ranking

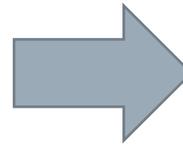
LexSite 5.0

Browse New Node Phrases A license

Noun Verb

license *noun*

официальное разрешение	✎	✕
патент на продажу спиртных напитков	✎	✕
отклонение от правил, норм <i>в искусстве, литературе</i>	✎	✕
злоупотребление свободой	✎	✕
удостоверение <i>о наличии определенных прав</i>	✎	✕
вольность <i>(излишняя)</i>	✎	✕
распушенность	✎	✕
права <i>(на управление автомобилем)</i>	✎	✕
патент	✎	✕
разрешение	✎	✕
диплом	✎	✕
лицензия	✎	✕



LexSite 5.0

Browse New Node Phrases A license

Noun Verb

license *noun*

лицензия	✎	✕
права <i>(на управление автомобилем)</i>	✎	✕
официальное разрешение	✎	✕
диплом	✎	✕
патент	✎	✕
разрешение	✎	✕
патент на продажу спиртных напитков	✎	✕
отклонение от правил, норм <i>в искусстве, литературе</i>	✎	✕
злоупотребление свободой	✎	✕
удостоверение <i>о наличии определенных прав</i>	✎	✕
вольность <i>(излишняя)</i>	✎	✕
распушенность	✎	✕

# What will it take to rank the senses?

- The scope of ranking efforts can be very big or even prohibitive
- These efforts should be evaluated for feasibility and planning purposes
- The initial ranking shall be based on sense frequencies
- To determine the scope of efforts and feasibility the analysis below was undertaken as shown below

# Lexical Dataset Subjected to the Analysis

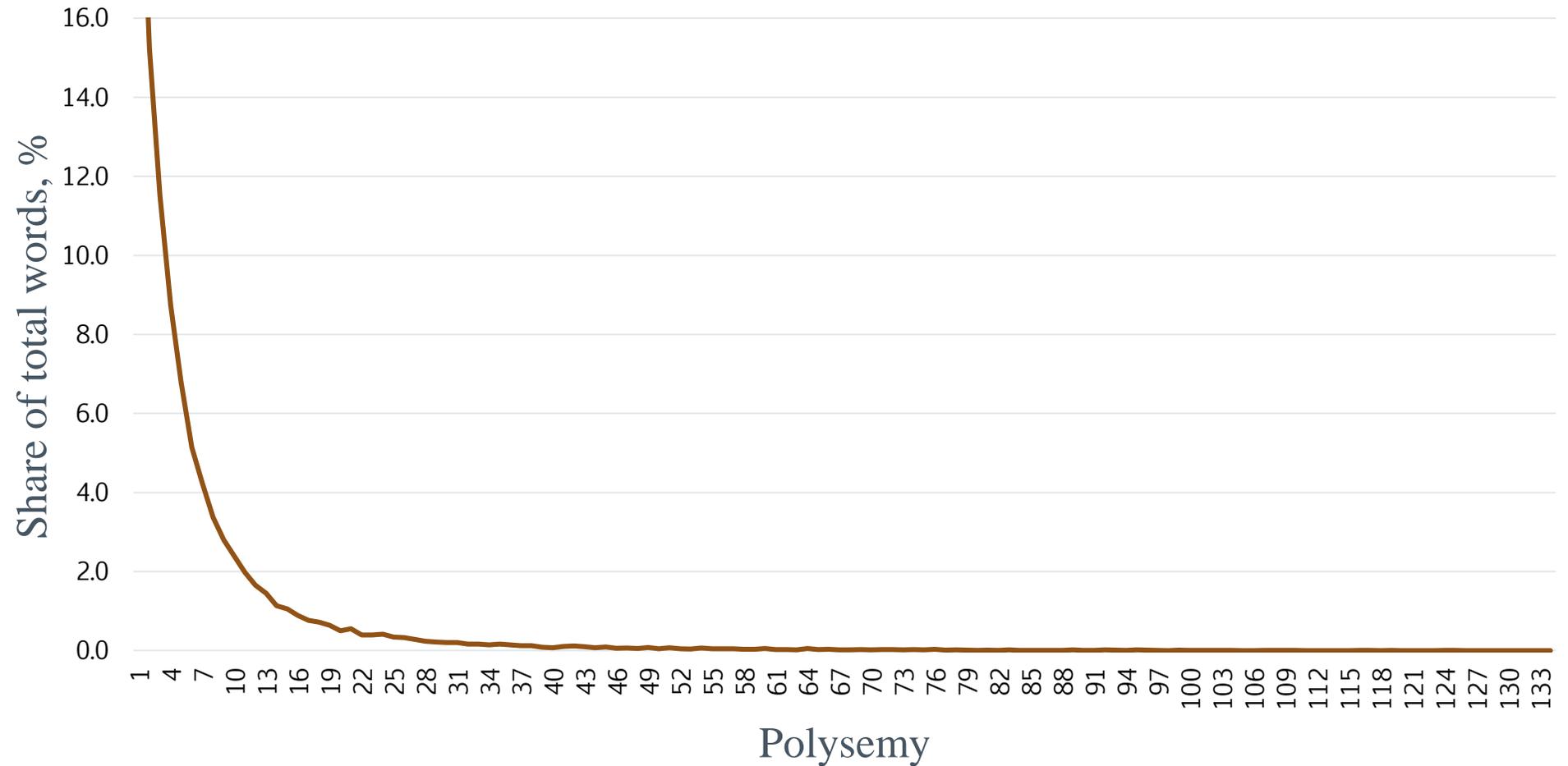
	English	Russian
Nouns	19,992	12,063
Verbs	4,958	5,700
Adjectives	8,919	5,004
Adverbs	1,275	1,124
Total	35,145	23,892

# Ranking Strategy Assumed

- Ranking is not required where a POS group includes a single translation of a source word
- Where a POS group includes a small number of translations (1-5) ranking is unimportant since it is easy to find the needed translation among the offered ones regardless of their order
- Words with greatest corpora-based frequencies should be ranked first since they are in the greatest demand

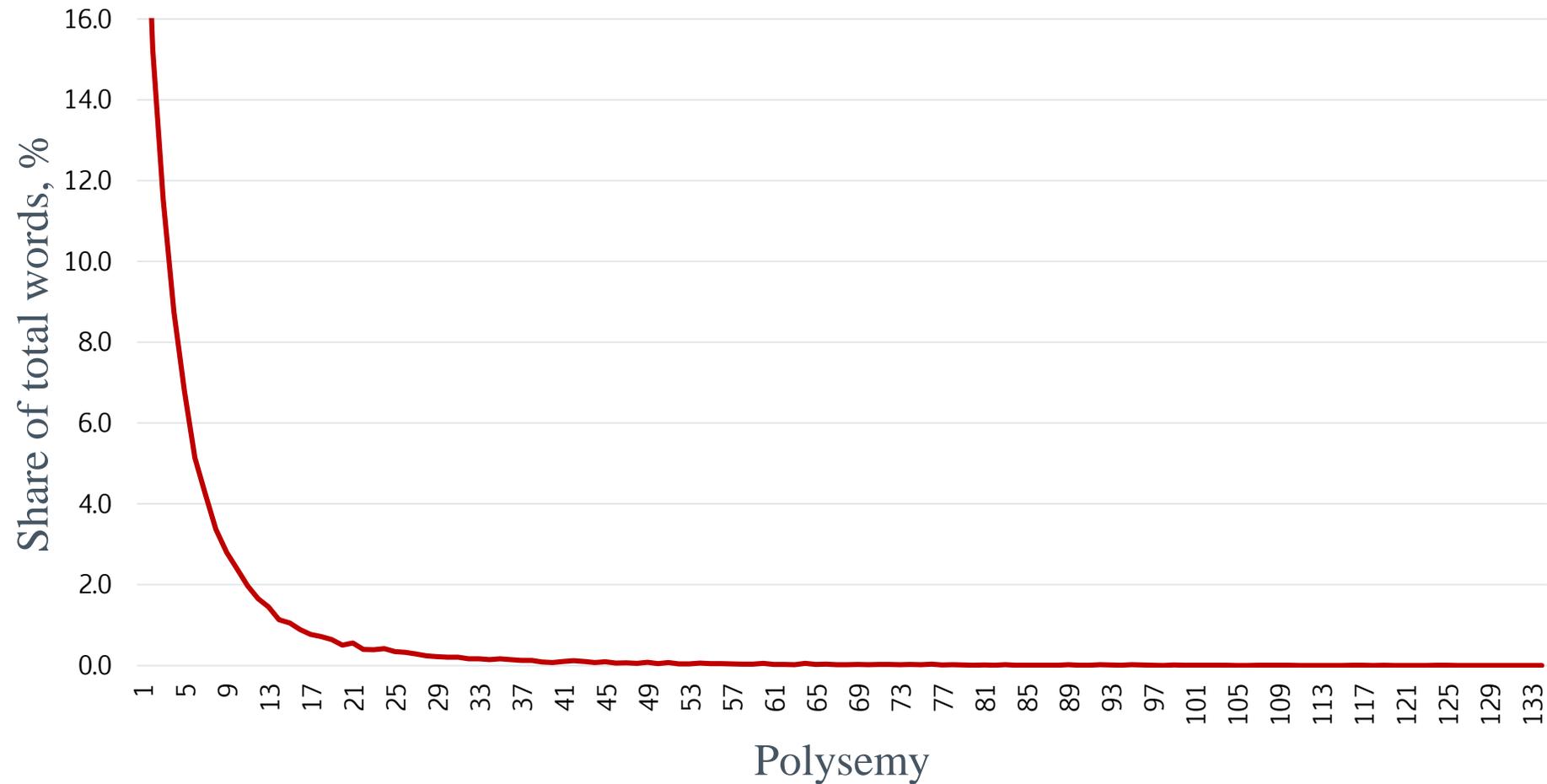
# The Fewer Meanings the More words

Distribution of share of words vs. word polysemy, English language



# Similar distribution in the Russian Language

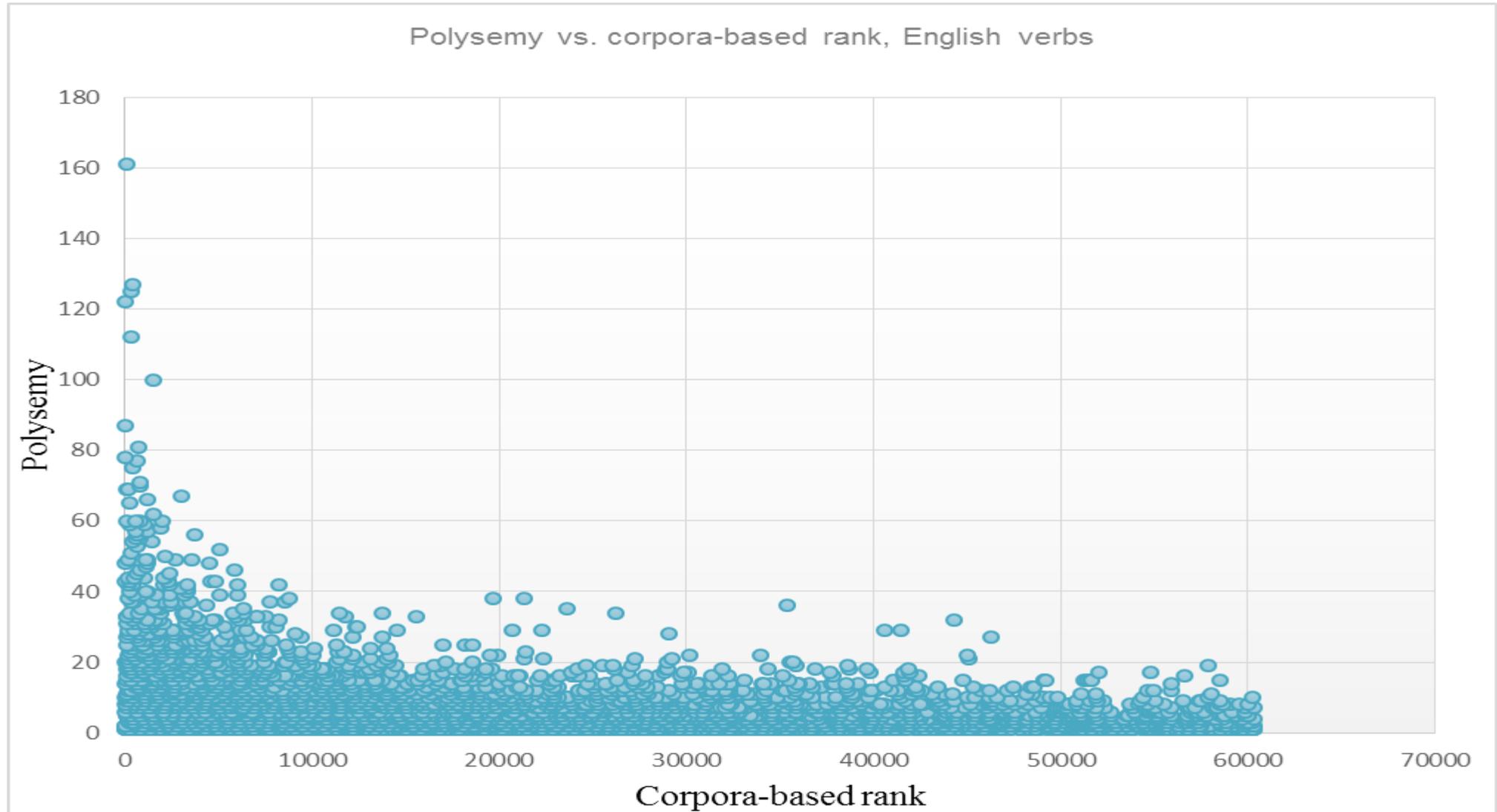
Distribution of share of words vs. word polysemy, English language



# The number of High Polysemy Words is Small

- English verbs with polysemy 30 and above: 3.2%
- English nouns with polysemy 30 and above: 3.4%
- Ranking low-polysemy words goes fast
- This implies that manual ranking is feasible even with large datasets

# Dependency between the corpora-based rank and polysemy



# Conclusion

- This study can form a basis for work planning when making dual-language dictionaries earning assignments
- The results of this study will help develop sense-ranking assignments for students studying foreign languages so that they would improve their lexical proficiency through ranking words on their own.
- The language teachers, on the other hand, can use this data when making learning dictionaries focused on specific pedagogical needs

Thank you!

