Annex to the paper *A new corpus annotation framework for Latin diachronic lexical semantics (REFERENCE)*

Helena Bermúdez Sabel, Francesca Dell'Oro and Paola Marongiu

University of Neuchâtel

This annex[1] contains the analysis of 18 words in relation to the vagueness study (section 7 of the paper). It is intended to be a working document on which we based the analysis published in the paper. Moreover, while in the paper we had to select the most interesting words, in this annex the interested reader will find more information on the vagueness results and the preliminary analysis of each word.

For each lemma we specify if it is a target or a control word (and its counterpart), the meanings that were given to the annotators, the derivation pattern, and the vagueness analysis.

The derivation pattern is defined in terms of sources and bases. The identification of bases and sources was made on the assumption that the order of the meanings is for the most part chronological, as dictionaries usually present older meanings before more recent ones. We are aware that this can be a limit of our analysis. With the term *source* we indicate a meaning that is at the origin of another one. By the term *base* we specifically refer to a meaning that is (potentially) the source of other meanings, but that was not derived from another one. A base meaning can be the source of other meanings or not be productive as a source at all. At the same time, a derived meaning (therefore not a base) can be the source of other meanings. We will illustrate how we analysed meanings in terms of bases and sources with the example of *regnum*.

Regnum has four meanings:

1 kingship, royalty

2 dominion, sovereignty, rule, authority

3 a kingdom

4 a territory, estate, possession

The first two meanings are abstract: while the first one is very specific referring to the status of a king or similar (base); the second meaning is less specific, potentially referring to the status of anyone holding authority over people. The third meaning is more concrete and depends – as meaning 2 – on the base. The fourth meaning is concrete and the concept of authority is absent. *Regnum* has therefore – according to our analysis – one base (meaning 1) and two sources (meaning 1 is the source of meanings 2 and 3, and meaning 3 is the source of meaning 4).

The analysed words are presented in descending order in relation to the vagueness score. The words are classified in three categories: high vagueness, medium vagueness and low vagueness.

---

[1] The attribution of the analyses is as follows: Helena Bermúdez Sabel (HBS) wrote *adsumo*, *beatus*, *consul*, *coohors*, *jus*, *simplex*; Francesca Dell'Oro (FDO) wrote *credo*, *fidelis*, *necessarius*, *regnum*, *sacramentum*, *templum*; Paola Marongiu (PM) wrote *ancilla*, *consilium*, *dubius*, *dux*, *oportet*, *scriptura*.

High vagueness score

Lemma with its target or control word	OPORTET (target word; control word: none)
Number of meanings and meanings	1. It is necessary 2. It is proper / it is becoming
Source pattern	2 = base 2 --> 1 Pattern: 1 base, 1 source
Commentary	The lemma *oportet* shows the highest vagueness score among all the lemmas annotated. In 55 out of 60 contexts the annotation is "4, 3": if "it is necessary" has a "4", "it is proper/it is becoming" has a "3" and vice versa. This is the reason why the vagueness score is high. This type of annotation shows that the annotator perceived the two meanings as being closely related. The possible explanation for this is that in most cases the two meanings are tied by a relation of implication, that could function in both ways: the "properness" is almost always implied in the notion of "necessity", and the notion of "properness" also seems to imply that of "necessity". There are some exceptional cases: in one of the contexts, *oportet* is annotated with a "4" for "it is necessary", and a "1" for "it is proper / it is becoming". We noticed that deviations from the most frequent pattern of annotation ("4, 3") are concerned with the type of modality expressed by the lemma in the relevant contexts. Specifically, in all cases of deviation from the established pattern, the annotator selects "it is necessary" as the only pertinent meaning, and "it is proper/it is becoming" is annotated as absolutely not related.

Lemma with its target or control word	DUX (target word; control word: HONOR)
Number of meanings and meanings	1. Leader 2. Commander, general-in-chief 3. Duke
Source pattern	1 = base 1 --> 2 2 --> 3

	Pattern: 1 base, 3 sources
Commentary	The vagueness score for this word is very high, probably because the meanings 1 and 2 are closely related from a semantic point of view: there is a strong relation of implication between them, as "commander, general-in-chief" is a hyponym to "leader". *Dux* is a target word. However, the vagueness score does not increase significantly for the texts from CE. This is probably due to the fact that the new meaning "duke" is almost always annotated with "1". There is only a passage from a CE text, where it is annotated with a "3". Meaning 1 and 2 are the ones showing the highest level of implication based on the annotation. On a closer analysis of the annotations, we noticed that whenever the meaning "commander, general-in-chief" receives a "4", the meaning "leader" receives a "3". For instance, in the passage from Hilarius Pictauiensis, *De martyrio Maccabaeorum, dux* refers to Moses. For this passage,"leader" is annotated with a "4" and "commander, general-in-chief" with a "3". The only exception to this is one annotation "4, 2" for a passage in Titus Livius' *Ab urbe condita,* where the context does not seem to play a difference with the other contexts annotated as "4, 3". The implication of "leader" in "commander, general-in-chief" is rarer, alternating "4, 3" with "4, 2" or "1". For what concerns the meaning "duke", it is always assigned the value "1", except for the XVI CE *Baldus* written by Teofilo Folengo*,* where the annotation is "leader" "4", "commander, general-in-chief" "3" and "duke" "3". To be noted: Gaffiot (1934/2016) does not give "duke" as an available meaning for *dux,* as it is not meant to go beyond Late Antiquity. Since the meaning "duke" for the word *dux* is attested later on, the source for this meaning was the DMLBS (Dictionary of Medieval Latin from British sources).[2]

Lemma with its target or control word	**REGNUM (control word; target word: *imperator*)**
Number of meanings and meanings	1. Kingship, royalty 2. Dominion, sovereignty, rule, authority 3. A kingdom 4. A territory, estate, possession
Source pattern	1 = base 1 --> 2 1 --> 3 3 --> 4

---

[2] Ashdowne, Richard K., David R. Howlett, & Ronald E. Latham (eds.). 2018. *Dictionary of Medieval Latin from British Sources.* Oxford: British Academy.

	Pattern: 1 base, 2 sources
Commentary	It appears as a highly vague term already in the BC period, though this could be related also to the annotation style. The vagueness score slightly diminishes in the CE period.

In a relevant number of cases (18) the annotator couldn't identify a precise meaning, so that for these 18 passages the value 4 is lacking. In a certain number of cases 2 meanings –

"kingship, royalty" (1) and "dominion, sovereignty, rule, authority" (2) once;

"kingship, royalty" (1) and "kingdom" (3) 11 times;

"kingdom" (3) and "territory, estate, possession" (4) twice;

– seem fitting (they are both marked 3, but note that there is also a case where both (1) and (3) are marked "4".

Rarely 3 meanings seem fitting: "Dominion, sovereignty, rule, authority" (2), "kingdom" (3) and "territory, estate, possession" (4) once; or "kingship, royalty" (1), dominion, sovereignty, rule, authority" (2) and "kingdom" (3) once. |

Lemma with its target or control word	**FIDELIS (target word; control word: *simplex*)**
Number of meanings and meanings	1. trusty, faithful, sincere, true,

2. to be depended upon, sure, safe, strong, firm,

3. Christian |
| **Source pattern** | 1 = base

1 --> 2

1 --> 3

Pattern: 1 base, 1 source |
| **Commentary** | The vagueness score slightly diminishes in the CE period. When meaning 3 is the fit choice, the context could imply the idea of being "trusty, faithful, sincere, true" (meaning 1, 7 times). When the fit choice is meaning 1, the context can imply meaning 3 (7 times meaning 3 is marked "3" and twice "2"). |

Lemma with its target or control word	IUS (control word; target word: CIVITAS)
Number of meanings and meanings	1. broth, soup, sauce 2. juice, mixture 3. right, justice, duty 4. a court of justice 5. justice, justness 6. legal right, power, authority, permission
Source pattern	1 = base 1 --> 2 3 = base 3 --> 4 3 --> 5 3 --> 6 Pattern = 2 bases, 2 sources
Commentary	This is not a polysemic word, but two homonymic ones. The homonym meaning "broth, soup" is related to Sanskrit *yū́ṣ* and maybe Greek *zúmē̄* (Proto-Indo-European *$i(e/o)uH$-s- 'broth, soup'); the other derives from the Proto-Indo-European root *h_2oi-u 'vital force' > *h_2ieu-os/es (De Vaan, 2008: 316). Therefore, there are two meanings that belong to the first homonym, and the second one presents four meanings. The distinction between the homonyms was clear, but the high vagueness score comes from the vagueness of the four meanings related to the semantic field of "justice". In fact, the word "justice" appears as a translation in two of the provided meanings. The contexts of annotation of this word go from Old Latin to Contemporary Latin. After classifying the annotation between BCE and CE, the level of vagueness of both periods is practically the same. As one of the control words, *jus* does not present a new meaning related to the Christian context, therefore it is not expected to present great differences in terms of vagueness between both eras.

Lemma with its target or control word	BEATUS (target word; control word: NECESSARIUS)

Number of meanings and meanings	1. Happy 2. Fortunate 3. Rewarded 4. Rich 5. Blessed (religion)
Source pattern	1 = base 1 --> 2 1 --> 3 2 --> 5 3 --> 4 Pattern = 1 base, 3 sources
Commentary	In Gaffiot (1934/2016), the first four meanings get simplified into three which can be paraphrased as follows: 'happy', 'rich (from a materialistic point of view)' and 'rich (in a metaphoric sense)'. Thus, the organization of the meanings in the lexicographical works illustrates in a way the lack of distinctness between the meanings "fortunate" and "rewarded" vis-à-vis "happy" and "rich". This is empirically confirmed by looking at the annotation of this word which shows that the higher level of vagueness is noted between "happy" and "fortunate". The vagueness score is much higher in the annotated classical passages than in the later ones. This is due to the distinctness of the new meaning: the annotation shows that the new meaning is clearly identified in the passages that contain it.

Medium vagueness score

Lemma with its target or control word	**SCRIPTURA (target word; control word: none)**
Number of meanings and meanings	1. Writing 2. Tax 3. Will

	4. Holy Scripture
Source pattern	1 = base 1 --> 2 1 --> 4 2 --> 3 Pattern: 1 base, 2 sources
Commentary	A possible factor for the vagueness score of this word might be the metonymic relation between the base "writing" and some of the meanings, that is "tax" and "will". *Scriptura* is also a target word, so we would expect a sharper distinction of meanings between BC and CE texts. The vagueness score is much higher in CE than for texts from BC. The reason for this might be that the meaning "Holy Scripture" always implies the meaning "writing". This is demonstrated by the fact that whenever the former gets a "4", the latter gets a "3". The annotator has also been careful with adjectives: a comment at the beginning of the sheet explains that if the word clearly means "Holy Scripture" but the nuance of holiness is given by an adjective (e.g. *sacrae Scripturae*), then "writing" gets a "4" and "Holy Scripture" gets a "3", because the word *scriptura* in itself carries the meaning "writing", and it is the presence of the adjective that adds the specification "Holy". The meanings "tax" and "will" are very rarely selected. There are a few cases where the annotator selected a "0" for "will", leaving a comment explaining that the context clearly indicated a legal document, but not necessarily a will. The same seems to be verified with patterns such as "4" to "writing" and "3" to "will" (all the others being "1"). In the – rare – cases where "tax" is assigned a "4", all the other candidates receive a "1". This does not impact the calculations as there are only two cases with this type of annotation pattern. But it is interesting to notice how our analysis of the derivation pattern does not completely overlap with the analysis of the annotator: the link between Holy Scriptures and the act (or the product) of writing is probably perceived as more straightforward than the one between the latter and the tax. The annotator also suggests adding another meaning, to distinguish the action "writing" from the product of writing.

Lemma with its target or control word	DUBIUS (control word; target word: none)
Number of meanings and meanings	1. Moving two ways, fluctuating 2. Uncertain, doubtful 3. Precarious, dangerous, critical, difficult
Source pattern	1 = base

	1 --> 2 2 --> 3 Pattern: 1 base, 2 sources
Commentary	*Dubius* is a control word, so we do not expect a substantial change in the vagueness score for texts from CE. However, vagueness increases from BC to CE texts. This could be due to the fact that the semantic relation between the first two meanings is very close. Looking at the *Thesaurus Linguae Latinae* (1900-) we can see that the second and the third meaning appear almost at the same time, in BC. Therefore, the rise of the vagueness score is not due to the emergence of another specific meaning in CE. On a closer analysis of the annotation, we noticed that when "moving two ways, fluctuating" or "uncertain, doubtful" receive a "4", "precarious, dangerous, critical, difficult" hardly gets a value different from 1 (7 annotations out of 60 deviate from this pattern, with annotations "2" or "3"); whenever "precarious, dangerous, critical, difficult" gets a "4" the other two meanings always receive a "3" or "2"; when meaning 1 receives a "4", meaning 2 often receives a "3" and vice versa. If we interpret this annotation from the point of view of semantic implications, we could hypothesize that: there is a reciprocal implication between the first two meanings "moving two ways, fluctuating" and "uncertain, doubtful"; the first two meanings are often implied in the third one "precarious, dangerous, critical, difficult", but they do not seem to imply – not necessarily at least – the third.

Lemma with its target or control word	**SACRAMENTUM (target word; control word: none)**
Number of meanings and meanings	1. a cause, a civil suit or process, 2. the military oath of allegiance, 3. an oath, a solemn obligation or engagement, 4. a secret, 5. a sacrament, 6. a mystery
Source pattern	1 = base, 2 = base 2 --> 3 3 --> 4 3 --> 5

	5 --> 6
	Pattern: 2 bases, 3 sources
Commentary	We individuated tentatively 2 bases and 3 sources. [however, the dictionaries Lewis & Short (1879) and Gaffiot (1934/2016) indicate that meanings 4, 5 and 6 are clearly associated]
	Meaning 3 seems a generalization of meaning 2, so that it is not surprising that, when meaning 2 is recognized, also meaning 3 receives a high score ("3").
	Meaning 1 appears only in BC. Meaning 4, 5 and 6 appear only in AD. In fact, these three meanings are clearly associated to the emergence of Christianism.

Lemma with its target or control word	CREDO (target word; control word: *licet*)
Number of meanings and meanings	1. to give as a loan, to loan, lend,
	2. to commit or consign something to one,
	3. to trust to or confide in a person or thing, to have confidence in, to trust,
	4. To trust one in their declarations, to believe,
	5. to believe a thing, hold or admit as true,
	6. to think, to suppose,
	7. to believe in God
Source pattern	1 = base
	1 --> 2
	2 --> 3
	3 --> 4
	3 --> 5
	5 --> 6
	5 --> 7
	Pattern: 1 base, 4 sources

Commentary	Meaning 7 is attested only AD, but very rarely (2 times). It must be also pointed out that in one of the two attestations there is an element implicitly indicating God: "non crediderunt in me" (Abaelardus).

Lemma with its target or control word	**ADSUMO (control word; target word: ITERO)**
Number of meanings and meanings	1. take to oneself 2. receive
Source pattern	1 = Base 1 --> 2 Pattern = 1 base, 1 source
Commentary	The analysis of the annotation of this word shows us that 20% of the contexts present a high level of vagueness. One of the provided meanings was considered to be identical to the meaning of the passage (with the other meaning annotated as being closely related) in only one of those vague passages. All the other uses of *adsumo* whose annotation corresponds to our interpretation of vague were evaluated by the annotator as its two senses being closely related to the meaning of the passage. There is a higher score of vagueness in the passages written during the Common Era. Although between the postclassical contexts that were annotated as vague we find different genres, three of them belong to *De medicina* by Aulus Cornelius Celsus. In this case, the annotator leaves in a comment the reason behind the lack of distinctness between the two meanings by highlighting the difficulty of the annotation of medical passages in which sometimes is unclear whether the patient has a passive or an active role.

Lemma with its target or control word	**SIMPLEX (control word; target word: FIDELIS)**
Number of meanings and meanings	1. simple, plain, uncompounded 2. without dissimulation, open, frank, straightforward
Source pattern	1 = Base

	1 --> 2 Pattern = 1 base, 1 source
Commentary	The comparison of the vagueness score between the two chronological periods reveals a higher score concerning classical Latin. The word *simplex* is a control word and as such we would expect a similar value in terms of vagueness in both eras. The annotation shows that most of the vague uses of *simplex* belong to rhetorical treatises. This is a context where "simple, plain, uncompounded" and "without dissimulation, open, frank, straightforward" can easily be seen as fitting. See as an example: "*Defensoris narratio simplicem et dilucidam expositionem debet habere* [...]" (*Rhetorica ad Herennium*, 2.3), 'The Statement of Facts of the defendant's counsel should contain a simple and clear account' (Caplan, 1954: 61).

Low vagueness score

Lemma with its target or control word	CONSILIUM (control word; target word: VIRTUS)
Number of meanings and meanings	1. A conclusion made with consideration, a determination, resolution, measure, plan, purpose, intention 2. understanding, judgment, wisdom, sense, penetration, prudence 3. the persons who deliberate, a council
Source pattern	1 = base 1 --> 2 1--> 3 Pattern: 1 base, 1 source
Commentary	*Consilium* is a control word: as such we do not expect a relevant change in the vagueness score between BC and CE. The vagueness score is in fact very low for the texts from BC and it does not increase significantly for the texts from CE. Closer analysis of the annotation: whenever the meaning "the persons who deliberate, a council" is selected with "4", the others almost never receive a value different from "1" (this happens in only one passage, where the annotator left a comment saying that the use was metaphorical and therefore meaning 2 was annotated with a "2"). There are a couple of annotations where the meaning

"understanding, judgment, wisdom, sense, penetration, prudence" gets a "4" and "a conclusion made with consideration, a determination, resolution, measure, plan, purpose, intention" gets a "2", and vice versa, but the rest of the annotation follows the pattern "4, 1, 1". We can conclude that the level of implication between the three meanings is very low. The reason why some vagueness is captured by the formula is that few passages are annotated with two "4", or some "0"s. The annotator justifies these cases in the comments field, addressing the difficulty of assigning a value to the meanings in those specific contexts.

Lemma with its target or control word	ANCILLA (control word; target word: HUMANITAS)
Number of meanings and meanings	1. maidservant, handmaid, female slave 2. someone servilely devoted to anything
Source pattern	1 = base 1 --> 2 Pattern: 1 base, 1 source
Commentary	The limited number of meanings might be interpreted as one of the reasons why the vagueness score for this word is so low. However, the case of *oportet* shows the weakness of this hypothesis. *Ancilla* is a control word, which is confirmed by the analysis of the vagueness score, which remains substantially unchanged in the texts from CE compared to those from BC. The finer analysis of the annotations shows an extremely regular pattern: the annotation is always "4, 1", except for one case of "4, 0", which seems to be a mistake of the annotator, who did not leave any comment explaining the anomaly – it is worth saying that the annotator is prolific in comments, which makes the absence of a comment for this passage peculiar. We can conclude that the relation of implication between the two meanings is almost absent. It is worth highlighting that the second meaning "someone servilely devoted to anything" is always annotated with a "1". This is true also for cases where the annotator commented on the presence of some metaphor, or specification of meaning not captured by the given meanings: "*Ancilla* as in 'servant of the gods'. Vestal virgins?", "This could be *ancilla* as the female version of *anculus* 'servant of the gods', seeing as the context is pagan religion."

Lemma with its target or control	NECESSARIUS (control word: correspondent target word: *beatus*)
Number of meanings and meanings	1. unavoidable, inevitable, indispensable 2. belonging, related, connected, bound
Source pattern	1 = base 1 --> 2 Pattern: 1 base, 1 source
Commentary	According to our calculation and as for *templum*, the meanings of *necessarius* are not at all vague. The annotator was always able to clearly distinguish the two meanings, though it must be noted that meaning 1 is much more frequent (54 out of 60) than meaning 2. Both meanings are attested BC and AD. Contrary to *templum* there does not seem to be a change in distribution between BC and AD.

Lemma with its target or control word	TEMPLUM (control word: target word: *casus*)
Number of meanings and meanings	1. a space marked out, an open place for observation, 2. a consecrated or sacred place, 3. a small timber
Source pattern	1 = base 1 --> 2 2 --> 3 Pattern: 1 base, 2 sources
Commentary	According to our calculation and as for *necessarius*, the meanings of *templum* have a vagueness score of 0. Meaning 1 and 2 are constitutively related as both indicate a kind of place. Meaning 3

	is more distant, as it designates an object. However, this meaning is never attested. There could be some bias in the case of this lemma, as in the rare cases where meaning 2 is not the fitting choice, meaning 3 is always excluded and meaning 1 is never selected as the fitting choice. Consider that meaning 2 is selected as the fitting choice in 50 cases out of 60.

Lemma with its target or control word	CONSUL (target; control: SENATUS)
Number of meanings and meanings	1. Consul 2. Proconsul 3. An epithet of Jupiter 4. Municipial official
Source pattern	1 = Base 1 --> 2 1 --> 3 3 --> 4 Pattern = 1 base, 1 source
Commentary	This is a target word and the meaning "municipial, official" emerged in the medieval era, however, this meaning was never assigned in the annotation task. The annotation shows a clear identification of the meaning "consul" for all contexts except one belonging to a Neo-Latin work. The use of *consul* in this passage was evaluated by the annotator as not corresponding to any of the proposed meanings, which explains why the vagueness score of this word in the Common Era has a negative value. The annotator left a comment saying that the passage contained a church-related meaning.

Lemma with its target or control word	COHORS
Number of meanings and meanings	1. A place enclosed around, a court, enclosure 2. a company of soldier, a division of an army, a cohort

	3. a crowd, multitude, company, throng, attendants 4. imperial court
Source pattern	1 = Base 1 --> 2 1 --> 4 2 --> 3 Pattern = 1 base, two sources
Commentary	*Cohors* is a target word with the meaning "imperial court" appearing linked to Christianity. The annotation of this word presents three contexts annotated with "cannot decide" for all meanings, two of them belonging to Titus Livius, which explains the negative value of the vagueness score in pre-classical/classical Latin. The formula did not include a step to account for the annotations with "0". Thus, in this case a passage with an annotation that we interpret as vague between "a company of soldiers, a division of an army, a cohort" and "a crowd, multitude, company, throng, attendants" was neutralized due to this caveat.

15