

(Dis)continuities in the diachrony of the Greek lexicon: The learned component in the light of a corpus analysis

Asimakis Fliatouras

Democritus University of Thrace/

Hellenic Open University

afliatou@helit.duth.gr

Konstantinos Sampanis

University of the Aegean

konstantinos.sampanis@yahoo.com

Abstract

A most notable feature of Standard Modern Greek (SMG) lexicon is the abundance of pairs of lexemes of similar meaning which belong to different registers. Thus, while one lexeme pertains to the colloquial language and is widely used both in oral and written speech, a semantically akin or even synonymous word of Ancient (i.e., Classical) Greek (AG) origin is also attested in specific contexts. The latter ones are usually described as part of the “archaic” or “learned” component of the Modern Greek (MG) lexicon. In the light of three case studies, our paper aims at investigating this variation by conducting research through diachronic corpora. We intend to dispense certain misconceptions concerning lexical change (which is erroneously interpreted as a straightforward replacement of an “older” word with a new one) and to introduce the concepts of adstratal and superstratal lexicon along with the theory of prototypicality and exaptation in the relevant research. Along with dealing with the diachrony of the Greek lexicon, our account sketches a first theoretical framework for investigating the lexicon of languages with a long-recorded history and diglossic contradictions. Accordingly, it emphasizes the vital role of corpora research in diachronic semantics and onomasiology.¹

Keywords: semantic change, onomasiological change, diachronic lexicon, Greek

1 Introductory Remarks: Onomasiological and semantic Change in Greek: the rise of lexical pairs

The Standard Modern Greek (MG) lexeme for the concept of HORSE is *άλογο άλογο*, a word etymologized as “the one without reason”², originally an adjective frequently preceding the Middle Greek (MidG) noun for ‘beast’, ‘animal’. So, the whole phrase was attested as *άλογον ζῶον άλογον ζῶον* ‘unreasoning beast’. The collocation was repeatedly referring to beasts accompanying soldiers, who were “the ones with reason”, into battle. Since the noun for ‘animal’ was often dropped and the phrase became elliptical³, and given that most of the beasts brought to military campaigns were horses, the word *άλογον* (or *άλογο* as pronounced in MG) shifted metonymically, through narrowing, from the initial sense into merely meaning ‘horse’.

If we consult a standard MG dictionary, we will also find the word *ίππος ίπος*, which derives from Ancient Greek (AG) *ίππος ήππος*, a word cognate to Latin *equus* (cf. Engl. *equine*) or Sanskrit *άśvaḥ* from Indo-European (IE) **h₁ékʷos* (cf. Beekes ²2011: 37). In MG, *ίππος ίπος* means ‘horse’ but ‘with a twist’: it is basically used either in concrete contexts, e.g. *δούρειος ίππος dhúrios ίπος* ‘Trojan

¹ We would like to thank the anonymous reviewers for their comments. We also cordially thank Christina Katsikadeli for her valuable feedback. Finally, we express our gratitude to Professor Emeritus Oswald Panagl for his remarks which have significantly enriched our text. Especially inspirational for our argumentation here has been Panagl (1989).

² Obviously associated with the word *logic*. The initial *a-* prefix is the so-called “alpha privativum” that negates a word.

³ According to our query in Thesaurus Linguae Graecae (TLG), the elliptical usage of the Middle Greek form *άλογον* is already attested in the 11th c., cf. *καί άνάπαυσον καί τόν στρατόν καί τά άλογα* “let the army and the beasts (horses?) rest” (Nicephorus Uranus, *Tactica*, 63.6). Retrieved from: <http://stephanus-tlg-uci-edu.uaccess.univie.ac.at/Iris/Cite?3091:001:3906>. In the same author: *άλλά οί έχοντες είτε βορδόνια είτε άλογα είτε όνικά* “those having either mules, or horses or asses” (Nicephorus Uranus, *Tactica*, 64.4), retrieved from: <http://stephanus-tlg-uci-edu.uaccess.univie.ac.at/Iris/Cite?3091:007:78406>. As of the 15th c. *άλογο(n)* the word is clearly attested frequently with today’s meaning.

Horse’ (literally ‘the wooden horse’ in AG) or metaphorically, e.g. as the unit of the horsepower measurement or the “knight” chess piece. The word stem is also used for morphological derivation and compounding, e.g. *ιπικό ipikó* ‘cavalry’, *ιπασία ipasia* ‘equestrianism’, *ιπποδύναμη ipodhinami* ‘horsepower’. In all the above cases, the word *ἵππος ipos* is generally considered an archaic and learned alternative to the colloquial word that has been preserved in the Greek lexicon due to the written tradition and a high degree of linguistic purism. We will return to that below.

Clearly, semantic and onomasiological change is not a novelty only to be found in Greek. Yet, what is “paradoxical” in MG is that numerous substituted lexemes did not just fade away but are still present in the modern lexicon even though they are characterized as archaic or learned lexemes and their usage is restricted to certain registers⁴. Nevertheless, we “dare” assume that – impressionistically – the abundance of +archaic/learned versus -archaic/learned lexical pairs in MG points to a great degree of systematicity of this lexical antithesis which is probably unattested in other standardized European languages. In Italian, for example, the lexeme for the concept of HORSE is *cavallo*, a word probably of Celtic origin⁵, while the Latin word *equus* is no longer in use except in derived lexemes, e.g. in the adjective *equino* ‘equine’. The “more modern” concept of ‘horsepower’ on the other hand was rendered in Italian as *cavallo-vapore*, so it seems that the Latin word is not productive in Italian even in the case of neologisms. On the contrary, in Greek both vernacular and learned lexemes for HORSE are systematically found, at least in written corpora and, as already mentioned, the learned word serves as the basis for morphological derivation and compounding. Compare the “hits” of the two words in two MG corpora of written speech:

Corpus	tokens of <i>álogho</i>	tokens of <i>ipos</i>
Hellenic National Corpus ⁶	1,488	778
Sketch Engine’s Greek Web 2014 (elTenTen14) Corpus	45,910	25,508

Table 1: Distribution of *álogho* and *ipos* in SMG

Evidently, both words are present in the modern language. As expected, *álogho*, i.e. the vernacular lexeme, is more frequent, yet *ipos* is widely attested as well. Despite the fact that the latter is frequently used in its technical sense as a unit of the horsepower measurement, still, it should be expected that the word’s semantics evoke the concept of a horse in the mental lexicon of a speaker, even if this “horse” bears a number of connotations or it is expected to be read (rather than heard) in particular registers and contexts. Simply said, no native speaker of MG will utter the word *ipos* in a trivial, daily conversation in reference to HORSE, yet almost everyone will understand what the word means. In fact, the word may also emerge in some contexts, e.g. when referring to the horses of the cavalry of a historical battle. Its etymology is then still totally transparent, but the two words are not substitutable and do not share a total synonymy connection. As has been shown, the learned MG lexicon is not always semantically accessible to primary school students⁷ and we could also expect similar results for illiterate or poorly educated individuals. Yet, in the case of several lexical pairs such as the one discussed above and those we are going to encounter in the following sections of this paper, both learned (or archaic/formal) and non-learned (or norm) lexemes are widely understood by most MG native speakers.

⁴ Cf. Fliatouras and Sampanis (forthc.).

⁵ Cf. <https://www.etimo.it/?term=cavallo&find=Cerca>, retrieved on 22/05/2020.

⁶ Accessible at http://hnc.ilsp.gr/index.php?current_page=main&lang=en, last count 24.11.20.

⁷ Cf. Martzoukou et al. (2019).

On the grounds of the previous introductory remarks, we now turn to a more detailed discussion on the dichotomy between the learned and non-learned lexicon in MG. In section 2, we refer to the historical origins of the MG lexical variation (2.1.), we introduce certain concepts that facilitate the understanding of onomasiological and semantic change within the diachrony of the Greek language (2.2. - 2.5.), while in the last subsection (2.6.) we present the corpora which we have used for tracing lexical changes, extracting examples and measuring statistical values. Section 3, in turn, is dedicated to three case studies of lexical pairs that illustrate the “diachronic antagonisms” and the semantic and onomasiological shifts in Greek. The examination of these three cases can also be considered as the basis for a future large-scale investigation of the factors and processes that are in play in this sort of “reorganization” of a language’s lexicon.

2 Theoretical and historical background and research framework

2.1 Historical context: Purism and Diglossia in Modern Greek

The attestation of an archaic/learned component in the lexicon of Standard MG is linked to the diglossia phenomenon and the parallel existence of a High (H), largely archaic, and a Low (L)/Vernacular variety (cf. Ferguson 1959). The dichotomy between a H and a L register has been pervasive in the history of the Greek language since Hellenistic times⁸ and has led to the “institutionalized” diglossia after the establishment of the independent Greek state in 1830, in which two varieties, the *dhimotiki* (= ‘Demotic’ roughly the “vernacular” and the *katharévusa* (= ‘the pure (language)’, i.e. the archaic one) competed with each other, with the latter being the official linguistic variety of the state until as late as 1976. There is a long list of publications on Greek diglossia, habitually referred to as the “Language Question” (for instance Browning 1982, Eideneier 2000 and most relevantly Mackridge 2009) which one can consult. It is beyond the scope of this paper to examine this sociolinguistic issue in detail, yet it is important to keep in mind that the standardization of the Modern Greek language reached its completion in the context of the antagonism between the H and the L variety, an issue not only linguistic but also deeply ideological/political that still stirs up emotions and conflict in Greece. This fact is self-evidently mirrored in the lexicon as well. Due to the long-standing and tenacious presence of diglossia, the impressionistic explanation for the rise of lexical pairs comprising an AG or archaic/learned versus a vernacular lexeme is that the “old word” still occurs as a relic of the once official archaic variety. “New words” entered Greek at some point either as loanwords or as novel formations of Greek origin. Once introduced into the vernacular, the new word replaced the AG one. The latter artificially survived only as an element of the learned written language as well as a statement in favor of the idea of the continuity of the Greek language⁹.

Still, there is a caveat: Although it is true that many words were reintroduced from AG to Standard MG through linguistic purism, it is also oversimplistic to explain all cases of “vernacular versus archaic” lexeme pairs along the scenario put forward above. Ancient words were not merely replaced by new ones, but there was a gradience as it is the case in all aspects of linguistic change. In reality, more meticulous corpora research reveals that:

⁸ Cf. e.g. Toufexis 2008 for Diglossia in Middle/Medieval (“Byzantine”) Greek.

⁹ Mackridge (2009: viii), “In the minds of almost all Greeks, their language has also been connected, in an unbroken continuity upheld by the traditions of the Orthodox Church and by the chanting of texts in the liturgy, with the New Testament and other early ecclesiastical texts. This has given educated Greeks a sense that their nation possesses a unique cultural heritage. Their language both distinguishes them from all other modern nations and connects them with the civilization of ancient Hellas, early Christianity, and Byzantium. It is largely this complex connection between contemporary and older culture that has given rise to the development of the Greek national identity in modern times. In no other area has this connection been so problematic as in language.”

a.) Certain learned/archaic lexical elements are attested throughout. Thus, they were not “retrieved” from the past or revived from their linguistic death, but they were constantly in use in the written speech. Surely, these archaic/learned words were what their designation indicates: archaic and/or learned. Nevertheless, there is a significant difference between a theory suggesting that all learned words were virtually obsolete and reintroduced and another theory showing that several learned words were frequently used, notwithstanding their high register “flavor”.

b.) In some cases, an AG lexeme shifted from a prototypical meaning to a more “peripheral” one. For example, the AG word *άρτος* *ártos* ‘bread’ was replaced by *ψωμί* *psomí*. The latter is the colloquial form, yet the former is now usually used in the religious vocabulary (e.g. at the Holy Communion) and in the bakery industry (somehow nuanced with a more professional or poetic flavor). Thus, although the words are semantically synonymous, their pragmatics are considerably different.

Given these remarks, we suggest that the Standard Modern Greek lexicon was developed under the influence of two opposite but intertwined tendencies: not only replacement of words but also the preservation of old words were factors. What seems to be a contradiction at face value can be actually clarified with the aid of the following theoretical concepts we present below.

2.2 The Standard Modern Greek lexicon and its learned component: Lexical adstrate and superstrate

In broad strokes, the MG lexicon consists of four major components (cf. Mackridge 1985: 310):

a.) AG lexicon. This includes words retained unchanged or slightly changed morphologically and semantically, e.g. *θάλασσα* (*thálasa*) ‘sea’.

b.) New formations through “internal borrowing” and/or re-borrowing via European languages, namely revived/reintroduced and (in most cases) semantically changed, e.g. AG *λύκειον* (*likion*) via French *lycée* ‘high school’ (cf. Petrounias 2017).

c.) Neologisms and derivatives of AG words coined in MG, e.g. *λεωφορείο* (*leoforío*) ‘bus’ (< Attic *λεώς* (*leós*) ‘people’ + root *φερ/φορ-* ‘carrying’), and

d.) Loanwords (“external borrowing”), namely words mainly stemming from Latin and Romance languages, Turkish, French, English (traditionally, loanwords already found in Classical Greek are not considered as such), e.g.: *κέφι* (*kéfi*) ‘mood; high spirits’ < Turk. *keyif* ‘joy’, ‘high spirits’.

While lexical innovations rendered a significant part of AG lexemes obsolete, a part thereof still constitutes the “learned component” of the Standard MG lexicon. The learned level involves an artificial language stratum, which leads to the maintenance of AG types due to the prescriptivism and the standardization of Ancient Greek. Due to the phenomenon of diglossia which we discussed in the previous section, there was a deep gap between the written discourse (the register of which was preponderantly archaic) and the oral discourse which –by definition– is associated with the vernacular. The latter considerably differed from the written language. Consequently, diglossia led to the “artificial”, erudite survival of AG varietal elements which after the establishment of the Greek state in 1830 were sumptuously used in the language of administration, high oral/written registers, the scientific register, the military language, as well as in the language of the church (and even in literature) and, hence, their usage was ubiquitous in all higher registers. This component of the lexicon was linked to the purist *katharévusa* we already referred to. This should be rather seen as a linguistic fossil (cf. Anastassiadis-Symeonidis & Fliatouras 2019) and an artificial interference in the modern vernacular language. Thus, unless an AG word survived as such in Hellenistic, Medieval and Modern Greek, the MG lexicon displayed two major sources of natural (i.e. without the intervention of scholars and

educational institutions) innovation: a.) inter-systemic (AG words are replaced by other Greek words) e.g. AG ὕδωρ *hýdōr* versus HK νερό *neró* ‘water’ and b.) extra-systemic replacement (AG words are replaced by loans, mainly Latin in the period of Hellenistic/Koinē Greek), e.g. AG πλαγγών (*plangōn*) versus HK κούκλα (*kúkla* < Lat. *cuculla* ‘puppet’).

Irrespective of the artificiality of the perseverance of AG lexemes in the modern language, it is a fact that anyone thoroughly examining the Modern Greek lexicon, for example as a non-native speaker learning the language, will be confronted with lexical pairs like *álogho* versus *ípos* (‘horse’) as discussed in the introductory section. For instance, in Table 2 we present certain AG lexemes that have been replaced by new formations or Latin loanwords in vernacular Greek.¹⁰ Yet, the ancient words, phonologically and morphologically adapted to the MG grammatical system, are still attested in higher registers, some of them more frequently and some in specific contexts with specialized semantics or pragmatics:¹¹

Core Meaning	AG Lexeme	MG Lexeme	Etymology of the MG Lexeme
FIRE	πῦρ (pýr)	φωτιά (fotiá)	< AG φῶς (phōs) ‘light’
WATER	ὑδωρ (hýdōr)	νερό (neró)	< coll. νεαρόν ὑδωρ (nearón hýdōr) ‘fresh water’
LIVER	ἥπαρ (hêpar)	συκώτι (sikóti)	< ἥπαρ συκωτόν (hêpar sykōtón) (< σῦκον ‘fig’) ‘liver full of figs’
NOSE	ῥίς (rhís)	μύτη (míti)	of unknown origin
EYE	ὄφθαλμός (ophthalmós)	μάτι (máti) ¹²	AG ὀμμάτιον (ommátion) < ὄμμα ‘look’
FISH	ἰχθύς (ikhthýs)	ψάρι (psári)	ὀψάριον (< ὄψον ‘cooked food’)
BONE	ὀστοῦν (ostoûn)	κόκκαλο (kókalo)	κόκκαλος ‘pumice, pine’
RED	ἐρυθρός (erythrós)	κόκκινος (kókinos)	κόκκος ‘painting granule from pine’
CYAN (SKY BLUE)	κυανοῦς (kuanoûs)	γαλάζιος (ghalázios)	κάλαϊς ‘blue precious stone’

¹⁰ It is noteworthy that the respective archaic words are retained in dialects, e.g. *ὄρνιθα* ‘chicken’ in Crete and Cyprus (cf. SMG *κότα*), *λευκός* ‘white’ in Tsakoni as *λεκός* (cf. SMG *άσπρος*) and *θύρα* ‘door’ (cf. SMG *πόρτα*) in the Cappadocian dialects with various variants, such as *θύρ(α)*, *χύρα*, *τύρα*, *σύρα* (cf. Dawkins 1916: 603). It will be of particular interest to investigate the diachronic course of archaic elements in dialects to support both the artificiality of “learnedization” and the register “de-learnedization” as far as SMG is concerned.

¹¹ It seems that specific semantic/word fields are more prone to lexical change/innovation, such as domestic animals and colors. For example, Romance languages borrowed only a small number of lexemes from Germanic, but they adopted several color terms, such as *blanc* or names for “hues of blue”; the investigation of these word fields and the “lexical solidarity” (according to E. Coseriu) of specific color terms goes beyond the scope of this paper; for further studies on the diachrony and borrowing of color terms cf. also Biggam 2012.

¹² In Modern Greek *μάτι*, the root *οπ-* *op-* has vanished and what is preserved is a combination of two suffixes: from **οπ-μάτιον* > *ὀμματιον* *ommátion* (after regressive assimilation), diminutive of *ὄμμα* *ómma* ‘eye’. The root *op-* is retained in learned lexemes, such as *οπτική* *optikí* ‘point of view’, cf. English *optics*.

MOUSE	μῦς (mûs)	ποντίκι (pontíki)	ποντικός μῦς (<i>pontikós mûs</i> = either mouse of the sea or Pontic (Black Sea) or mouse of the sea in general)
SHIP	ναῦς (naûs)	πλοίο (plío)	πλοῖον (< πλέω ‘sail’) ‘merchant/transport ship’

Table 2: Examples of lexical change in the Greek language

In the examples above, we observe that the learned lexicon was not always “resurrected” in the MG written language but in many cases some lexemes exhibit a continuity in their usage. As a matter of fact, following Anastasiadis-Symeonidis & Fliatouras (2019), the learned level stems from two distinct but interacting paths:

a.) The adstrate lexical layer: this involves the continuous maintenance of AG words via written speech and/or written tradition. The adstrate emerges beginning in the era of the Hellenistic Koinē (HK, 3rd B.C. – 6th AD) and, mutatis mutandis, it is still going on to this day. In the purist form of the Katharévusa archaic register, the adstrate has been dominant. The adstrate originates from this part of the Attic Greek lexicon that has been replaced in the Vernacular of the HK. It was during that linguistic period that the first reactions to language change became vocal, and the most conspicuous example is the revivalist linguistic movement of “Atticism” (first quarter of the 1st c. BC) that emerged as a reaction to the Koinē Greek of the Hellenistic and Roman Period¹³. The “tradition” of reaction to the vernacular led to the systematization of Diglossia up to the modern times, as already mentioned above. The *ἵππος* *ípos* example we examined in the introduction is such an adstrate survival case. Similarly, instances of the adstrate lexical layer are all the archaic/learned words we are going to examine in our case studies below.

b.) The superstrate lexical layer: This was shaped as the outcome of a lexical revival/renaissance of AG words from 18th c. AD onwards, imposed by movements of nationalist linguistic purism that swept most modern languages on their way to standardization¹⁴. Contrary to other languages (such as German or Turkish, for example), the purist movement in Greek was not instantiated by the rise of the Vernacular to the status of standard language but by the revival of AG grammatical features and lexicon which was reflected on the aforementioned dichotomy between the Demotic (“dhimotikí”) and the Archaic (“katharévusa”) varieties. The impetus for the usage of AG lexemes was naturally the wish to remove foreign loanwords but also the effort to emphasize the continuity of the Greek nation since the ancient era as a counterargument against Fallmerayer’s theory which claimed that Ancient and Modern Greeks were not racially linked. In this process of artificial “ancientization”, Homeric and Attic words were now used for concepts of the modern times, e.g. the Hellenistic *ἐφημερίς* *ephēmeris* (‘military diary’) was “mobilized” in the meaning of ‘journal’, ‘newspaper’ in order to replace the colloquial loanword *γάζετα* *ghazéta*.¹⁵ Similarly, the word *διαβατήριον* *dhiavatírion*, which was associated with sacrifices before a journey to foreign lands in Ancient Greece, was used instead of colloquial *πασαπόρτο(ν)* *pasapóрто(n)* ‘passport’. Interestingly, the once colloquial loanwords we referred to in the examples are

¹³ Horrocks (2010: 133ff and 155f). Cf. also Kim (2010: 470): “The most striking contrast between Atticist and colloquial language, however, is in vocabulary; Atticizing writers avoid using words not attested in Classical texts, substituting the Attic equivalent. To take a popular example (Browning 1983: 47; Zgusta 1980: 126; Horrocks 1997a: 94), the late second-century lexicographer Phrynichus (Selection 10) recommends the Attic *χάριν εἰδέναι* “to give thanks” instead of *εὐχαριστεῖν*, a postclassical word that frequently appears in the New Testament.”

¹⁴ On linguistic purism, cf. Thomas (1991), Brunstad (2003), Hock and Joseph (2009: 265), Langer and Nesse (2012).

¹⁵ An interesting parallel to *ἐφημερίς* as a military diary is Latin *commentarius* in its use in the title of Caesar's work.

barely used nowadays in Standard MG, a fact that underscores the success of imposing this superstrate on the modern language.

Summing up this section, we can draw a sketch of the diachrony of the “indigenous” (i.e. not borrowed from other languages) Greek lexicon: Along with the natural evolution of the AG vocabulary into Hellenistic Koinē (HK), Medieval Greek (MEDG) and MG, there is a superstrate of words that were reintroduced in the modern language from AG in an effort to “purify” the modern lexicon. Yet, an adstrate of archaic/learned lexemes was constantly used in the written speech and, as will be illustrated later, in some cases the “old” words grew semantically and pragmatically distinct from their new counterparts. In this sense, the pairs we are going to examine were never perfectly synonymous neither in terms of their stylistic distribution¹⁶ – and most crucially – nor in their denotation, since subtle connotational differences are in play. This point will become clear when we present our case studies below.

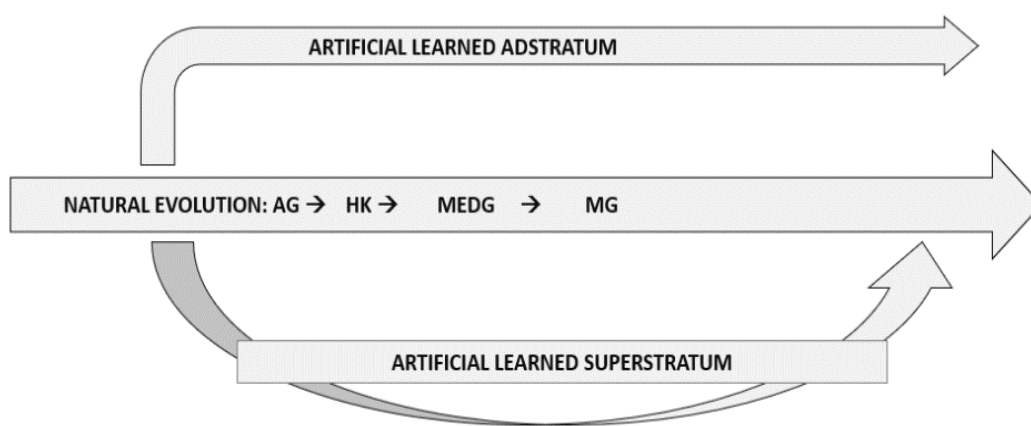


Diagram 1: The learned lexicon as an adstrate and superstrate layer

2.3 Prototypicality and Lexical Change

Another crucial concept in our discussion is the one of *prototypicality*. In cognitive semantics, prototypicality is usually understood as the inclination of the human mind to evoke certain mental images rather than others when a word is used in speech¹⁷. This is rather conspicuous in the case of common nouns. For example, whenever the word ‘bird’ is uttered or read, the mental image that will pop up into our mind would be a sparrow or a pigeon rather than a penguin or an ostrich. A prototype is then “the characteristic instance” of a category, a hyponym that can represent a whole category (in the case of birds a species or a genus as representative of the class of birds) par excellence. Consequently, the prototypical meaning is closer to the “conceptual” (i.e. literal) rather than to the “associative” (i.e. metaphorical, metonymical, synecdochical or connotational) use of a word.

In our analysis, we use the notion of prototypicality in a modified fashion. This does not refer to the “characteristic instant” of a hyperonym but to the most literal connotation-free meaning of a lexeme, the one usually describing a whole class, a hyperonym itself or, simply, one concrete category (item, quality, action etc.) straightforwardly. Based on the idea that “linguistic categories may be fuzzy at the edges but clear in the center” (Geeraerts 2010: 183), we propose that lexemes that newly enter a

¹⁶ That there is no “absolute” synonymy is of course barely a novel idea. Saeed (³2009: 65) remarks: “As Palmer (1981) notes, the synonyms often have different distributions along a number of parameters.”

¹⁷ Cf. Saeed (³2009: 37f., 355ff.) and Geeraerts (2010: 182ff.). The concept of *prototypicality* goes back to Rosch’s (1973) seminal study on prototypes and basic level categories. Cf. also Taylor (²1995) and Dressler (1987).

lexicon typically appear in the periphery of prototypicality, in the sense that they are imbued with metaphorical/ metonymical/ ironical/ synecdochical/ hyperbolic¹⁸ co-meanings (mostly connotational), or they just refer to specific hyponyms of a superordinate class. Once these lexemes enter the lexicon, they may tend to move to the core of the prototypical meaning, depending on the frequency of their usage¹⁹ and the diffusion in various contexts. In turn, the new “prototypical” word starts elbowing the original prototypical word to the periphery and the latter gets more nuanced with connotations, more specified or it just turns archaic and is then attested in certain categories of contexts and registers. The systemic process of lexical change described above may be also enforced by extra-systemic reasons, for example prestige or ideology connected with historical conditions, such as conquest, language contact etc.²⁰

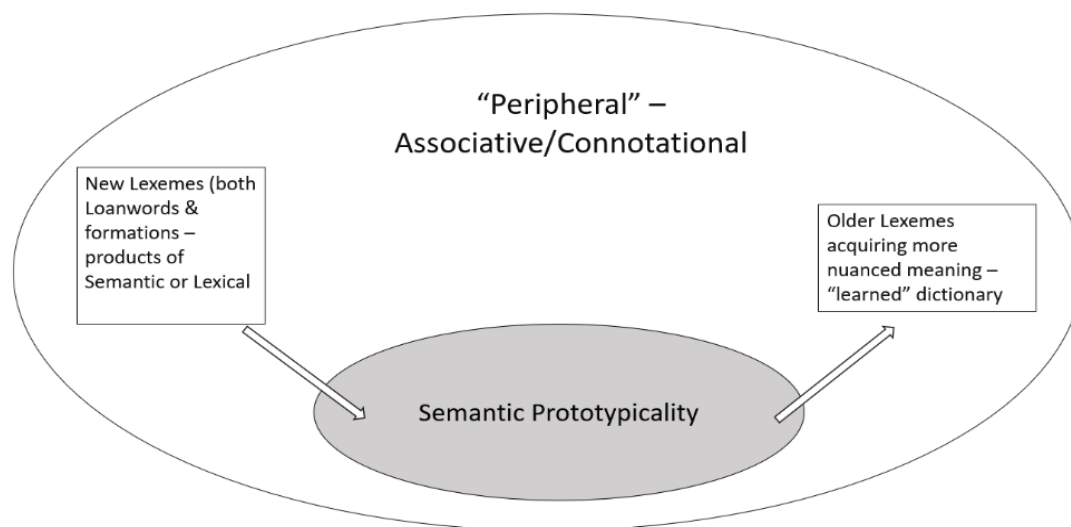


Diagram 2: Shift towards semantic prototypicality in cases of lexical pairs

Although our paper will examine three case studies in Greek, it is interesting to note an example in German that illustrates that this lexical, semantic and onomasiological process could possibly constitute a kind of pattern and the base for a theoretical model. Specifically, in Old and Middle High German, the common word for ‘head’ was *houbit* or *houbet* which is of course cognate with English *head* and further with Lat. *caput*. This word survives in modern Standard High German (SHG) as *Haupt* and it can still mean ‘head’ but as a marker of higher (“gehoben”) register or nuanced with irony (“umgangssprachlich”, “scherzhaft”) or it is used metaphorically (e.g. *das Haupt der Kirche* ‘the head of the church’). Yet, *Haupt* is mostly found in the sense ‘main’ or ‘principal’ and usually as a compound constituent, e.g. *Hauptbahnhof* ‘main railway station’. It was actually under the influence of Old/Middle

¹⁸ Traugott (2010: 281), “Lehmann (1985) suggested that speakers wish to be ‘expressive’, Haspelmath (1999) that they wish to be ‘extravagant’...”.

¹⁹ It is enormously difficult to provide a threshold of how frequently a word should occur in order to start losing “peripheral” connotations and start shifting towards the prototypical meaning. Such research would demand a large-scale examination of historical corpora and vigorous statistics. Still, it would not be reliable about the usage in the vernacular varieties of the past for self-evident reasons that have to do with the orality of vernacular varieties. In our study, we show that lexical replacement of basic concepts used in daily speech and semantic changes in general can only be traced after meticulous observations. This is especially true for Greek, a language with a diachronically persistent diglossic tradition.

²⁰ Our take on the semantic shift of certain newly introduced lexemes towards prototypicality does not disregard the role of hedging (cf. Lakoff 1972) in semantic change. On the contrary, it presupposes it, since this enables new lexemes to enter a language (bearing initially “peripheral” meanings with ironic or comical connotations, for example), while “older” words start becoming hedges that express a variety of the speakers’ stances. Nevertheless, as we show below in the case of co-existence of lexical pairs, we clearly observe how the new lexemes headed toward describing prototypical, less connotational concepts. Thus, our model seems to work for cases of lexical variation which is based on a +/-learned dichotomy.

High German forms such as *houbetstat* ‘capital city’ (SHG: *Hauptstadt*) that the word *Haupt* acquired the metaphorical meaning, which is in the modern standard language the main one.²¹

Today’s prototypical word for HEAD in SHG is *Kopf* deriving from Lat. *cuppa* ‘cup’. In Middle High German the word *kopf* was often found as a “sarcastic” metaphor for an enemy’s head cut off in the battle (Cf. Fritz 2006: 107) and, gradually, its usage was generalized so that it is now the default word for ‘head’. For example, one can use both *Haupt* or *Kopf der Firma* with a meaning differentiation (the former means ‘head of the company’ while the latter can be rendered in English as “the mastermind of the company”). Still, replacing *Kopf* with *Haupt* in the mundane expression *Mein Kopf tut weh* “my head hurts” is blocked. Thus, we observe the pattern: *Haupt* was replaced by *Kopf*, the latter stemming from the periphery of prototypicality (not a simple head but an enemy’s loathed head that looks like a cup when cut!) and gradually moving towards the core of prototypicality and pushing *Haupt* to the periphery (*Haupt* constituting now a learned alternative for ‘head’ or meaning ‘main’, ‘principal’).

2.4 The Concept of Exaptation

In broad strokes, adaption involves the cases of branching/non-linear language change. Lexical adaptation can be traced, when a lexical change in norm does not lead to obsolescence/replacement of the origin word but to its restriction to the peripheral lexicon. As a result, the learned component in SMG could easily be considered a case of an extended artificial lexical exaptation, since it is equivalent to the archaic high register as a hyponym of the high register.

Specifically, according to Anastassiadis-Symeonidis & Fliatouras (2019), most learned items in SMG display some semantic/pragmatic/stylistic specialization, since they develop a more metaphoric, intensive, or metonymic meaning, are semantically ameliorated representing ‘higher’ or ‘more prestigious’ notions of science, trade or religion and can be found in ‘politically correct’ utterances, e.g. *παλαιός paleos* (= learned) means ‘unforgettable, gorgeous, valuable’²², whereas *παλιός palyós* (= non-learned) means ‘useless’, *πιπεριές Φλωρίνης piperies FlorinisGEN* (=learned) is used as a brand name and refers to a specific kind of pepper from the town of Florina (periphery of Macedonia in Northern Greece) in comparison to *πιπεριές Φλώρινας piperies FlórinasGEN* (= non-learned) which refers to all peppers from Florina, *οίκος ikos* (learned) ‘house’ (vs non-learned *σπίτι spíti* ‘house’) refers to buildings with prestigious or specialized function, such as *εκδοτικός οίκος ekdhotikós ikos* ‘publishing house’, etc. (see Anastassiadis-Symeonidis & Fliatouras 2019). The distribution of the learned variation in SMG may be either supplementary (X or Z) or overlapping (X and/or Z):

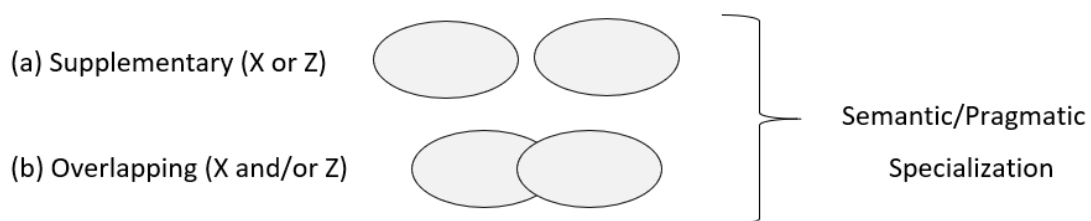


Diagram 3: Distribution of learned variation in SMG

²¹ The designation for the human head in German is sociolinguistically conditioned, since it has gradually seeped into the standard language from a special language of the military via its distribution in the families of the soldiers.

²² Oswald Panagl indicated an interesting semantic/pragmatic parallel in German: the adjective *historisch* (‘historic(al)’) can also bear the meaning *bedeutungsvoll* (‘significant’). Any event of the past is recalled if it is worthy for being recalled, hence a reference to the past survives through its significance.

2.5 The corpus research

To testify the validity of the theoretical framework, we have presented in the previous section (2.) that it is essential to examine how the tripartite methodological apparatus of the adstrate versus superstrate lexical layers, prototypicality and exaptation can be applied in semantic and onomasiological change in order to account for the emergence of the learned vs. vernacular lexical pairs. Towards this goal, we should search corpora for a) the very first attestations and usages of a “new word”, b) the distribution and the semantics of the “new word” with respect to the AG “old word”, c) the gradual decline and archaization of the AG word, d) finally, the (adstratal) continuity in the usage of the AG word, the distribution of the AG/learned and the vernacular word in Standard MG. We focus on writers using both lexemes in parallel within the same text. While the diachronic semantic content of the lexical pairs is significant, in our diachronic analysis we primarily take heed of the cooccurrence of the lexemes at a certain period of the language. Thus, in early stages the lexemes were not synonymous in most cases and only gradually they became semantically akin to each other. Our research does not capture the gradience in semantic change, yet it provides a methodological tool for doing so in the future. Obviously, this research agenda is ambitious, thus it cannot be discussed exhaustively in this article, since our main goal is to discuss how investigation into digital corpora facilitates the examination of lexical change. Nevertheless, our work presented herein serves to pave the way for more systematic, large-scale research in the future.

The main tool we use for diachronic investigation is the renowned *Thesaurus Linguae Graecae* (link: <http://stephanus.tlg.uci.edu/>), probably the most important historical corpus not only of Greek but of any language. The corpus covers the period from the 8th c. BC to the 18th c. AD²³ and it is still expanding. One of its assets is that, apart from its vast collection of texts, it offers a number of tools facilitating thorough textual investigation, such as lexica, intertextual browsing, n-grams, and lexical statistics.

Especially for the period between the 16th – 18th c. AD, we made use of a pre-published version of an anthology of Early Modern Greek Texts (edited by Kakoulidi-Panou et al., in press) which comprises a representative collection of texts of that period, not yet included in the TLG.

For the examination of the distribution of each part of the lexical pairs in the Modern language, we utilized the two corpora: a.) The *Hellenic National Corpus* developed by the Institute for Language and Speech Processing/R.C. “Athena” and b.) the *Greek Web 2014 (elTenTen14) Corpus* available on Sketch Engine. For both corpora, we provided relevant links in the introductory section.

3 Case studies

In our introductory chapter we dealt with the example of a lexical pair comprising the colloquial *álogho* versus the learned *ípos* ‘horse’. The history behind the rise of this lexical pair is quite straightforward. The etymology of the “new” word and the semantic path it followed from the ancient to the modern meaning is crystal-clear and well-attested in the corpora. In this chapter, we are going to investigate the diachronic distribution and the semantic shifts of three additional pairs of lexemes in which the process of gradual lexical change may be more obscure than the already examined case.

In accordance with the historical and theoretical framework provided in the previous sections, the lexical pairs we will examine contain a learned (+learned) and a non-learned (-learned)²⁴ component: (a) θύρα

²³ For the history of TLG, visit <http://stephanus.tlg.uci.edu/history.php> and cf. Pantelia 2000. For the contribution of TLG in Digital Humanities, cf. Brunner 1991.

²⁴ One should consider that a learned lexeme is necessarily of AG origin, but not every AG word is a learned one. Words such as θάλασσα *thálasa* ‘sea’ or ουρανός *uranós* ‘sky’ are perfectly colloquial irrespective of their AG origin and therefore non-learned. The “tags” learned and non-learned refer to both the register status and peripheral/fossilization/markedness of the

thíra vs. *πόρτα πόρτα* ‘door’, (b) *λευκός lefkós* vs. *άσπρος áspros* ‘white’, (c) *οίκος íkos* vs. *σπίτι spíti* ‘house’. All non-learned lexemes under examination entered Greek as loanwords from Latin. As we will demonstrate, the new loanwords *πόρτα*, *άσπρος* and *σπίτι* have been incorporated into the norm of SMG, whereas the AG origin/initial ones have been maintained in the peripheral learned component of SMG via the learned adstratal written discourse.

3.1 Case study I: *θύρα thýra* vs. *πόρτα pórta*

The AG word *θύρα thýra* (pronounced *thíra* in MG) belongs to the inherited Indo-European lexicon (cognate to Sanskrit *dvār-as*, Latin *forēs*, German *Tür* ‘door’). In its prototypical meaning DOOR, the ancient word has been largely replaced by *πόρτα πόρτα* in SMG. The word straightforwardly derives from Latin *porta*, denoting ‘a gate of a city or a town’²⁵. In Greek, the first attestation of *πόρτα πόρτα* can be traced to Claudius Ptolemeus 2nd c. AD and his reference to *Πόρτα Αύγουστα* ‘Porta Augusta’ in his *Geographia*²⁶, obviously a direct transfer from Latin.²⁷

The written records of the Late Ancient, Middle and Early Modern Greek periods point to a stepwise replacement of the ancient word with the new loanword. Self-evidently, these attestations do not depict the exact pace of the lexical shift in the vernacular, due to the conservative character of the register of the written records of most of the examined era. Nevertheless, according to the TLG corpus it seems that the usage of the lexeme *θύρα* was persistent for a long period of time, as the scheme below suggests:

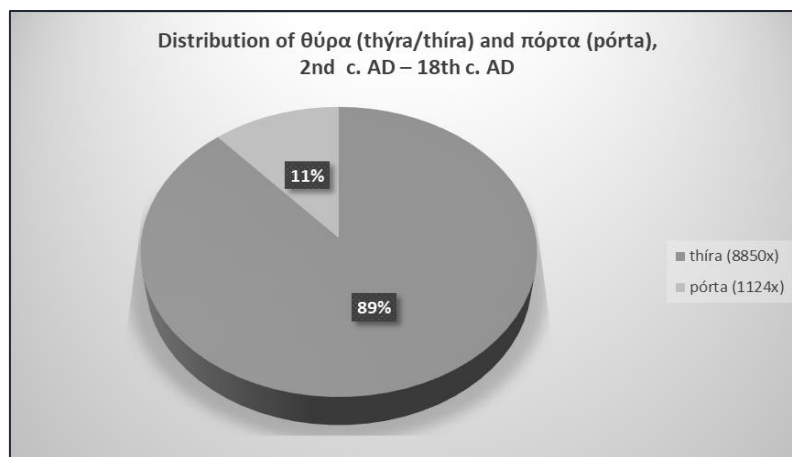


Table 3: Distribution of *θύρα (thýra/thíra)* and *πόρτα (pórta)*, 2nd c. AD – 18th c. AD in TLG

As can be seen from the anthology of Early Modern Greek, however, the difference in distribution between the two words decreases significantly after Medieval Greek, proving that the new word tends to gradually replace the older one:

words within Modern Greek and not to the etymology per se, although of course this can explain why a word is learned or non-learned.

²⁵ Cf. Glare ²2012: s.v. *porta*.

²⁶ Retrieved from: <http://stephanus-tlg-uci-edu.uaccess.univie.ac.at/Iris/Cite?0363:015:135365>.

²⁷ We should note that this first attestation does not bear the later generic meaning. As we will show later, the borrowing of *porta* was reinforced by proper names. Namely it is possible that microtoponyms constituted the initial form pattern, to which the semantic change was applied.

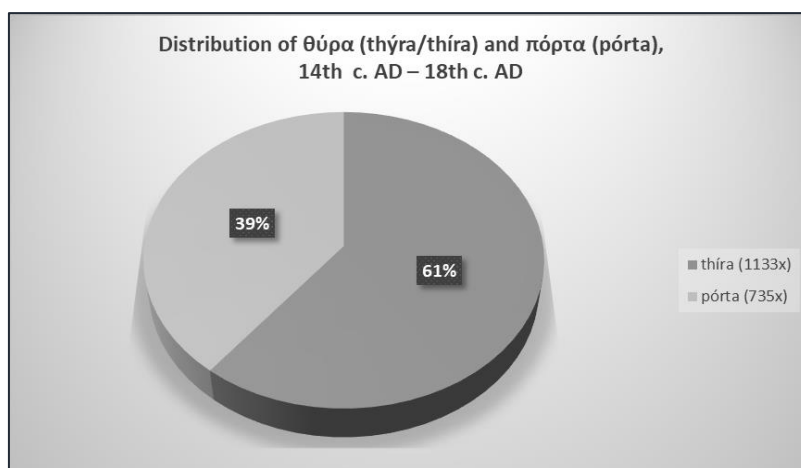
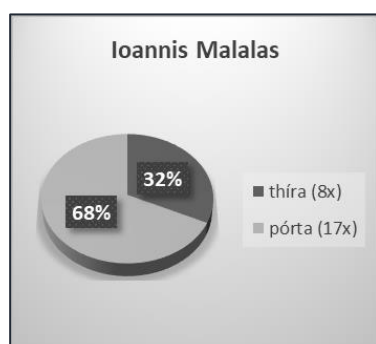


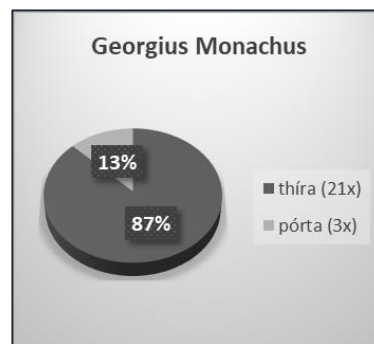
Table 4: Distribution of θύρα (*thýra/thíra*) and πόρτα (*pórta*), 2nd c. AD – 18th c. AD in EMG

Patterns of distribution of the two lexemes may vary in the work of different authors, which use both competitive words, as the tables below illustrate²⁸:

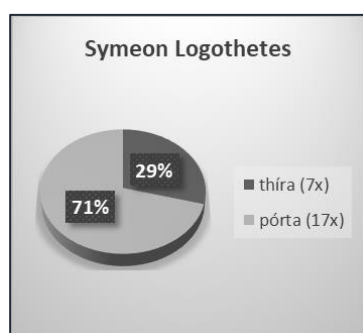
i) Ioannis Malalas (5th – 6th c.):



ii) Georgius Monachus (9th c.):



iii) Symeon Logothetes (10th c.)



iv) Anna Komnene (11th – 12th c.)

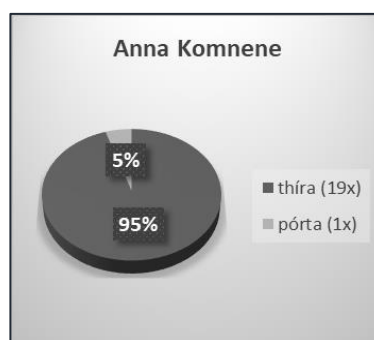


Table 5: Distribution of θύρα (*thýra/thíra*) and πόρτα (*pórta*) in various authors

The corpora investigation reveals that the word *porta* entered the Greek texts in reference to various gates or entryways of Constantinople. Apparently, this lexical usage corresponded to toponyms of

²⁸ The distribution of the lexemes through the centuries does not capture semantic shifts, which means, for example, that the sense of the word *pórta* changed gradually. The semantic change of the lexical pair is discussed below. The tables however are useful because they present the contextual co-occurrence of both lexemes. How the exact semantic change took place in the vernacular is not always easy to postulate, yet below we present some scenarios for each lexical pair.

similar emblematic landmarks of Rome which equally bore the Latin word *porta*. Thus, the lexeme was probably imbued with a certain degree of “Roman ideology”, so that Constantinople, the new capital of the Roman Empire (also called “the New Rome”) was linked to the old capital, the Old Rome²⁹. The “*porta*-landmarks” in Rome were abundant: e.g. *Porta Fontinalis*, *Porta Trigemina*, *Porta Capena*³⁰, whereas Varro (in his work *De lingua Latina: V 163*) mentions ‘Porta Naevia’ in Rome. The cases in Greek are also numerous, for example Πόρτα Αὐγούστα (< Lat. *Porta Augusta*), Χρυσέα/Χρυσή Πόρτα Chrysea/Chrysē *Porta* ‘the Golden Gate’ (< Lat. *Porta Aurea*), etc. The designation “Sublime/High/(Ottoman) Porte” for the main gate of the *Topkapı Palace* and, synecdochically, for the Ottoman Government does not seem to be a coincidence.

As observed in early texts (namely in Malalas) the two words θύρα *thýra* and πόρτα *pórta* co-occur within a sentence. Obviously, the two words are not synonymous, with *pórta* clearly still bearing the original Latin meaning:

(1)

“δὲ αὐτὸς Θεοδοσίος ἐχρύσωσε καὶ τῆς Δαφνητικῆς Πόρτας δύο θύρας τὰς χαλκᾶς καθ’ ὁμοίότητα ἣς ἐχρύσωσε πόρτας ἐν Κωνσταντινουπόλει, ἣτις καλεῖται ἕως ἄρτι ἡ χρυσέα πόρτα”. (Malalas, *Chron.* LXIV 62)

“Theodosius also gilded the two bronze **doors** of the Daphnetic Gate [in Antioch], after the pattern of the gate he had gilded in Constantinople, which up to the present, is still called the Golden Gate”.

During the span from the 2nd c. AD to the 14th c. AD, θύρα *thýra/thíra* attestations are still more frequent, which means either that *pórta* retained its initial meaning of ‘gate’ for a longer time and the two lexemes *thíra* and *pórta* kept referring to different things, or that the word *pórta* was not considered appropriate for the written language; it was more “colloquial”. Yet, since *pórta* was introduced into the texts with specific semantics, the meaning ‘door’ gradually evolved in colloquial speech. By the 16th c., the new semantics of the lexeme seem to have been established in common parlance, as the Early Modern Greek example below reveals:

(2)

εκάθετον εἰς τὴν πόρταν τοῦ σπιτίου καὶ ἐκλαίειν
“He was sitting at the house door and was crying”

Compare an excerpt belonging to a more “learned” register:

(3)

Τετάρτη θύρα ἐστὶν ἡ πόρτα τῶν Ψαροπουλιῶν, καὶ οἱ Ἀγαρηνοὶ λέγουσιν αὐτὴν κατὰ τὴν διάλεξιν τῆς γλώττης αὐτῶν Μπαλοῦκ καπισί. (*Anthology* 118, 16 c.)

“The Fourth Gate is the seabirds’ (seagulls’) *Porta*, which Agarenes (= i.e., the Turks) call Baluk Kapisi (i.e., = Balık Kapısı).”

²⁹ The influx of Latin words in Early Middle (i.e., “Byzantine”) Greek was less a matter of language contact in the colloquial registers. What seems to have applied in this case was the prestigious character of the Latin vocabulary as an integral part of the legal, administrative, and military life (cf. Adrados 2005: 250). As Alexiou (2002: 23) phrased it, “With Latin as the language of central administration from the first century B.C. to the sixth century A.D., changes in spoken Greek were rapid, giving rise to new dialect differentiation within the koine still evident in modern dialects”. The political ideology underlying the constant attempt to associate the Eastern Roman Empire with its past is lucidly described in Kaldellis (2019).

³⁰ Cf. Richardson 1992: passim.

In SMG, the same lexemes have opposite meanings: *πόρτα* *pórta* refers to DOOR (introduced in written speech, probably not colloquial initially), whereas *θύρα* *thíra* refers to GATE, ENTRANCE (e.g. of a huge building, a stadium, a city, etc.), also reserved for technological items, e.g. USB *θύρα* ‘USB port’ or it is an archaic/learned word for ‘door’. *Thíra* is also used in compounding, e.g. the learned/coined compound form *θυροτηλέφωνο* *thirotiléfono* ‘door/entry phone’.

Corpus	tokens of <i>thíra</i>	tokens of <i>pórta</i>
Hellenic National Corpus ³¹	1,089	7,586
Sketch Engine’s Greek Web 2014 (elTenTen14) Corpus	33,439	147,584

Table 6: Distribution of *thíra* and *pórta* in SMG

3.2 Case study II: λευκός *lefkós* vs. άσπρος *áspros*

The AG word λευκός (AG pronunc. *leykós*) has been inherited from the IE Proto-Language (IE *leuk- ‘white’, cf. Skt. *roca-* ‘bright’ (Beekes 2010 s.v.)). It is already attested in Mycenaean (re-u-ko). In MG, we also find the word άσπρος *áspros* with the same meaning. It is firstly attested in the 5th c. AD and originates from the Latin word *asper* ‘rough’, ‘harsh’, ‘uneven’, often referring to surfaces or skin.³² Both the continuation of the initial word *lefkós* and the competition of the lexical pair on the adstratal level of the written learned discourse are attested in the following quantitative diachronic data, extracted from the TLG corpus:

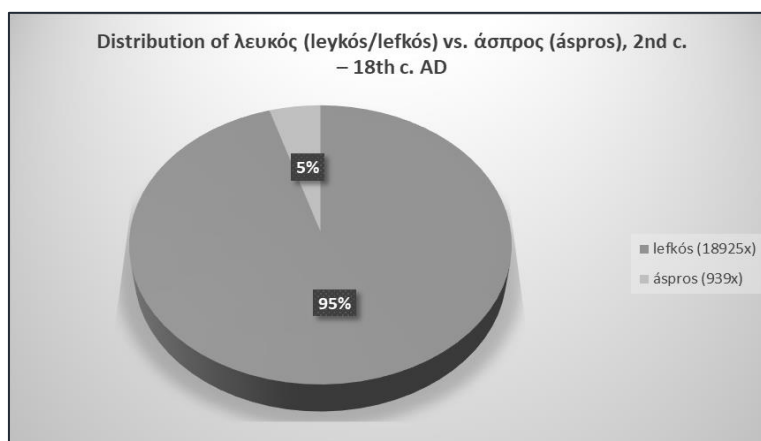


Table 7: Distribution of λευκός (*lefkós/lefkós*) vs. άσπρος (*áspros*), 2nd c. – 18th c. AD in TLG

However, as can be seen in Early Modern Greek, the difference in distribution between the two words decreases significantly after Medieval Greek, proving that the new word tends to gradually replace the older one, see Table 8:

³¹ Accessible at http://hnc.ilsp.gr/index.php?current_page=main&lang=en, last count on 24.11.20.

³² For *áspros* as a Greek loanword in Jewish Palestinian Aramaic see Katsikadeli/Klampfl/Slepoy in this volume.

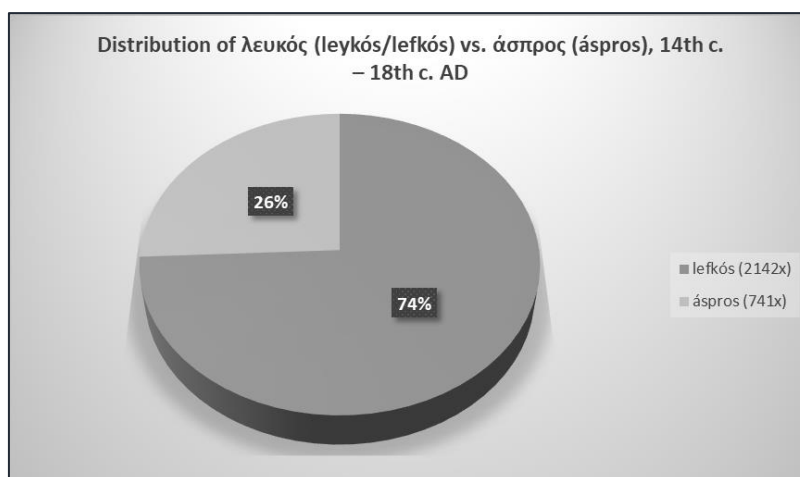


Table 8: Distribution of λευκός (*lefkós/lefkós*) vs. άσπρος (*áspros*), 14th c. – 18th c. AD in EMG

The patterns of distribution of the two lexemes may vary in the works of different authors, which use both words, as the tables below illustrate:

i) Malalas (5th – 6th c.)

ii) Anna Komnene (11th – 12th c.)

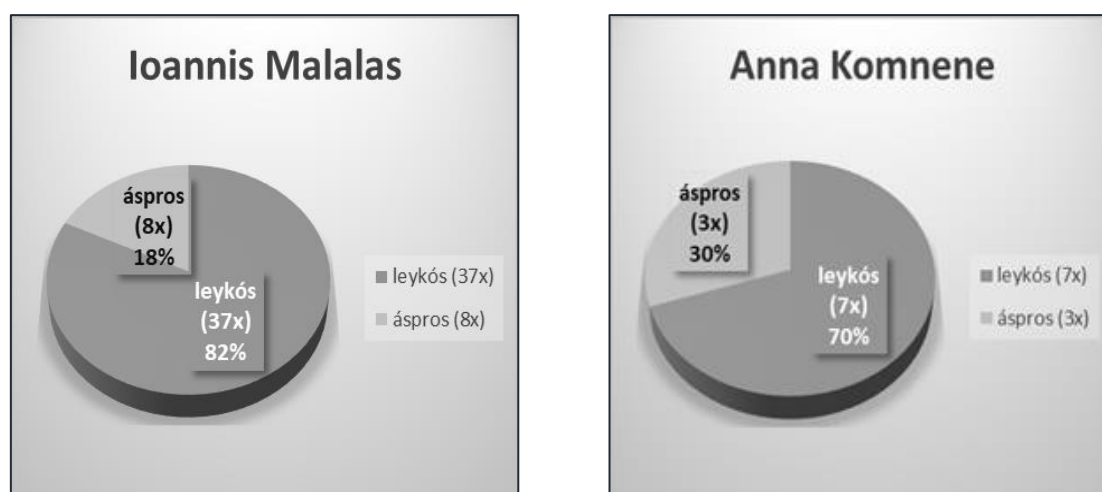


Table 9: Distribution of λευκός (*lefkós/lefkós*) vs. άσπρος (*áspros*) in various authors

The lexeme *lefkós* is attested in Malalas in order to describe human complexion and the appearance of several Ancient Greeks (Diomedes, Chryseida, Kalchas) or Romans (Dometianus, Helius Antonius Pius). Also, in collocations describing the air, animals, a pavement and in translating a toponym (thus, it is still the default option for ‘white’). In Anna Komnene, it is employed in descriptions for eyes, skin and a gem, in other authors of that period, it is used for describing milk, teeth, a hat, the skin, a larnax (small coffin or cinerary urn), a horse, gems, tunics and other garments. In SMG the word is frequent, although it belongs to the learned vocabulary and in several registers of the written language is the only one to be found.

The distribution is mostly complementary, e.g. *lefkós* → ‘white’ → ‘white’ (as an adstratal element) → MG ‘white’, but it is usually used in multi-word compounds or as a compound constituent (mostly calques), mainly meaning ‘neutral, clear, lucid’, e.g. Λευκός Οίκος *lefkós íkos* ‘White House’, λευκό ποινικό μητρώο *lefkó pinikó mitróo* ‘white criminal record’, λευκή σημαία *lefkí siméa* ‘white flag’. The lexeme *áspros* → ‘white’ is then undoubtedly more colloquial, and although is excluded in some

contexts, such as the compounds and collocations we mentioned above, it is the word that will be spontaneously uttered in common parlance. Nevertheless, within the MG corpora we encounter both lexemes, and *lefkós* is still preponderant in the written language.

Corpus	tokens of <i>lefkós</i>	tokens of <i>áspros</i>
Hellenic National Corpus ³³	5,827	1,659
Sketch Engine's Greek Web 2014 (elTenTen14) Corpus	151,231	43,930

Table 10: Distribution of *lefkós* and *áspros* SMG

3.3 Case study III: οίκος *íkos* vs. σπίτι *spíti*

The AG word οίκος (*oikos*, MG: *íkos*) also belongs to the inherited IE lexicon (*weyk- ‘to settle’, dial. φοῖκος *woikos*, cognate to Lat. *vīcus*) and it was already attested in Mycenaean (wo-i-ko-de ‘homeward’). The word σπίτι originates from the Latin *hospitium* ‘lodging’ (> also English *hospital*). It exhibits orthographic variation: ὀσπήτιον/ὀσπήτιον, ὀσπήτιον/ὀσπίτιον, σπίτιον (σπιτίον), σπήτι, σπήτι/σπίτι (SMG: σπίτι). Once more, according to the TLG corpus, it seems that the usage of the lexeme θύρα was persistent for a long period of time, as the diagram below, in Table 11, shows:

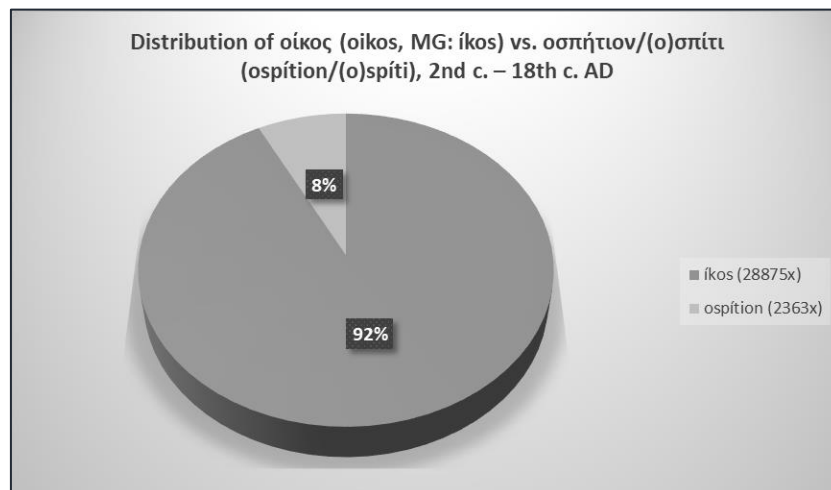


Table 11: Distribution of οίκος (*oikos/íkos*) vs. οσπήτιον/(ο)σπίτι (*ospítion/(o)spítì*), 2nd c. – 18th c. AD in TLG

It is again in Early Modern Greek that the new lexeme becomes more frequent similarly to the previous cases we already examined, see Table 12:

³³ Accessible at http://hnc.ilsp.gr/index.php?current_page=main&lang=en, last count on 24.11.20.

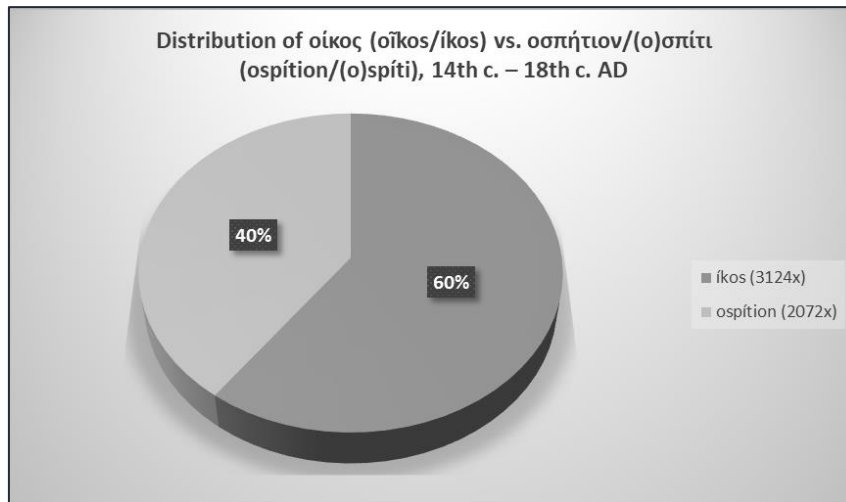
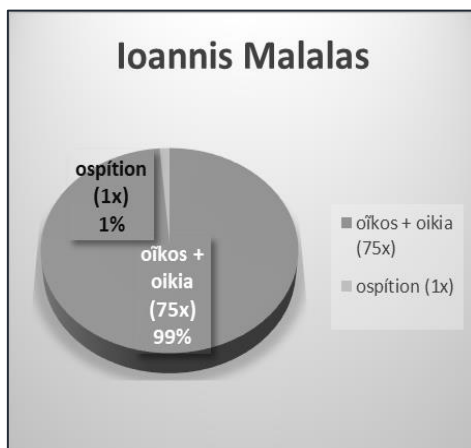


Table 12: Distribution of οίκος (*oikos/ikos*) vs. οσπήτιον/(ο)σπίτι (*ospition/(o)spiti*), 14th c. – 18th c. AD in EMG

Patterns of distribution of the two lexemes may vary in the works of different authors, which use both competitive words, as the tables below illustrate:

i) Malalas (5th – 6th c.)



ii) Anna Komnene (11th – 12th c.)

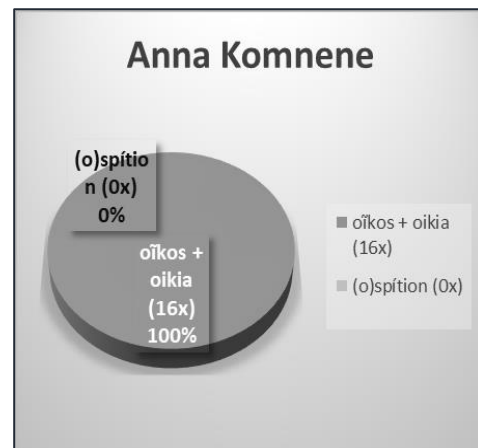


Table 13: Distribution of οίκος (*oikos/ikos*) vs. οσπήτιον/(ο)σπίτι (*ospition/(o)spiti*) in various authors

The first attestations can be traced in Thessalus of Tralles (*De virtutibus herbarum*: 1.10.5, 1st c. AD) referring to a doghouse (a) and Aelius Herodianus (*Partitiones*, 2nd c. AD, a sort of Dictionary) (b):

(4)

a. “ἐὰν δὲ ὑπὸ κυνὸς λυσσῶντος δηχθῆ τις ... κλύσον αὐτὸν εἰς ὀσπήτιον...

b. “τοῖχος: ὁ τοῦ ὀσπιτίου.” (137) “The wall, the one of a house”

In some early attestations, the word still reflected the meaning of the Latin loanword, namely “poor-house”:

(5)

τὸν δὲ τῆς Ἀφροδίτης ναὸν ἐποίησεν....κτίσας πέριξ ὁσπήτια καὶ κελεύσας δωρεὰν μένειν ἐν αὐτοῖς τὰς πᾶνυ πενιχρὰς πόρνας.

“And he founded an Aphrodite’s temple....constructing lodgings around it and ordering that the poorest prostitutes reside there”

(Malalas, *Chron.* 13,38,11 – 5th c. AD).

However, it seems that already in the 10th c. A.D. the “prototypical” meaning ‘house’ was generalized:

(6)

αὐλὴν περιοριζομένην μετὰ πόρτης ἔχουσαν καὶ ὁσπήτια δέκα

“yard marked by boundaries having a gate and ten buildings/houses”

(*Acta Monasterii Lembiotissae*, 12th -13th)

In SMG, the distribution is mostly supplementary (consider also Table 14 below). *ikos*: AG *oikos* ‘house’ → ‘house’ (as an adstratal/superstratal element) → in MG usually preserved in multi-word compounds or as a compound constituent (mostly calques), meaning mainly ‘lodge/building, usually big/gorgeous with specific function’, e.g. Οἶκος Θεοῦ *ikos theú* ‘house of God’, εκδοτικός οἶκος *ekdhotikós ikos* ‘publishing house’, οἶκος εὐγηρίας *ikos evghirías* ‘nursing home’, οικό-πεδο *ikópedho* ‘plot’ (introduced into the written language, probably not colloquial initially). Earlier colloquial compound: νοικοκύρης *nikokíris* “homemaker”, νοικοκυρά *nikokirá* ‘housewife’ (with initial /n/ from reanalysis of the preceding article τὸν οικοκύρη(v) → τὸ νοικοκύρη(v) *ton ikokíri(n)* → *to nikokíri(n)*). *spíti* → ‘house’ → ‘ordinary family house’.³⁴

Corpus	tokens of <i>ikos</i> + <i>ikia</i>	tokens of <i>spíti</i>
Hellenic National Corpus ³⁵	4,288 + 591	22,150
Sketch Engine’s Greek Web 2014 (elTenTen14) Corpus	119,730 + 29,779	624,481

Table 14: Distribution of *ikos* and *spíti* in SMG

4 Conclusions and open issues

On the basis of the three cases we examined in the previous section of this paper, we observe that the idea that lexical shift is a sort of replacement is inaccurate (at least in the cases we investigated) for two reasons: a) the “older” lexeme is not an obsolete or fossilized item, but is still in use bearing more specific/less prototypical semantics, and b) it is attested on the adstratal level and in specific registers, thus it is not “re-introduced” as an archaic superstratal word but modified through regular linguistic practices. Hence: i) *pórta* attained the prototypical meaning DOOR while *thíra* has been semantically specified, ii) the word *lefko* never ceased to abundantly appear in texts constituting a genuine synonym of *áspro*, despite its register differentiation and partially distinct distribution in collocations, iii) (*o*)*spíti*

³⁴ Compounds very often tend to preserve archaic vocabulary and/or meanings, e.g. Germanic: Engl. *werewolf*, German *Werwolf* (cf. *Wer-* ~ Lat. *vir* ‘man’, Vedic *vira* ‘man; hero’), German *Weichbild* ‘city landscape; later: urban area’ (cf. *Weich-* ~ Lat. *vicus*).

³⁵ Accessible at http://hnc.ilsp.gr/index.php?current_page=main&lang=en, last count on 24.11.20.

usage was rapidly diffused in texts (reflecting developments in the vernacular language) while *ikos* or *ikia* became a word of the high register or got semantically more specified.

Along with the cases we examined, we regard the theoretical framework we provided as a meaningful contribution to the overall semantic and onomasiological research. The concept of *prototypicality* illustrates tendencies in meaning and lexical change that can be tested in other languages that exhibit phenomena that are germane to the Greek ones. In a same vein, the concept of *exaptation* accounts for the integral presence of the archaic/learned component either as a marker of higher register or, most importantly, as a case of lexical complementary distribution, according to which otherwise synonymous words of a lexical pair have discrete semantic or stylistic functions.

Furthermore, corpus analysis reveals that the impressionistic idea that lexical change has been diffused from vernacular to higher registers is not always valid: e.g. *pórta* was probably introduced into Greek not due to language contact but as part of the official vocabulary that reflected a certain ideology. Due to a long history of systemic diglossia, one part of the adstratal vocabulary in Greek is considered “archaic” or “learned” and is thus used artificially. Although this may be true in cases of superstratal/reintroduced words, in the pairs we examined, we see that there has been a continuity of usage and a semantic distinctiveness. Thus, the adstratal vocabulary is an indispensable component of the overall Greek lexicon.

Given that our research has been only recently begun and is “work in progress”, there are still several open issues:

a) A typology of lexemes pairs based on semantic characteristics is needed. It is anticipated then that the meticulous examination of more pairs can reveal patterns of lexical shift.

b) The examination of semantic gradience in semantic change similarly is a desideratum. This constitutes a major challenge, since we are called to trace subtle diachronic semantic shifts. Nevertheless, this task is utterly required for locating a replacement of one lexeme by another in the vernacular in the history of the language.

c.) Geographical distribution and diffusion of changes. The diachronic investigation through corpora (and crucially also through dialectal material) will reveal the gradual change of meaning of the lexemes. In turn, a more detailed examination of the texts will also lead to a more sophisticated theoretical framework of semantic and onomasiological change.

Finally, it should be emphasized that the ability to conduct research on lexical, onomasiological and semantic change is only feasible by making use of the digital means we have at our disposal today. Large-scale lexical investigations have been either exceedingly strenuous or merely impressionistic in the past, and this is the reason why we still lack a sufficient methodology for explaining cases of MG-like lexical variety. The application of digital tools does not only allow to posit new questions but also to deal with “more traditional” philological and linguistic issues afresh.

References:

- Adrados, Francisco Rodriguez. 2005. *A history of the Greek language: from its origins to the present*. Leiden, Boston: Brill.
- Alexiou, Margaret. 2002. *After Antiquity: Greek Language, Myth and Metaphor*. Ithaca and London: Cornell University Press.
- Anastassiadis - Symeonidis, Anna and Asimakis Fliatouras. 2019. “The learned level in Modern Greek: synchronic and diachronic tendencies”. In Asim. Fliatouras and Anna Anastasiadi Simeonidi (eds.) *The learned level in Modern Greek: Theory, history, application*. Athens: Patakis (in Greek).

- Bakker, Egbert J. (ed.). 2010. *A companion to the Ancient Greek language*. Oxford: Wiley-Blackwell.
- Beekes, Robert S.P. 2011. *Comparative Indo-European Linguistics. An Introduction*. (Revised and corrected by Michiel de Vaan). Amsterdam/ Philadelphia: John Benjamins Publishing Company.
- Biggam, C. P. 2012. *The Semantics of Colour: A Historical Approach*. Cambridge: Cambridge University Press.
- Brunner, Theodore F. (1991). "The Thesaurus Linguae Graecae: Classics and the Computer". *Library Hi Tech* 9, no. 1: 61–67.
- Brunstad, Endre. 2003. "Standard language and linguistic purism". *Sociolinguistica* 17 (1), pp. 52 – 70.
- Dawkins, Richard M. 1916. *Modern Greek in Asia Minor: A study of the dialects of Silli, Cappadocia and Pharsa with grammar, texts, translations and glossary*. Cambridge: Cambridge University Press.
- Dressler, Wolfgang U. 1987. *Leitmotifs in natural morphology*. Amsterdam: Benjamin.
- Eideneier, Hans. 1966. *Sogenannte christliche Tabuwörter im Griechischen*. München: Institut für Byzantinistik und Neugriechische Philologie der Universität (Miscellanea Byzantina Monacensia, 5), zugleich Dissertation München.
- Ferguson, Charles. 1959. "Diglossia". *Word* 15: 325 – 340.
- Fliatouras, Asimakis & Konstantinos Sampanis (forthc.). "Towards the necessity to revise the notion and the lexicographic use of the term 'learned' in Standard Modern Greek" (in Greek). To appear in Anastasiadi & Katsoyannou (eds.) collective volume *Lexicography and Morphology*. Athens: Patakis.
- Fritz, Gerd. 2006. *Historische Semantik*. Stuttgart: Springer-Verlag.
- Geeraerts, Dirk. 2010. *Theories of Lexical Semantics*. Oxford: Oxford University Press.
- Glare, P. G. W. 2012. *The Oxford Latin Dictionary*. Oxford: Oxford University Press.
- Hock, Hans Henrich & Brian D. Joseph. 2009. *Language History, Language Change and Language Relationship. An Introduction to Historical and Comparative Linguistics*. Berlin/ New York: Mouton De Gruyter.
- Jeffreys, Elisabeth; Michael Jeffreys and Roger Scott. 1986. *The Chronicle of John Malalas. A translation*. Byzantina Australiensia 4. Melbourne: Australian Association for Byzantine Studies.
- Kakoulidi-Panou, E., E. Karantzola & K. Tiktopoulou (in press). *Demotic prose texts of the 16th c.*, [in Greek]. Thessaloniki: Centre for Greek Language.
- Kaldellis, Anthony. 2019. *Romanland. Ethnicity and Empire in Byzantium*. Cambridge, Massachusetts: Harvard University Press.
- Kim, Lawrence. 2010. "The Literary Heritage as Language: Atticism and the Second Sophistic". In Bakker (ed.), v.s., pp. 468 – 482.
- Lakoff, George. 1972. "Hedges: a study in meaning criteria and the logic of fuzzy concepts". *Chicago Linguistic Society* 8: 183–228.
- Lakoff, George. 1987. *Women, Fire and Dangerous Things*. Chicago: University of Chicago Press.
- Langer, Nils and Agnete Nesse. 2012. "Linguistic Purism". In Hernández-Campoy, Juan Manuel & Juan Camilo Conde-Silvestre. (eds.). *The Handbook of Historical Sociolinguistics*. Blackwell, pp. 607-625.
- Martzoukou, Maria, Stathis Selimis, Athanasia Katsalirou, Asimakis Fliatouras (2019). "The learned/formal register of Modern Greek as a basis for learning Ancient Greek: Experimental data and educational considerations.". In Gavriilidou, Zoe, Lydia Mitits, Chrisa Dourou (eds), *Situating Strategy Use in the Greek Setting*. Kavala: Saita.
- Palmer, Frank Robert. 1981. *Semantics*. Cambridge: Cambridge University Press.
- Panagl, Oswald. 1989. „Bahuvrihi, Ellipse, Synekdoche“. In Heller, Karin, Panagl, Oswald & Johann Tischler. *Indogermanica Europaea*. Festschrift für Wolfgang Meid zum 60. Geburtstag am 12. 11. 1989, GLM 4. pp. 185-192.
- Pantelia, Maria. 2000. 'Noûs, INTO CHAOS': The creation of the Thesaurus of the Greek Language. *International Journal of Lexicography*, Volume 13, Issue 1, Pages 1–11, <https://doi.org/10.1093/ijl/13.1.1>.
- Petrounias, E. 2017. "The lexicon of Modern Greek: Origin of categories". In Tzitzilis Christos, Papanastasiou Georgios (eds), *Greek Etymology*. Thessaloniki: Institute of Modern Greek Studies, 240-259.
- Richardson, L. (ed.). 1992. *A New Topographical Dictionary of Ancient Rome*. Oxford: Oxford University Press.
- Rosch, Eleanor H. 1973. "Natural categories". *Cognitive Psychology*. 4 (3): 328–350.
- Saeed, John I. 2009. *Semantics*. Malden, MA: Wiley-Blackwell.
- Taylor, John R. 1995. *Linguistic Categorization: Prototypes in Linguistic Theory*, 2nd ed. Oxford: Oxford University Press.
- Thomas, George. 1991. *Linguistic purism*. London & New York: Longman.
- Toufexis, Notis. 2008. "Diglossia and Register Variation in Medieval Greek". *Byzantine and Modern Greek Studies* Vol. 32 No. 2, 203–217. London: Maney Publishing.
- Traugott, Elizabeth Closs. 2010. "Grammaticalization". In Luraghi, Silvia and Vit Bubenik (eds.): *Continuum Companion to Historical Linguistics*. London: Continuum.