



## FORMULAICITY WITHIN TURKISH WORDS<sup>1</sup>

Philip Durrant<sup>2</sup>

*University of Exeter*

**Abstract:** One of the main insights to emerge from the last fifty years of corpus linguistics has been a greater understanding of the pervasiveness of formulaic language. Rather than exercising the full generative capacity of language, speakers and writers have been shown to rely to a great extent on conventional, pre-constructed phrases drawn from memory. Turkish presents a particularly interesting and challenging case because its agglutinative structure means that messages which are spread across several orthographic words in English are often expressed within a single word in Turkish. While it is possible that this difference in structure will mean that new types of formulaicity will emerge in Turkish, a good starting place may be to consider the extent to which types of formulaicity which are known to exist in English at the multi-word level exist in Turkish at the sub-word level. The research discussed here set out to examine this possibility, looking in particular at three types of formulaicity: collocations, lexical bundles and collostructions.

**Keywords:** *Formulaicity, collocations, collostructions, lexical bundels*

1 This study was supported by TÜBİTAK (Grant no: 113K039).

2 University of Exeter, Graduate School of Education, Exeter, United Kingdom, P.L.Durrant@exeter.ac.uk

Makale gönderim tarihi: 12 Eylül 2015; Kabul tarihi: 30 Mart 2016

## TÜRKÇE SÖZCÜKLERDE KALIPLAŞMA

**Öz:** Derlem dilbilim çalışmalarının son elli yılda ortaya koyduğu önemli bir gözlem, kalıp anlatımların dildeki yaygınlığıdır. Dil kullanıcıları, dilin üretici olanaklarını kullanmak yerine büyük ölçüde belleklerinde saklı duran, geleneksel, önceden kurulmuş öbekleri kullanma eğilimindedirler. Türkçe bu açıdan ilginç zorluklar taşımaktadır. Sondan eklemeli yapısının bir yansıması olarak İngilizcenin birden çok sözcüğe yaydığı bir anlatımı tek bir sözcükbirimde toplamaktadır. Türkçede kalıp anlatıları betimlemek için yeni birimler düşünmek sözkonusu olsa da, zengin ek yapısına bakarak İngilizce için sözcüksel olan çoksözcüklü birimlerle yapılan anlatımın Türkçe için sözcük-altı birimlerle yapıldığını söyleyebiliriz. Bu çalışma üç tür kalıplaşmaya bakarak bu olasılığı inceleyecektir: eşdizimlilik, sözcük kümeleri ve eşyapı.

**Anahtar sözcükler:** *Kalıplaşma, eşdizimlilik, eşyapı, sözcük kümeleri*

### 1. INTRODUCTION

#### 1.1. CONCEPTUALIZING FORMULAIC LANGUAGE

A problem facing anyone researching formulaic language is that of deciding what the term ‘formulaic language’ should refer to. As Wray (2002) pointed out in her landmark review of the area, linguistic phenomena which might loosely be described as ‘formulaic’ have been studied by researchers in a wide range of fields and for a wide range of purposes. This has led to a proliferation of terminology and of perspectives, with different researchers defining their objects of study in different ways. In an attempt to be inclusive, Wray formulated a now widely-cited definition which aims to capture the common ground between the different approaches to formulaicity, coining the term formulaic sequence to refer to:

“a sequence, continuous or discontinuous, of words or other elements, which is, or appears to be, prefabricated: that is, stored and retrieved whole from memory at the time of use, rather than being subject to generation or analysis by the language grammar (Wray, 2002, p. 9)”.

However, while it has been influential, this definition is not as inclusive as Wray had intended since it entails a very specific psycholinguistic model (i.e., that formulas must be ‘stored and retrieved whole from memory at the time of use’) which many researchers would question (Siyanova-Chanturia & Martinez, 2014 provide an excellent review of the issues).

Any definition which aims at inclusivity needs to leave room for research which prefers one of the many alternatives to Wray’s ‘holistic recall’ model of formula processing. It also needs to leave room for research which prefers to remain agnostic about the psycholinguistic correlates of formulas. Much research into formulaic language is interested less in the psycholinguistic status of formulas than in what they tell us about grammar, lexicography, discourse, or language pedagogy (Durrant & Mathews-Aydınlı, 2011). In work of this kind, linguistic sequences would be of interest regardless of their psycholinguistic status.

An overarching definition of formulaicity needs, therefore, to recognize that psycholinguistic status is only one reason amongst others why we might want to treat a linguistic sequence as a whole. For this reason, I advocate a more open definition of formulaic language as “sequences, continuous or discontinuous, of linguistic elements which, for one reason or another, can usefully be treated as a whole, rather than being analyzed into their component parts”.

On this model, formulaic language is not seen as a delimitable set of items. There is no theoretical limit to the reasons why we might choose to treat sequences of language holistically and so there can be no definitive list of formulas. Formulaicity can be seen, rather, as an approach to language study which recognizes that it is not always appropriate or useful to analyse sequences into minimal component parts; a recognition of the value of holism.

## 1.2. FORMULAICITY IN TURKISH

Working within this broad definition, why should formulaicity be of interest to scholars of Turkish? On the one hand, it is interesting for the same reasons that it has been of interest to researchers working on other

languages. Formulaic approaches can provide psycholinguistically - plausible and pedagogically-useful models of features of language which cannot be satisfactorily dealt with analytically. The recent history of research into formulaicity in English has provided significant insights into the nature of language, language processing and language learning and is now used extensively in countless dictionaries, grammars, and language teaching materials (the 2012 special issue of the *Annual Review of Applied Linguistics* (Polio, 2012) provides an excellent overview). It is to be hoped that formulaic approaches to the study of Turkish will yield similar benefits for students and scholars of that language.

From a broader perspective, work on formulaicity in Turkish has the potential to make a substantial contribution to debate on the nature of formulaicity in general. Its agglutinative structure opens up the possibility of types of formulaicity different from those found in morphologically-poor language such as English. Any general theory of formulaicity as a feature of language (rather than of a few well-studied languages) will need to take account of, and explain, what happens in such languages. Work on Turkish can therefore make a crucial contribution to our knowledge of formulaicity as a general feature of human languages.

It was this latter consideration which motivated my own exploratory research into formulaicity within inflected Turkish verbs (Durrant, 2013). In that study, I considered the extent to which three formulaic phenomena which have been productively studied in English could be found at the sub-word level in Turkish verbal inflections. The three phenomena I looked at were syntagmatic associations between linguistic items (as seen in English in collocations between words), fixed extended sequences of items (seen in English in lexical bundles) and associations between particular lexical and grammatical forms (see in English in collocations).

The study utilized a corpus of newspaper texts, collected over a period of six months. In contrast to many corpora, this collection was not intended to be representative of a particular realm of discourse, but rather to represent the newspaper text to which a typical reader might be exposed. The rationale for this choice was that the frequency features of the range of language with which any individual interacts is likely to be

different in type from those of broader realms of discourse (Durrant & Doherty, 2010). As I was interested primarily in formulaicity as a psycholinguistic phenomenon – a property of the language systems of individual speakers – a corpus of this type was therefore more relevant to my purposes. The corpus covered all of my own reading of online Turkish-language newspapers over a six-month period. It comprised a total of 374,590 words, from 765 separate articles and opinion pieces published in seven different newspapers.

The analysis focused on the inflected forms of 20 different verbs with widely varying frequencies of occurrence (see Table 1). All occurrences of these verbs were retrieved from the corpus and their inflectional suffixes manually tagged. The outcome of the analysis was a spreadsheet listing each form, along with its frequency, and separate columns representing each suffix (illustrated in Table 2). This enabled an analysis of the frequency of particular verb forms, of inflection combinations and of the relationships between inflections and verb roots.

**Table 1.** Verb stems studied

Verb root	Translation	Cumulative stem frequency <sup>3</sup>	Total types	% total tokens covered by top 5% of types	% types appearing once only
ol	be	8,540	438	72.06	39.27
et	do/make	4,161	423	56.50	42.08
yap	do/make	3,189	355	57.54	42.54
ver	give	1,836	256	49.35	39.45
de	say	1,232	108	60.71	44.44
çık	go/come out; emerge	1,112	145	53.15	42.76
çalış	work	964	167	38.90	43.11
konuş	speak	790	157	53.29	50.96
geç	pass	768	188	43.88	54.26
yaşa	live/experience	736	156	45.92	51.92
gir	enter/go into	474	133	38.19	55.64
bak	look	381	111	37.27	55.86
bırak	leave	341	114	31.67	56.14

<sup>3</sup> In this paper, the term ‘cumulative stem frequency’ is used to refer to the combined frequency of all inflected forms of a verb.

Verb root	Translation	Cumulative stem frequency <sup>3</sup>	Total types	% total tokens covered by top 5% of types	% types appearing once only
anlat	explain	313	80	48.56	50.00
geliş	develop	312	70	37.82	51.43
sağla	provide/obtain	271	97	32.47	56.70
yarat	create	207	73	37.20	60.27
koru	protect	190	64	43.68	64.06
paylaş	share	76	41	18.42	68.29
önle	prevent	42	22	21.43	68.18

**Table 2.** Sample analysis

Word	Freq	Root	Suffix 1	Suffix 2	Suffix 3
					<i>POS.3</i>
anlatmaması	1	anlat	<i>NEG-mA</i>	<i>SUB-mA</i>	<i>&lt;s&gt;I&lt;n&gt;</i>
anlatmıyor	1	anlat	<i>NEG-mA</i>	<i>IMP-&lt;I&gt;yor</i>	
anlatmadım	1	anlat	<i>NEG-mA</i>	<i>PRF-DI</i>	<i>1-m</i>
anlatın	3	anlat	<i>2PL-&lt;y&gt;In</i>		
anlatsın	1	anlat	<i>3-sIn</i>		

Analysis of these data revealed a number of key findings:

1) The frequencies of individual verb forms were highly skewed, such that a small number of very frequent forms made up a high percentage of each verb's occurrences (see Table 1). This was taken to suggest that a cognitively-efficient language system would require some kind of formulaic storage or processing of particular forms.

2) Strong collocational relationships were found between suffixes. To take one example, 19.8% of occurrences of the suffix *NEG-mA* were directly followed by *SUB-DIK*; a further 17.3% were followed by *AOR-z* and 10.13% by *SUB-<y>An*. Thus, over 47% of occurrences were followed by one of only three other suffixes. Looking to the other side of *NED-mA*, 18.6% of occurrences were directly preceded by *POSS-<y>A*; a further 6.7% were preceded by *PASS-il* and 0.9% by *PASS>I>n*. Thus, over 26% of cases were preceded by one of only three other suffixes. Generalizing these calculations across the 29 most widely-used suffixes, it was found that, on average, 40% of cases of

each suffix were directly followed, and 38% directly preceded, by one of three other suffixes.

3) Particular combinations of suffixes were also found to occur with very high frequency. This is exemplified in Table 3, which shows the ten most frequent three-morpheme bundles used with these verbs. A number of points can be noted about these bundles. First, they all appear with very high frequency. If the verbs sampled for this study are typical of those in the rest of the corpus, the most frequent bundle (SUB-DIK POS.3-<s>I<n> ACC-<y>I) is used in almost one in twenty verb tokens, while the top ten bundles together are used in around one in eight verbs. This suggests that these bundles are highly likely to be formulaic for newspaper readers and writers and that these would be an excellent focus for learners of the language.

Second, the majority of these bundles are used with a wide range of verb roots – all but two being found with at least three-quarters of the twenty verbs studied. This lends further credibility to the idea that an efficient language system would include some kind of formulaic storage of these items.

Third, it is notable that one structural type dominates the list of frequent bundles. Specifically, nine of the ten bundles involve combinations of subordinators plus person markers. This points to an interesting cross-linguistic regularity. It is known that English lexical bundles often consist of parts of embedded clauses, such as I don't know why or I thought that (Biber et al., 1999, p. 991). Both English word bundles and Turkish morpheme bundles appear therefore to be primarily used for the structural job of anchoring complex sentences. This lends support to Pawley and Syder's (2000) claim that one aim of formulaic language is to enable speakers to fluently process language which spans clauses.

4) Though morpheme bundles were used across a range of verb forms, it was also found that particular bundles were biased towards (or away from) particular roots. This is analogous to the relationship of collustruction described in other languages, whereby relationships of attraction or repulsion are seen to exist between particular grammatical forms and particular lexical items (Stefanowitsch & Gries, 2003). Table

4 shows the verb roots which are attracted/repelled by the ten high-frequency morpheme bundles. Two key of points can be made about this. First, all bundles except one showed patterns of strong attraction/repulsion towards particular roots, suggesting that lexis and syntax are to a certain extent co-selected. Second, associations appear to hold not only between particular root and particular morpheme bundles but between roots and more abstract grammatical categories. For example, all active-voice bundles including the two-morpheme bundle SUB-DIK POS.3-<s>I<n> are attracted to the root ol ('be'), but repelled by yap ('do'/'make') and et ('do'/'make'), while passive-voice bundles are all attracted to the roots yap and et.

Table 3. 10 most common three-morpheme bundles

Row No.	Bundle	Examples	Types	Roots	% verb form tokens
11.1	SUB-DIK POS.3-<s>I<pl> ACC-<y>I	"Çocuklar-a okul yetiştir-e-mi-yor-uz" "child-PL-DA schoo keep up-POSS-NEG-PROG-1 PL" I claimed that he said that "We can't build enough schools for the children"	59	17	4.57
11.2	PASS-III SUB-mA POS.3-<s>I<pl>	Çünkü <b>brak-ı-l-ma-sı</b> Because <b>release-PASS-SUB-POS</b> Because it is requested that he be released	67	16	1.50
11.3	NEG-mA SUB-DIK POS.3-<s>I<pl>	olumlu bir <b>me-dig-i</b> positive a messag <b>give-NEG-SUB-PO</b> it certainly didn't give a positive message	51	16	1.48
11.4	PASS-II	darbe hazırlık-ları-m konu-yl-dig-u toplantı-da	46	16	0.97



SUB-DIK POS.3-<s>I<nl>	coup N	preparation-POS.3PL-GE N	discuss-PASS-SUB-PO S.3	meeting-LOC meeting-LOC	16	53	16	0.86
11.5 SUB-<y>AcAK POS.3-<s>I<nl> ACC-<y>I	görişm e	stürec-in-de e	bir k	tükanklı k	ol-acag- <i>mi</i> -I	san-mı-yor-um	53	0.86
	meeting C	process-POS.3-LO C	a C	hold.up C	be-SUB-POS.3-AC C	think-NEG-PROG-1		
	I don't think there will be a hold up in the meeting process							
11.6 SUB-mA POS.3-<s>I<nl> ACC-<y>I	Hem de Also	parlamento parliament	çalış-ma-sın-ı work-SUB-POS.3-ACC	bil-iyor-um know-PROG-1		38	18	0.82
11.7 SUB-mA POS.3-<s>I<nl> DAT-<y>A	olay-lar-in incident-PL-GE N	çık-ma-sın-a occur-SUB-POS.3-DAT	engel obstructio n	ol-ma-ya be-SUB-DAT	çalış-tı try-PRE.3	27	16	0.61
	tried to stop incidents occurring							
11.8 POSS-<y>A NEG-mA AOR-z	Öğrenci Student	sınıf-a class-DA	şapka ile hat-CYB	git-e-me-z enter-POSS-NEG-AOR.3		39	13	0.58
	A student cannot go to class wearing a hat							

11.9	SUB-DIK POS.3PL-1Arl< n> ACC-<y>I	bir p mektn k yolla-yara k send-CVB letter	yaşa-dık-ların-ı experience-SUB-POS.3PL-A CC	biz-let-le us-PL-OB L	paylaş-m ak share-INF	iste- miş- ler want -PR F-3P L	22	15	0.51
11.1 0	SUB-DIK POS.3-<s>I<n> DAT-<y>A	Başkan-ın Chair:perso-GEN According to the chair person	anlat-tu-ın-a explain-SUB-POS.3-DA T	göre according to			18	12	0.48
		meeting	hold up	a	be-SUB-POS.3-AC C	think-NEG-PROG-I C			

**Table 4.** Associations and repulsions between three-morpheme bundles and verb roots

Row No.	Bundle	Attracted			Repelled			Fisher's test (p)
		Root	Translation	Fisher's test (p)	Root	Translation	Fisher's test (p)	
15.1	SUB-DIK POS.3-<s>I<n> ACC-<y>I	ol	be	3.88E-142	yap	do/make	1.04E-13	
					et	do/make	1.30E-10	
					ver	give	2.55E-9	
					konuş	speak	8.27E-8	
					de	say	1.73E-6	
					çık	go/come out; emerge	9.32E-6	
					anlat	explain	1.43E-5	
					geç	pass	1.44E-4	
					çalış	work	2.07E-4	
					bırak	leave	9.25E-4	
15.2	PASS-II SUB-mA POS.3-<s>I<n>	et	do/make	2.06E-24	çık	go/come out; emerge	1.57E-6	
		yap	do/make	4.57E-15	çalış	work	1.8e-4	
		bırak	leave	2.59E-9				
		yarat	create	6.34E-7				
		ver	give	4.37E-4				
15.3	NEG-mA SUB-DIK POS.3-<s>I<n>	ol	be	2.11E-62	yap	do/make	4.67E-8	

15.4	PASS-II SUB-DIK POS.3-<s>I<n>	yap et	do/make do/make	9.89E-23 3.27E-8	çık çalış	do/make speak work	4.46E-5 1.40E-4 5.67E-4
15.5	SUB-<y>AcAK ACC-<y>I	sağla	provide/ obtain	5.77E-4	çalış	work	3.77E-3
15.6	SUB-mA POS.3-<s>I<n> ACC-<y>I	konuş brak	speak leave	9.08E-6 7.43E-3	ol	be	5.27E-3
15.7	SUB-mA POS.3-<s>I<n> DAT-<y>A	konuş	speak	3.32E-4			
15.8	POSS-<y>A NEG-mA AOR-z	et	do/make	1.67E-5			
15.9	SUB-DIK POS.3PL-IArI<n> ACC-<y>I						
15.10	SUB-DIK POS.3-<s>I<n> DAT-<y>A	ol	be	2.06E-23	et	do/make	1.13E-5
		yap	do/make		yap	do/make	8.56E-3

### 1.3. FUTURE PROSPECTS FOR RESEARCHING FORMULAICITY IN TURKISH

The research described in Durrant (2013) was intended to be exploratory and, as such, raised rather more questions than it answered. A usage-based model of language (Ellis, 2003; Kemmer & Barlow, 2000) would suggest that the various types of frequency-based phenomena discussed above – the skew towards particular word forms; the existence of collocational relations between suffixes and of extended morphological bundles; the preferences of particular morpheme bundles for particular verbal roots – are likely to be reflected in language users' mental linguistic representations and processing. Previous studies of morpheme-processing in agglutinating languages (e.g. Niemi et al., 1994) have proposed a 'dual route' processing model, whereby words may be either processed morpheme-by-morpheme or stored as single holistic chunks. However, the patterns seen in Turkish suggest the existence of intermediate levels of representation – larger than morphemes, but smaller than words – and of associations between those 'morphemic chunks' and specific lexical roots, which cannot be readily accounted for on such models. These point towards ways in which models of processing in agglutinating language could be enriched.

However, it is crucial to note that this possibility requires independent verification in the form of more directly psycholinguistic studies. While corpus data of the sort described above can give clues as to the types of psycholinguistic mechanisms which may be in place, and can draw attention to patterns of language use for which psycholinguistic models may need to account, the precise nature of those models needs to be spelled out, and their existence confirmed, through well-designed studies of language processing in action (Durrant & Siyanova-Chanturia, 2015). This should be a key focus of future research in this area.

While the primary focus of my previous study was on formulaicity as a psycholinguistic construct, another area in which Turkish formulaicity could be productively researched is in the study of discourse variation. Formulaic language has become a key focus in studies of variation for at least three reasons: formulaic combinations are highly sensitive to contextual variation; they often have distinctive semantic functions;

and they can be identified by automatic means across large numbers of texts. This means that analysis of formulas in a corpus can give an excellent insight into both formal and functional variation in language use (Durrant, 2015).

In Durrant (2015), for example, I used the technique of quantifying overlaps between writers in their use of four-word sequences to map the relations between a large number of university-level writers from a range of disciplines (see Figure 1). This analysis showed a clear pattern of difference between arts/social science disciplines on the one hand and science/technology disciplines on the other. Applied disciplines related to commerce (e.g. Business Studies, Agriculture) and health (e.g. Medicine, Psychology) were found to fall midway between these poles. This quantitative analysis provided the basis for a further, qualitative analysis, focusing on the nature of the recurrent sequences which were distinctive of the two main poles of the corpus (see Figure 2) to give an insight into the nature of the differences found in the initial map.

With the advent of reliable morpheme-level tagging for Turkish language corpora, enabling texts to be broken down into strings of component morphemes, research of this sort might be productively applied using the types of morphemic bundles described in the previous section to quantify and characterize discourse variation in Turkish texts. Following the methodology of Durrant (2015), for example, texts could be broken down into series of overlapping morpheme n-grams. For example, using 4-grams, the sequence (from my newspaper corpus, described above) *GDO yönetmeliğinde yapılan değişikliği değerlendirirken* could become (informally <sup>4</sup>):

<b>gdo</b>	<b>yönet</b>	<b>me</b>	<b>lik</b>
<b>yönet</b>	me	lik	in

<sup>4</sup> I have transcribed morphemes orthographically here. A full analysis of this sort would probably represent them using form/function notations of the sort used in earlier sections in order to overcome problems of ambiguity. An interesting question which future research should address is that of the optimal level of representation for particular research purposes. For example, for some purposes it may be appropriate to distinguish between formally different realisations of a morpheme (e.g. to distinguish between the third-person possessive endings *sı*, *si*, *ı* and *i*) while for other purposes, these might be combined (the approach I took in Durrant 2013).

<b>me</b>	lik	in	de
<b>lik</b>	in	de	yap
<b>in</b>	de	yap	ıl
<b>de</b>	yap	ıl	an

As in Durrant (2015), such series of n-grams could be created for each text (or collection of texts) and percentage overlaps in the use of n-grams defined to quantify similarities between texts. Follow-up analysis could then identify the n-grams which are distinctive of particular groups. These could be analysed qualitatively to understand the patterns of similarities and differences between groups of texts. Work of this sort would enable both linguistically-related clusters of texts to be identified in a bottom-up way (rather than specified in advance by the analyst) and the formal/functional features which characterise this variation to be determined.

## 2. CONCLUSION

Though the area of formulaicity in Turkish goes back some time (e.g. Doğancay, 1990; Tannen & Öztekin, 1981) the possibilities for exploring formulaic patterns through corpus methods are only starting to be explored (e.g. Doğruöz & Backus, 2009; Oflazer, Çetinoğlu and Say, 2004). The development of new technologies for morphological-level tagging and resources such as the Turkish National Corpus (Aksan et al., 2012) make it an exciting time to be working in this area and rapid developments can be hoped for in the years to come.

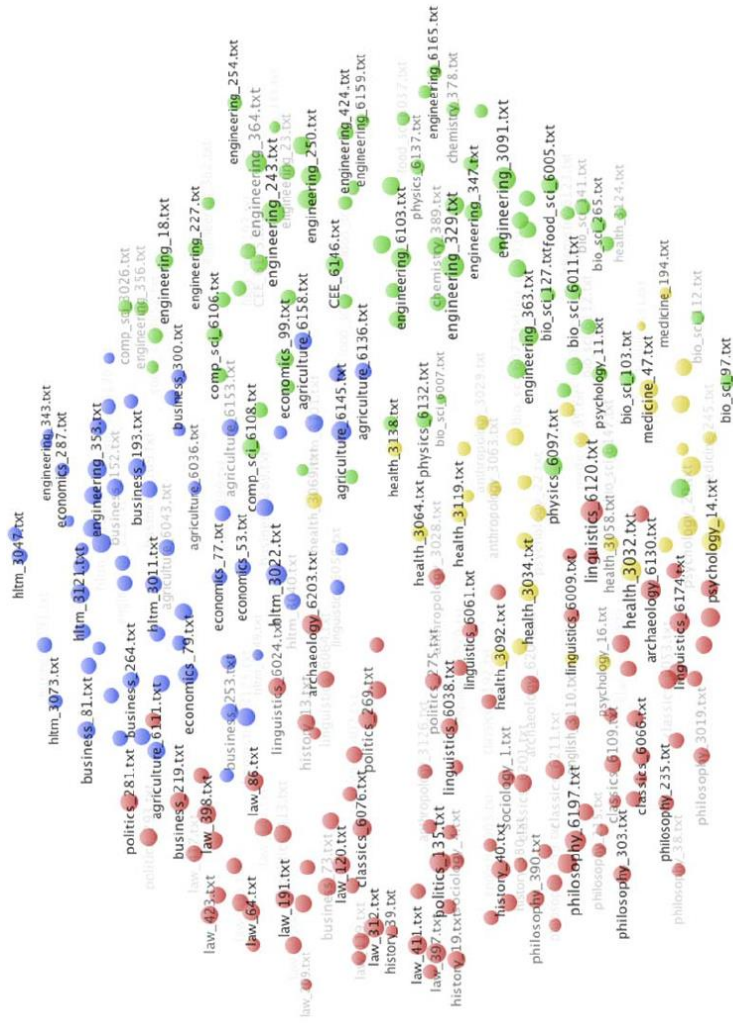


Figure 1. VOS map showing variation in 4-gram use across academic disciplines (Durrant, 2015)



Arts & Humanities/Social Sciences	Science/Technology
<ul style="list-style-type: none"> <li>• A focus on abstract constructs</li> <li>• A focus on historical moments/points in a process</li> <li>• Emphasizing the role of unique autonomous agents in processes that are difficult to control</li> <li>• Showing multiple contingent viewpoints</li> <li>• Evaluation</li> <li>• Establishing centrality</li> <li>• Setting things in interpretive/limiting context</li> <li>• Setting ideas in relationship with each other</li> </ul>	<ul style="list-style-type: none"> <li>• A focus on the physical world</li> <li>• Emphasizing the role of passive, interchangeable, instruments in processes that are tightly controlled by the researcher</li> <li>• Quantification; data presented in figures and tables</li> <li>• Received knowledge</li> <li>• Cause and effect</li> </ul>

**Figure 2.** Summary characterisation of distinctive bundles in Arts and Humanities/Social Sciences vs. Science/Technology

#### REFERENCES

- Aksan, Y., Aksan, M., Koltuksuz, A., Sezer, T., Mersinli, Ü., Demirhan, U. U., Yılmaz, H., Kurtoğlu, Ö., Atasoy, G., Öz, S., & Yıldız, İ. (2012). Construction of the Turkish National Corpus (TNC). In N. Calzolari, K. Choukri, T. Declerck et al. (Eds.), *Proceedings of the 12<sup>th</sup> International Conference on Language Resources and Evaluation (LREC)* (pp. 3223-3227). İstanbul, Turkey: LREC 2012.
- Doğançay, S. (1990). Your eye is sparkling: formulaic expressions and routines in Turkish. *Penn Working Papers in Educational Linguistics*, 6 (2), 49–65.
- Doğruöz, A. S., & Backus, A. (2009). Innovative constructions in Dutch Turkish: An assessment of ongoing contact-induced change. *Bilingualism: Language and Cognition*, 12 (1), 41–63.
- Durrant, P. (2013). Formulaicity in an agglutinating language: the case of Turkish. *Corpus Linguistics and Linguistic Theory*, 9 (1), 1–38.
- Durrant, P. (2015). Lexical Bundles and Disciplinary Variation in University Students' Writing: Mapping the Territories. *Applied Linguistics*, 1–30. <http://doi.org/10.1093/applin/amv011>
- Durrant, P., & Doherty, A. (2010). Are high-frequency collocations psychologically real? Investigating the thesis of collocational priming, 6 (2), 125–155.

- Durrant, P., & Mathews-Aydinli, J. (2011). A function-first approach to identifying formulaic language in academic writing. *Journal of English for Specific Purposes*, 30 (1), 58–72.
- Durrant, P., & Siyanova-Chanturia, A. (2015). Learner Corpora and Psycholinguistics. In S. Granger, G. Gilquin, & F. Meunier (Eds.), *The Cambridge Handbook of Learner Corpus Research* (pp. 57–78).
- Niemi, J., Laine, M., & Tuomainen, J. (1994). Cognitive morphology in Finnish: foundations of a new model, 9 (3), 423–446.
- Oflazer, K., Çetinoğlu, Ö., & Say, B. (2004). Integrating morphology with multi-word expression processing in Turkish. In T. Takaaki, A. Villavicencio, F. Bond, & A. Korhonen (Eds.), *Proceedings of the ACL Workshop on multiword expressions: Integrating processing, Barcelona, Spain* (pp. 64–71). Retrieved from internal-pdf://oflazer-1-3155256832/oflazer-1.pdf internal-pdf://Oflazer & El-Kahlout 2007 level of representation article-2266114560/Oflazer & El-Kahlout 2007 level of representation article.pdf on 9 March 2010.
- Polio, C. (2012). Editor's Introduction. *Annual Review of Applied Linguistics*, 32, vi-vii.
- Siyanova-Chanturia, A., & Martinez, R. (2014). The idiom principle revisited. *Applied Linguistics*. <http://doi.org/10.1093/applin/amt054>
- Stefanowitsch, A., & Gries, S. T. (2003). Collocations: Investigating the interaction of words and constructions, 8 (2), 209–243.
- Tannen, D., & Öztek, P. C. (1981). Health to our mouths: Formulaic expressions in Turkish and Greek. In F. Coulmas (Ed.), *Conversational Routine* (pp. 37-54). The Hague: Mouton.
- Wray, A. (2002). *Formulaic language and the lexicon*. Cambridge: Cambridge University Press.