

# *Çok Sözcüklü İfadeler*

GÜLŞEN ERYİĞİT

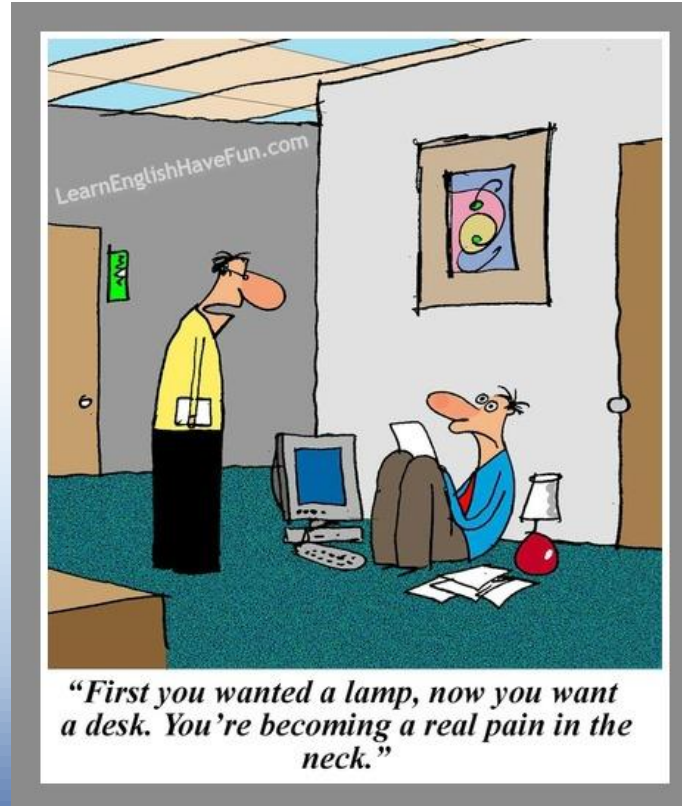
İstanbul Teknik Üniversitesi, Bilgisayar ve Bilişim Fakültesi



# «MultiWord Expressions: A pain in the neck for NLP»

pain in the neck

A person or thing that is extremely annoying or inconvenient.



karın ağrısı



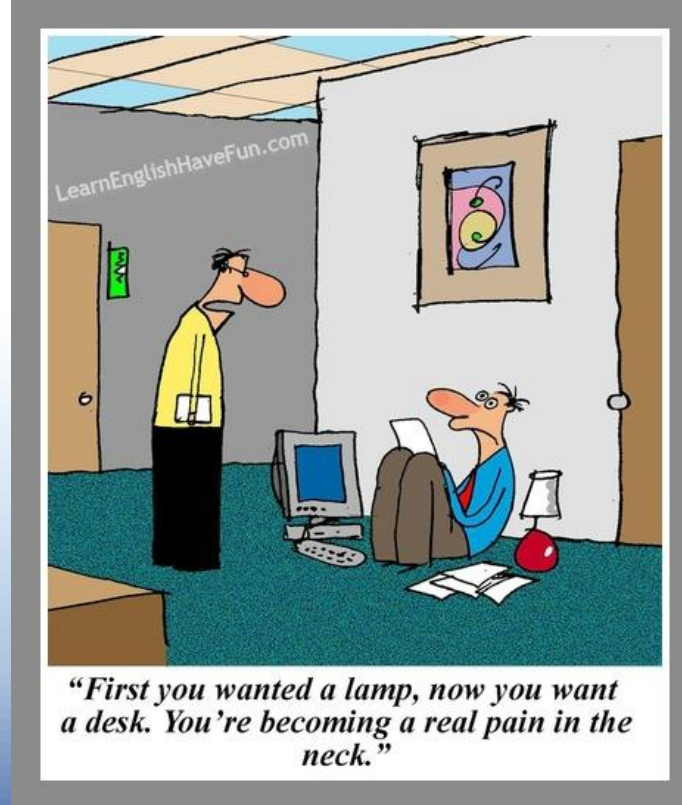
1. *isim* Karında duyulan ağrı.
2. *isim, mecaz* Çekilmeyen, seilmeyen kimse.
3. *isim, mecaz* Adı, niteliği bilinmeyen şey.
4. *isim, mecaz* Huzursuzluk, rahatsızlık veren iş, olay veya düşünce.



# «Çok Sözcüklü İfadeler Doğal Dil İşleme için tam bir karın ağrısı»

pain in the neck

A person or thing that is extremely annoying or inconvenient.



karın ağrısı



1. *isim* Karında duyulan ağrı.
2. *isim, mecaz* Çekilmeyen, seilmeyen kimse.
3. *isim, mecaz* Adı, niteliği bilinmeyen şey.
4. *isim, mecaz* Huzursuzluk, rahatsızlık veren iş, olay veya düşünce.



# *Çok Sözcüklü İfade*

- *“İstatistikî olarak tesadüften daha sık bir arada ortaya çıkan sözcükler”  
(Carpuat ve Diab 2010)*
- *“Bir dilbilimsel analiz düzeyinde tek bir birim olarak hareket eden bir sözcük dizisi” (Calzolari ve ark. 2002)*
- *“Kelime sınırlarını geçen özel yapılar” (Sag vd. 2002)*
- *“(a) birden çok sözcüğe ayrıştırılabilir ve  
(b) sözcüksel, sözdizimsel, anlamsal, pragmatik ve / veya  
istatistiksel deyimliliği gösteren sözcüksel öğeler”  
(Baldwin ve Kim 2010)*

# *Sunum İerięi*

- *ok Sözcüklü İfade Türleri*
- *Kavramsal ereve*
- *Keşif ve Tespit Yöntemleri*
- *Türke ve Diğer Diller için Kaynaklar*
- *alışmalarımız*

# *Çok Sözcüklü İfade Türleri*

- *Çok Sözcüklük Adlandırılmış Varlıklar*

*(örn. Gülşen Eryiğit, İstanbul Ticaret Odası)*

- *Deyimler*

*(örn. ayvayı yemek)*

- *Yardımcı Fiiler*

*(örn. sohbet etmek, engel olmak)*

- *Parça – Fiil*

*(örn. dönmek, -e dönmek –den dönmek)*

- *Tamlamalar*

*(örn. kırmızı kart)*

- *Çok Sözcüklü Terimler*

*(örn. Doğal Dil İşleme)*

- *Çok sözcüklü işlevsel yapılar*

*(örn. ya da, -ar –maz)*

# *Çok Sözcüklü İfadelerin Özellikleri*

- *Birliktelik (sert kahve ✓ güçlü kahve X)*
- *Devamsızlık (Kapıyı iki kere çaldı.)*
- *Birleşmeme (ayvayı yemek)*
- *Değişiklik (kafaları çekelim, kafayı çektik)*
- *Belirsizlik (Dolaptaki üç ayvayı yedim.)*

# *Topluluk*

- *Avrupa Cost Aksiyonu «PARSEME» 2013-2017  
(PARSing and Multi-word Expressions)*

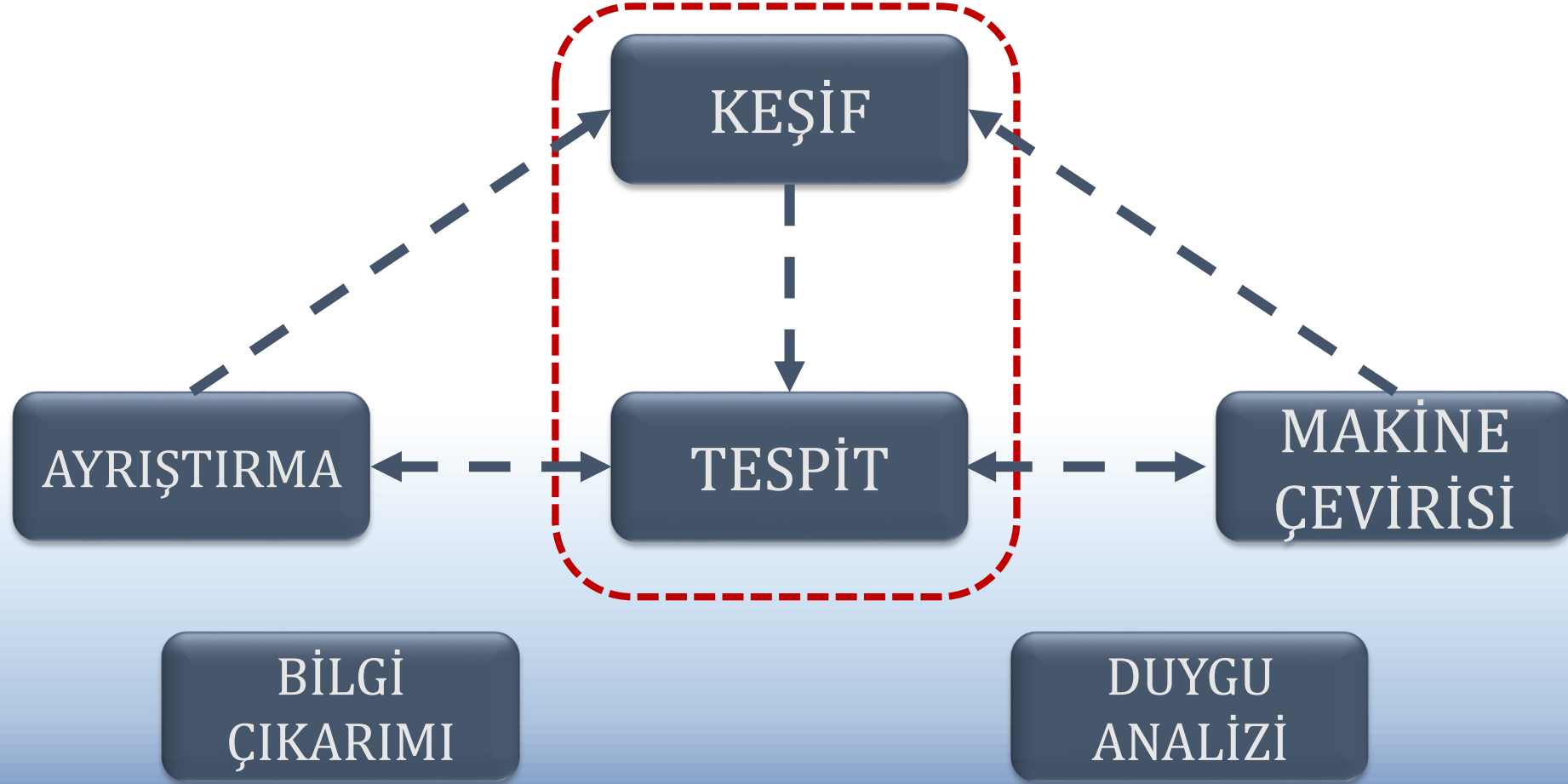


- *ACL özel ilgi grubu SIGLEX-MWE 2017-...*

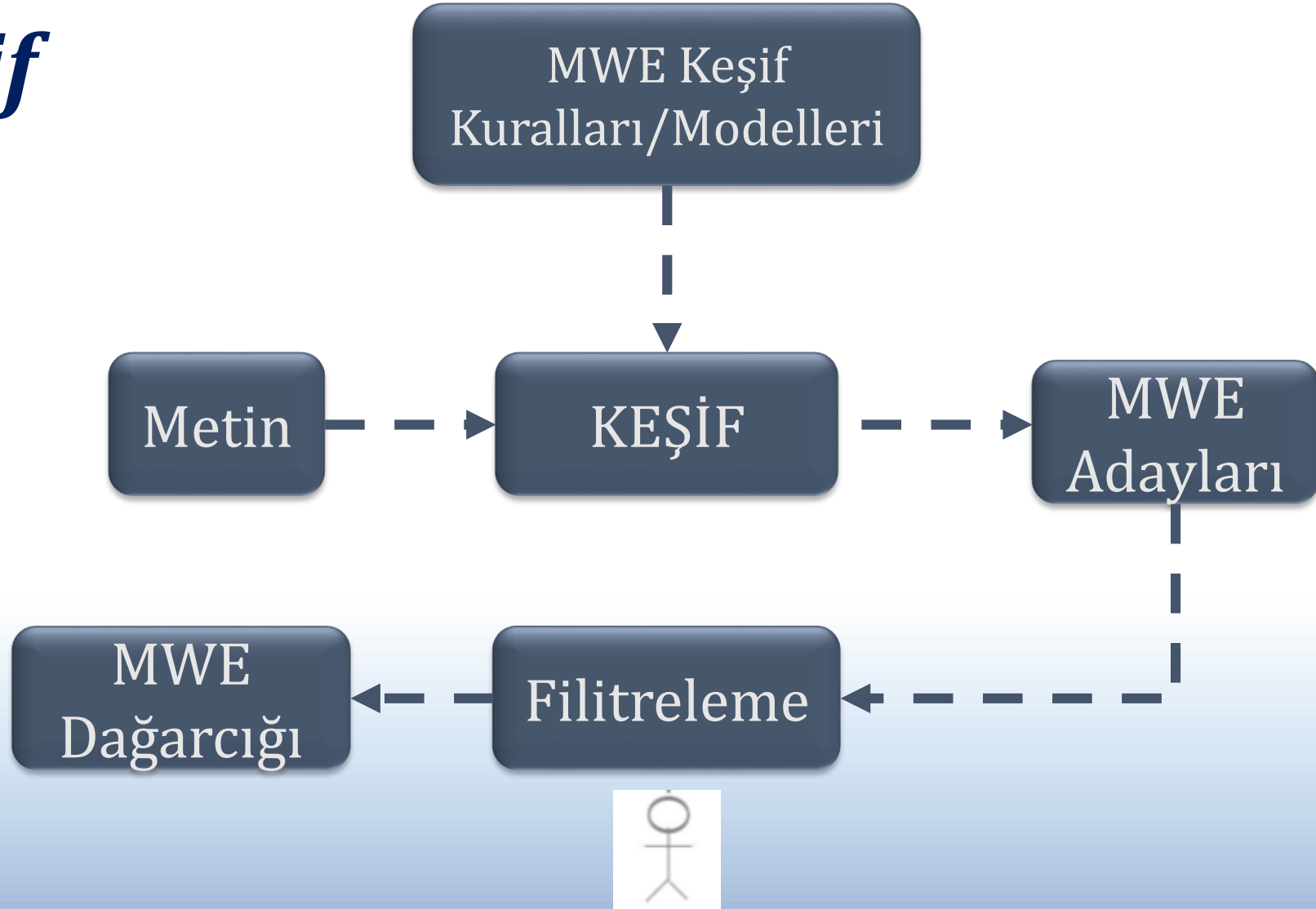


# *Çok Sözcüklü Terimlerin İşlenmesi*

## *Kavramsal Çerçeve*



# *Keşif*



# *Keşif Yöntemleri*

- *Birliktelik Ölçütleri : Nokta Tabanlı Ortak Bilgi, Hipotez Testleri (t-test, Kikare testi ...)*
- *Yer değiştirme*
- *Sözcük Vektörleri – Anlamsal Benzerlik*
- *Denetimli öğrenme*

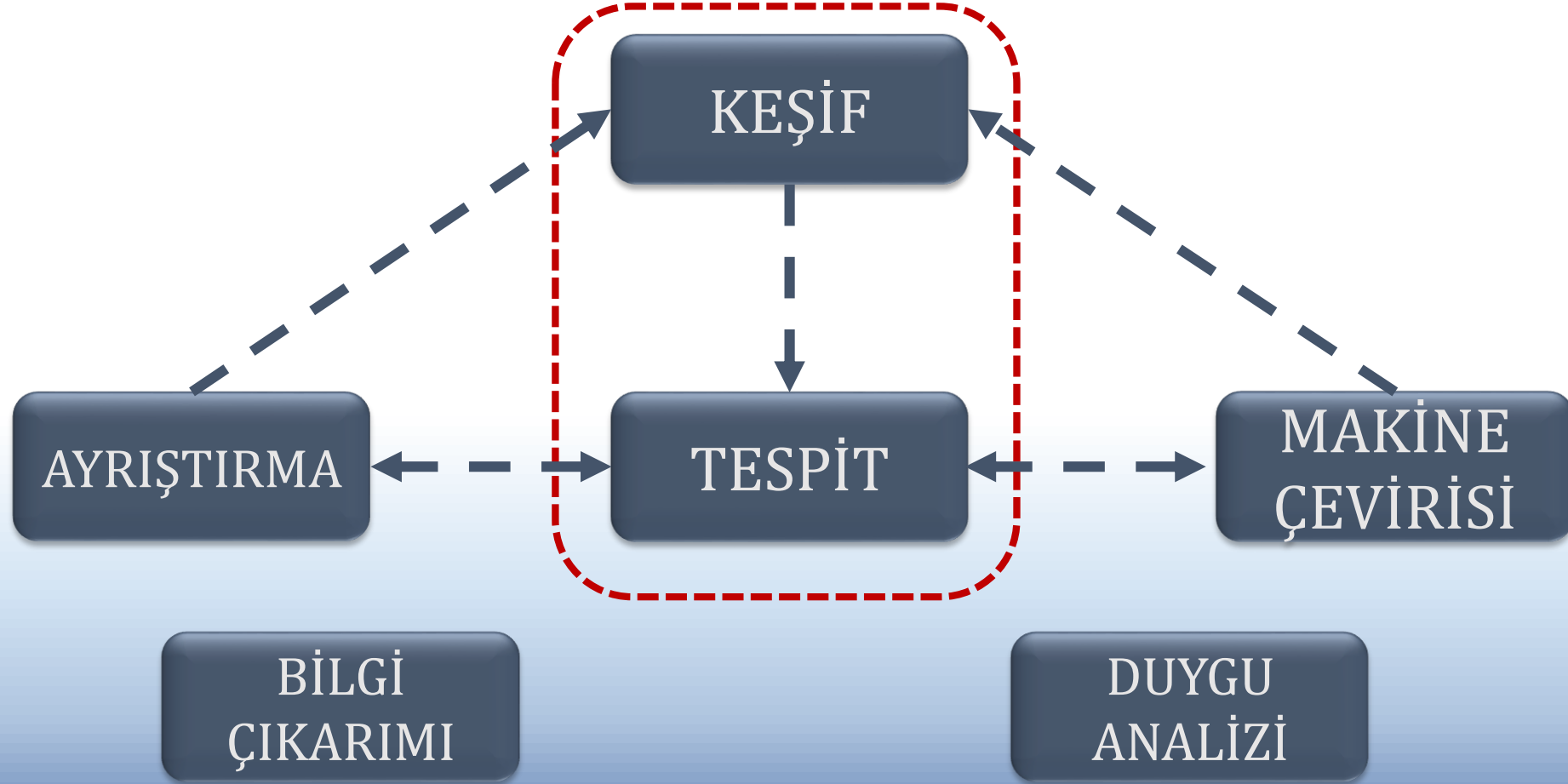
# *Tespit Yöntemleri*

Now	that	I	looked	the	dirty	word	up	,	I	understand	.
1	1		2		3	3	2				
B	I	O	B	o	b	i	I	O	O	O	O

- *Kural Tabanlı*
- *Anlamsal Belirsizlik Giderme*
- *Dizilim Etiketlemesi*

# *Çok Sözcüklü Terimlerin İşlenmesi*

## *Kavramsal Çerçeve*



# Çalışmalarımız ve Kaynaklar

- Eryigit, G., Ilbay, T., & Can, O. A. (2011, October). **Multiword Expressions in Statistical Dependency Parsing**. In *The Second Workshop on Statistical Parsing of Morphologically Rich Languages (SPMRL 2011)* (p. 45).
- Kübra Adalı, Tutkum Dinç, Memduh Gokirmak and Gülşen Eryiğit. 2016. **Comprehensive annotation of multiword expressions for Turkish**. In *TurCLing 2016, The First International Conference on Turkic Computational Linguistics at CICLING 2016*, pages 60–66, Konya, Turkey, April.
- Umut Sulubacak and Gülşen Eryiğit. 2018. **Implementing universal dependency, morphology and multiword expression annotation standards for Turkish language processing**. *Turkish Journal of Electrical Engineering & Computer Sciences*, 26(3):1662-1672.
- Mathieu Constant, Gülşen Eryiğit, Johanna Monti, Lonneke van der Plas, Carlos Ramisch, Michael Rosner and Amalia Todorascu. 2017. **Multiword Expression Processing: A Survey**. *Computational Linguistics*, 43(4):837–892
- Savary, A. Agata Savary, Marie Candito, Verginica Barbu Mititelu, Eduard Bejček, Fabienne Cap, Slavomir Čéplö, Silvio Ricardo Cordeiro, Gülşen Eryiğit, Voula Giouli, Maarten van Gompel, Yaakov HaCohen-Kerner, Jolanta Kovalevskaitė, Simon Krek, Chaya Liebeskind, Johanna Monti, Carla Parra Escartín, Lonneke van der Plas, Behrang QasemiZadeh, Carlos Ramisch, Federico Sangati, Ivelina Stoyanova, Veronika Vincze (2018). **PARSEME multilingual corpus of verbal multiword expressions**. In *Multiword expressions at length and in depth: Extended papers from the MWE 2017 workshop (Vol. 2, p. 87)*. Language Science Press.
- Constant, M., Eryiğit, G., Ramisch, C., Rosner, M., & Schneider, G. **Statistical MWE-aware parsing**. *Representation and parsing of multiword expressions: Current trends*, (Vol. 3, p.147). Language Science Press.

# Questions

