Anlam Belirsizliği İçeren Türkçe Sözcüklerin Hesaplamalı Dilbilim Uygulamalarıyla Belirginleştirmesi

Zeynep Altan
Maltepe Üniversitesi
Bilgisayar Mühendisliği Bölümü
zaltan@maltepe.edu.tr

Zeynep Orhan
Fatih Üniversitesi
Bilgisayar Mühendisliği Bölümü
zorhan@fatih.edu.tr

Özet: Birden fazla anlama sahip herhangi bir sözcüğün tümce içerisindeki yeri önemli olmaksızın, hangi anlamda olduğu ile ilgili genellikle bir belirsizlik mevcuttur. Kişi, anlam belirsizliği olan bir tümceyi anladığı zaman belirsizliğe neden olan sözcüğün diğer anlamlarını elemiş, sadece bir anlamını göz önüne almıştır. İnsanoğlu anlama işlemini gerçekleştiren bilişsel bir sisteme sahip olduğu için, belirsizlik içeren bir tümcenin anlaşılması insan dil anlama sisteminde olası anlam kümesi içinden uygun anlamın seçilmesidir. Bu belirsizlik durumunun sözcüğün uygulanma şekline göre uygun algoritmaların kullanılması ile çözümlenmesi doğal dil işlemleme alanındaki çalışmalarla başlamış, hesaplamalı dilbilim çalışmalarının yaygınlaşması ile daha da önem kazanmıştır. Sözcük anlamlarının belirginleştirilmesi çalışmaları, dilin doğasını inceleyen dilbilim çalışmaları, metin veya konuşmaların çevirisinin yapıldığı bilgisayarlı çeviri sistemleri, bilgi yönetimi teknolojisi olarak kullanılan bilgi geri getirimi ve çıkarımı, İnternet üzerinde arama motorlarının tasarımı gibi çok geniş uygulama alanına sahiptir.

Anahtar Sözcükler: Doğal dil işleme, hesaplamalı dilbilim, kelime anlamını belirginleştirme, belirsizlik, bütünce.

Summary: Any word including more then one senses which is unimportant its part of speech generally contains ambiguity. When a person comprehends an ambiguous sentence, he or she has eliminated the other meanings of obscure word. Since human beings possess a cognitive system realizing the judgment process, comprehension an ambiguous sentence with human language understanding principle means to choose the appropriate meaning from a probable sense set. This ambiguity situation was partially solved with the initiating of natural language processing studies by applying the appropriate algorithms to the complex words, and became more emphasis as the computational linguistics researches extended. Word sense disambiguation encompasses very broad application areas such as linguistic studies which analyze the features of a natural language, machine translation systems translating the texts or speeches, information retrieval and extraction utilized as the knowledge management technology and the design of research engines on Internet.

Key Words: Natural language processing, computational linguistics**,** word sense disambiguation, ambiguity, corpus.

1.Giriş

Doğal dil işlemleme çalışmaları ilk dönemlerinde yapay usun küçük bir uygulama alanı olarak sınıflandırılırken, oldukça kısa bir sürede araştırma konuları genişleyerek tek başına incelenen bir disipline dönüşmüştür. Bu hızlı değişimin nedeni son yıllarda bilginin ön plana çıkarak bilgi toplumu kavramını doğuran bilişim alanındaki önemli gelişmelerdir. İnternet üzerinden bilgiye ulaşmanın ve iletişimin kolaylaşması ile birlikte, erişilecek bilgi miktarı da aynı oranda artmıştır. Öte veri (meta data) olarak adlandırılabilecek geniş kapsamlı bilginin düzenlenerek yeniden elde edilmesi, uygun olanlarının çıkarılması, özetlenmesi, hatta belli bir kategoriye göre indekslenmesi doğal diller üzerinde dilbilim çalışmalarının kısmen ya da tamamen otomatikleştirilmesini gerektirmiştir. Ayrıca, farklı toplumların çeşitli ticari ve kültürel ilişkileri çoklu diller arasındaki çeviri sistemleri ile gerçekleşmektedir. Çeviri sistemleri İnternet üzerinden iletişimde de büyük kolaylıklar sağlamaktadır. Birer doğal dil işlemleme uygulama alanı olan tüm bu dilbilimsel incelemelerde, diğer dillerde olduğu gibi Türkçe için de pek çok farklı anlama sahip sözcüklerin anlamalarının belirginleştirilmesi, çözülmesi gereken önemli dilbilimsel problemlerden biridir. Bu problemin çözümü ise, insan-makine iletişimi arasındaki sorunlar kaldırılabildiği ölçüde basitleşecektir. Örneğin, her gün ortalama yirmi milyon sözcük içeren teknik bilginin İnternet ortamına aktarılması ile gittikçe büyüyen bir bilgi okyanusu oluşmaktadır. Kişinin dakikada ortalama bin sözcük okuyabildiği düşünüldüğünde, her gün eklenen bilginin okunabilmesi için günde sekiz saatten bir buçuk ay süre gerekir. Bu süre içerisinde eklenen yeni bilgiler için de beş buçuk yıla gereksinim vardır. Özetle, kişinin elektronik ortamdaki yenilikleri normal koşullarda takip edebilmesi ancak ilave yardımcılarla mümkündür (Bird, 1996).

Yapay us, dilbilim, mantık, psikoloji ve bilişsel bilimler gibi farklı bilim dalları arasında yer alan hesaplamalı dilbilim ise, doğal dilleri mantıksal olarak modeller ve hesaplamalarını gerçekleştirir. Hesaplamalı dilbilim, dilin işlenmesini matematiksel olarak ifade etmek üzere anlam ve anlatım arasındaki hesaplama yeteneğini araştırır. Bu özelliği ile dilbilim ve bilişsel bilimin bir parçasıdır. Diğer taraftan hesaplamalı dilbilim, anlam ve biçim arasındaki dönüşümü gerçekleştiren bilgisayar programlarının kullanılması ile mühendislik bilminin bir alt dalı olarak sınıflandırılır. Ayrıca hesaplamalı dilbilimin matematiksel dilbilim ve teorik bilgisayar bilmi ile yakın ilişkisi, anlam ve biçim arasındaki dönüşüm hesaplamalarının çeşitli özelliklerinin incelenmesi gerekliliğini ve bunun genel hesaplama teorisi ile bağlantısını açıklar (Uszkoreit, 2000).

Aklın çalışmasının araştırılması bilişsel bilim ve dilbilim arasındaki ilişki olarak tanımlanabilir. Burada, dinamik ve disiplinler arası bir yaklaşım olarak dil ve düşüncenin kökenlerinin araştırılmasına önem verilir.Bilişsel bilim insanları bir taraftan bilgisayarların düşünüp düşünmediği sorusunu cevaplamaya çalışırken, diğer taraftan öğrenme ve hatırlamanın nasıl gerçekleştiği, çevreyi algılama yetisi, beyin ve us ilişkisi, usun gelişimi gibi zihinsel prosesleri de araştırır. Dilbilimciler ise dilin yapısı, tarihi, felsefe ve psikolojisini inceler. Bu alandaki araştırmalara örnek olarak dillerin özellik ve edinimleri, dillerin gelişimi, zaman içerisindeki değişimleri ve dilin beyindeki örgütlenişi verilebilir.

Doğal dil işlemleme uygulamalarında da bu yaklaşımlar göz önüne alınarak, her bir uygulamanın ait olduğu disiplinin gözetiminde, farklı teknik ve yaklaşımlarla çözümleme gerçekleştirilir. Araştırma yöntemlerindeki farklılıklara rağmen, bilişsel bilim insanları usun beynin bir fonksiyonu olduğu, düşüncenin bir hesaplama türü olduğu, dil ve bilişin uzmanlaşmış bir dizi işlem ve betimlemelerle anlaşılabileceği şeklinde ortak bir fikre sahiptir. Kelime anlamlarının açıklaştırılmasının gerekli olduğu uygulamalarda ise, örneğin özelliklerine göre farklı yaklaşımlar içeren algoritmalardan yararlanılır.

2. Sözcük Anlamının Belirginleştirilmesi

Doğal dil işlemleme çalışmaları ana ve ara uygulamalar olarak iki gruba ayrılabilir. Ana uygulamalar bilgisayarla çeviri, otomatik özetleme,bilgi çıkarımı, bilginin yeniden eldesi gibi kendi başına bir uygulama oluşturan örneklerdir. Ara uygulamalar ise, tümceyi öğelerine ayırma, çözümleme, biçimbilimsel analiz (sözcük ek ve köklerini bulma), sözcük anlamını belirginleştirme gibi ana uygulamalar için gerekli işlemleri gerçekleştirirler. Bir sözcüğün tümce içinde hangi öğeye karşılık kullanıldığının bilinmesi ayırt edici bir özelliktir (Agirre ve diğerleri, 2001). Örneğin, "yüz" sözcüğü "yüz lira" ve "denizde yüz" kullanımlarında ad ve eylem olmasına göre anlamlandırılacaktır. Kök sözcüklerle türemiş sözcükler arasındaki ilişkiler bir başka ara uygulama olarak, sözcüğün anlamının belirginleştirilmesine katkı sağlar. "Git" sözcüğünün "gittik" şeklinde kullanıldığında eylem olduğu, "gideri" sözcüğünün ise ad olarak kullanıldığı biçimbilimsel analiz sonucunda çıkarılacaktır. Farklı anlamları olan sözcüklerden "kara" ise, "karaya çıkmamıza çok az kaldı" ya da "kara kara düşünmek" şeklinde kullanıldığında, sözcüğün hangi anlama geldiğini açık olarak belirler. Bu tür sözcükler yardımcı sözcükler olarak sınıflandırılır. Anlamsal sözcük birliktelikleri ise, "yüz-sayı" birlikteliği ile bir taksonomi, "yüz-deniz"birlikteliği ile durum, "yüz-spor" birlikteliği ile konu tanımlar.

Sözcük anlamının belirginleştirilmesi farklı bir uygulama alanı olarak mesaj anlama, insan-makine iletişimi gibi amacın anlama olduğu uygulamalarda mutlaka gerçeklenmelidir. Ayrıca, amacın anlama olmadığı çalışmalar da belirginleştirmeyi gerektirir (Ide ve Veronis, 1998). Örneğin, bilgisayarlı çeviri sistemlerinde sözcük kaynak dilde belirsiz olabilir veya hedef dile birden fazla şekilde çeviri yapılabilir. "Yüz" sözcüğü kullanıldığı yere göre, İngilizce'ye "swim", "float", "skin", "face", "surface", "cheek", "hundred" gibi farklı şekillerde çevrilebilir. Doğru olan sözcük belirsiz sözcüğün anlamının belirginleştirilmesi ile seçilecektir. Bilgi çıkarımının yapıldığı bir uygulamada önceden belirlenmiş bir anahtar sözcük taranırken, sözcüğün farklı anlamlarını elemek arama sonuçlarının kalitesini arttıracaktır. Örneğin "fare" sözcüğü bilgisayar terimi olarak arandığında, hayvan olarak kullanıldığı anlamın elenmesi çözüme ulaşmayı kolaylaştırır. Sözcük anlamlarını belirginleştirmeye bir başka yeni yaklaşım biçimi ise, metnin içinde incelenen sözcükten önceki ve sonraki sözcüklerin kavramsal olarak sınıflandırılmasıdır. Sözcük kategorilerini oluşturan bu ontolojik sıradüzen aynı zamanda bir anlamsal ağ yapısı oluşturur. Bu tür sözcük anlamını açıklaştırma yaklaşımları günümüzde pek çok doğal dil uygulamasında kullanılmaktadır (Mihalcea R. ve diğerleri 2004, Altıntaş ve diğerleri 2005, Ide N. ve diğerleri 1998).

3. Sözcük Anlamını Belirginleştirmede Yararlanılan Kaynaklar

Sözcük anlamlarını belirginleştirmede incelenen dilde düzenlenmiş elektronik sözlüklerden, sözcüklerin kavramsal ilişkilerine göre düzenlendiği ontolojik sözlüklerden, herhangi bir konuda analizi kısmen ve tamamen oluşturulmuş derleme metinlerden yararlanılabilir ya da elektronik sözlük ve ontoloji, elektronik sözlük ve derlem, ontoloji ve derlem gibi farklı kaynakların birlikte kullanımı gerçeklenebilir. Derleme metinler terimi derlem ya da bütünce olarak ta adlandırılmaktadır. Bu konudaki ilk araştırmalar İngilizce için yapılmış olduğu için, bu dilde pek çok kaynağa ulaşmak mümkündür. Özellikle Princeton Üniversitesi Bilişsel Bilimler Laboratuarı'nda 1985 yılında Prof. A.G. Miller tarafından başlatılan WordNet projesi anlamsal bir sözlük olarak İngilizce sözcükleri eşanlamlılar kümelerinde (synsets) sınıflandırır ve sözcüklerin kısa, genel tanımlamalarını yaparak bu eşanlamlılar kümeleri arasındaki çeşitli anlamsal ilişkileri oluşturur (Fellbaum C., 1998). Burada amaç iki farklı işlevi yerine getirmektir: İlki, sözcüklerin tanımlarının verildiği bir listeyi (dictionary), sözcüklerden kavramları, kavramların özelliklerini ve kavramlar arasındaki ilişkileri (thesaurus–ontology) oluşturmak iken, ikincisi yapay us uygulamalarını ve özellikle otomatik metin analizini desteklemektir. WordNet başlangıcından itibaren ücretsiz olarak kullanılabilmektedir; yeni sürümü olan WordNet 2.1 ise, ad, eylem, sıfat, belirteç olarak sınıflandırılmış toplam 155327 farklı öğenin üst kavram (hyperonym), alt kavram (hyponym), eşanlamlılık (synonym), zıtanlamlılık (antonym), parçanın (bölümün-üyenin) bütünü (holonomy) ve bütünün parçası (bölümü –üyesi) gibi çeşitli anlamsal sınıflandırma sonuçlarını vermektedir.

Senseval Projeleri pek çok dilde sözcük anlamlarının açıklaştırılması çalışmalarının yaygınlaşmasına neden olmuştur. İlk Senseval Projesinde (1998) İngilizce, İtalyanca ve Fransızca için çalışma grupları oluşturmuştur. Senseval 2 ise 2001 yılında incelendiği dil sayısını arttırarak Baskça, Çince, Çekçe, Danimarkaca, Hollandaca, İngilizce, Estçe , Japonca, Korece, İspanyolca ve İsveççe dillerinde farklı kategorilerde düzenlenmiştir. Senseval 3 çalıştayı 2004 yılında Barselona'da yapılmış ve kapsamına diğer çalıştaylara ek olarak anlamsal rollerin tanınması, çok dilli açıklamalar, mantıksal biçimler, alt sınıflandırma edinimi gibi konular daha fazla dili kapsayacak şekilde eklenmiştir. Senseval 4 çalıştayı için duyurular yapılmaya başlanmış olup, 2007 yılında gerçekleştirilecektir. Bu çalıştaya Türk dilinin de takım olarak katılımı için öneri verilmiştir.

4. Türkçe Sözcük Anlamını Belirginleştirme Araştırmaları

Sabancı Üniversitesi'nde BalkaNet Projesi'nin bir parçası olarak Türkçe bir kavramsal sözlük hazırlanmıştır. (Bilgin O. ve diğerleri, 2004). Bulgarca, Çekce, Yunanca, Romence, Türkçe ve Sırpça olarak 6 farklı Balkan dilinde uygulanan BalkaNet projesi temel olarak Princeton WorldNet modelini kullanmıştır. BalkaNet projesi için kurulan

konsorsiyum projenin ilk aşamasında EuroWordNet[1] projesinin 1310 temel kavramını her bir çalışma takımının diline çevirmiştir. Bu kavramlar sıradüzendeki düzey sayısının yüksekliği ve pek çok alt kavram içermesi nedeni ile tüm dillerde oldukça önemli bir yapı taşı olmuştur. Birinci aşama Türkçe için eş anlamların, zıt anlamların ve alt kavramların elektronik Türkçe dilbilgisi sözlüğünden otomatik çıkarımı şeklinde gerçekleşmiştir. Daha sonra konsorsiyum incelenecek kavramların sayısının beşbine çıkarılmasını kararlaştırmış; böylece Türkçe dahil tüm takımlar bütünce sıklıkları (corpus frequencies), sözcük dağarcığının tanımlanması, tek dilli sözlükler, çoklu anlamlar (polysemy) gibi farklı kriterleri de ekleyerek alt kümelerini genişletmişlerdir.

Bir doğal dil işlemleme alanı olarak bilgi-tabanlı tekniklerle olasılı modellerin bütünleşmesi, veri tabanı sorgulamalarıyla sınırlı dil uygulamalarını zenginleştirmiştir. Böylece metinlere uygulanan istatistiksel yöntemlerle en olası yorumun tahmini mümkün olmaktadır. Bunun için de ayrıntılı olarak işlemlenmiş derleme metinlere (bütünce) gereksinim vardır. Metin örnekleri kullanarak birden fazla anlama sahip kelimelerin, özellikle eylem türündeki kelimelerin anlamlarını çıkarabilmek için, bu metinler üzerinde sözcüksel ve anlamsal bilginin doğru olarak işaretlenmiş olması önemlidir. Örneğin Türkçe için her biri yaklaşık 25000 sözcükten oluşan 7 farklı metin koleksiyonunun[2] bi-gram model üzerinde test edildiği bir çalışma yapılmıştır (Altan Z., Yanık E., 2001). Burada tümcelerin sözdizimsel ve anlamsal sınıflandırmasında sadece incelenmek istenen sözcükten önceki sözcük işaretlenmiştir. Bütünce üzerinde uygulanan olasılı dil modeli, elle tanımlanan kurallara ek bir öğrenme bileşeni olarak en olası çözümü tahmin edebilmekte ve dili işlemlemedeki belirsizlikleri de büyük ölçüde azaltmaktadır. Bu çalışmada eylemlere ait kavramsal sınıflandırma yol alma, yönelme ve terk etmeden biri şeklinde devinim (motion), kavrama (perception), duygu (emotion), fonksiyon (bodily care and functions), bağlantı (contact) gibi WordNet'in eylemler için grupladığı kavramsal özelliklerinden yararlanarak gerçekleştirilmiş ve eylemlerin sözcük anlamları bunlara göre numaralandırılmıştır. Tümcelerin işaretlenmeleri "git" eylemi için Tablo 1'de görüldüğü gibidir. Artık herhangi bir eylemin anlamı olasılığa bağlı olarak tahmin edilebilir. Tahmin için kullanılacak yöntem en olası maksimumun kestirimi (Maximum Likelihood Estimation- MLE) olabilir. MLE, işlemlenmiş bütünce içinde aranılan sözcüğün eğitilme sayısını hesaplar. Eğitim sadece bir önceki sözcüğe göre yapıldığı için araştırılan kelimeden önceki kelimenin öğelerine ayrılmış olması önemlidir. Bu sınıflandırmadan elde edilen değerler, işaretlenmiş bu bütünce üzerinde farklı anlamların belirlenmesi için bir Bayes sınıflandırması oluştururlar. Bu bütünce tümce öğeleri elle işaretlenerek elde edilmiştir.

Benzeri problemler Türkçe dil işlemleme çalışmalarının pek çoğunda mevcuttur. Fakat ODTÜ derleme metninin kullanıma açılması ile birlikte gerek sözdizimsel, gerekse biçimbirimsel olarak çözümlenmiş; bütünce bulma problemi kısmen de olsa çözümlenmiştir.

---

[1] EuroNet Projesi 1996 yılında başlayıp üç yıl süren WordNet'in Avrupa dillerine uyarlaması şelinde bir konsorsiyumun gerçekleştirdiği ortak bir çalışmadır

[2] Dünya klasiklerinden örnek hikayeler: Guliver Devler Ülkesinde , Candide , Ivan Nikiforoviç, Tours Papazı , Mozart Prag Yolunda , Mektuplar, Kır Atlı

Tablo 1: “git” eylemi için örnek işaretleme

No	Tümce
1	ülkeyi keşfetmek için yazar da [birlikte] (DuZ) $gidiyor$ {1} ve karada kalıyor
2	gemi, [Suratya] (YeZ) $gidiyordu$ {1}
3	Umut Burnuna kadar [rüzgâr] (ÖZ) [çok iyi] (DuZ) $gitti$ {3}
4	Biraz daha kuzeye dönerek Tataristanın kuzeybatısına ve [Buzdenizine] (YeZ) $gitmek$ {1} [olasılığı karşısında] (ZaZ), bulunduğumuz rotayı izlemenin daha iyi olacağını düşündük
5	[merakımı] (KEyl) $giderecek$ {5} bir şey de göremediğimden
6	[olanca hızımla] (DuZ) [o önce] (ZaZ) $gittiğim$ {1} [yana] (YeZ) koşmuştum
7	[sesim ve işaretlerim] (ÖZ) [hoşuna] (KEyl) $gitmiş$ {4} [gibiydi] (Eyl)
8	Fakat nasıl davrandığımı, kocasının işaretlerine göre ne kadar iyi davrandığımı görünce bana alıştı ve $gitgide$ {12} [artan bir sevgi] (DoT) beslemeye başladı
9	Bu yaptığım [pek] (MiZ) [hoşlarına] (KEyl) $gitmişti$ {4}
10	Ben de $gittim$ {1}, elini öptüm

Bu bütünce, ODTÜ-BAP ve TÜBİTAK tarafından desteklenmiş ve ODTÜ-Sabancı Üniversiteleri işbirliği ile gerçekleştirilmiştir. Çalışmada bir ana derleme metin oluşturulmuş; ayrıca farklı kullanımlar için bu ana derleme metinden bazı farklı özellikleri olan bir de ağaç bankası derleme metni geliştirilmiştir (Oflazer ve diğerleri, 2003). Derlemde kullanılan metinler 1990 yılı sonrası basılan eserlerden seçilmiştir. Derlemde yaklaşık olarak 2.000.000 sözcük bulunmaktadır. 201 kitap, 87 makale ve 3 tane günlük gazeteden seçilmiş haberlerden oluşan 999 farklı yazılı metin kullanılmıştır. Derlemde bulunan metinlerin çoğunluğu biçimbirimsel olarak çözümlenmiştir. Fakat yapısal belirsizlikler tamamen çözülmememiş olduğu için kullanımda bazı problemlerle karşılaşılmaktadır.

5 . Sözcük Anlamını Belirginleştirmede Kullanılan Yöntemler

Sözcük anlamının belirginleştirilmesinde en etkili çalışmalardan biri, olasılı başka deyişle bilgisayarla öğrenme (machine learning) algoritmalarının kullanılmasıdır. Bu yaklaşımlardan öngörmeli (supervised) yaklaşımların, öngörmesiz (unsupervised) yaklaşımlardan daha iyi sonuçlar verdiği gözlenmiştir[3].

Standart bilgisayarla öğrenme algoritmalarından pek çoğu öngörmeli öğrenme yaklaşımında kullanılabilir. Sonuçlar genellikle başarılı olmasına rağmen, öngörmeli

[3] Öngörmeli öğrenmede çalışılan verinin her parçasının gerçek durumu bilinmesine rağmen, öngörmesiz öğrenmede eğitim örneği içindeki verinin sınıflandırılması bilinmez. Öngörmesiz öğrenme çoğunlukla kümelendirme olarak bilinirken, öngörmeli öğrenme sınıflandırma olarak adlandırılır

yöntemler anlamsal olarak işaretlenmiş derlemlerin az olmasından veya hiç olmamasından dolayı dezavantaj oluştururlar.

Tablo 1'de küçük bir örneği verilmiş olan bütünce, sözcük anlamını belirginleştirmek üzere kullanılmış ve öngörmeli öğrenme gerçekleştirilmiştir. Sonuçlar Tablo 2'de istatistiksel bir yaklaşım olan Naïve Bayes (NB) ve örnek tabanlı Exemplar Based algoritmalarının uygulamaları olarak görülebilir (Altan ve Orhan, 2003). Tablo 3 yöntemlerde kullanılan özellikler açıklamaktadır. Küçük bir bütünce üzerinde uygulanan bu algoritmalardan örnek tabanlı yaklaşım, Naïve Bayes'e göre biraz daha iyi sonuç vermiştir .

Tablo 2: Bütünceden çıkarılan özelliklere bir örnek

No	L1P	L2P	R1P	L1M	L2M	R1M	Git Kok	Git Ek	Anlam	CL	CR	NB	FB
1	duz	-	-	-	-	-	gidiyor	gidiyor	1	birlikte	-	1	1
2	yez	-	-	e	-	-	gidiyordu	gidiyordu	1	surat	-	1	1
3	duz	öz	-	-	-	-	gitti	gitti	3	iyi	-	3	3
4	yez	-	zaz	e	-	de	gitmek	gitmek	1	buzdenizi	olasılık	1	1
5	keyl	-	-	i	-	-	giderecek	giderecek	5	merak	-	5	5
6	zaz	duz	yez	-	-	e	gittiğim	gittiğim	1	önce	yan	1	1
7	keyl	öz	eyl	e	-	-	gitmiş	gitmiş	4	hoş	gibi	4	4
8	-	-	dt	-	-	-	gitgide	gitgide	12	-	artan	1	12
9	keyl	miz	-	e	-	-	gitmişti	gitmişti	4	hoş	-	4	4
10	-	-	-	-	-	-	gittim	gittim	1	-	-	1	1

Tablo 3: Öngörmeli eğitimin yapıldığı algoritmaların özellikleri

Özellik	Açıklama
L1P	Hedef sözcüğün solundaki birinci tümce öğesi
L2P	Hedef sözcüğün solundaki ikinci tümce öğesi
R1P	Hedef sözcüğün sağındaki birinci tümce öğesi
L1M	Hedef sözcüğün solundaki birinci tümce öğesinin hal eki
L2M	Hedef sözcüğün solundaki ikinci tümce öğesinin hal eki
R1M	Hedef sözcüğün sağındaki birinci tümce öğesinin hal eki
GK	Git sözcüğünün kökü
GM	Git sözcüğünün eki
CL	Soldan birinci kalıp sözcük
CR	Sağdan birinci kalıp sözcük

İkinci bir Türkçe sözcük anlamını belirginleştirme çalışması, Bölüm 4'de açıklandığı gibi, ODTÜ-Sabancı Türkçe ağaç bankası kullanılarak gerçekleştirilmiştir. Bütünce olarak ODTÜ derleme metinlerinin kullanıldığı çalışmada, anlamları belirginleştirilecek sözcük sayısı ve tipi arttırılmış; bu sözcükler için yeni anlam sınıfları eklenmiştir (Tablo 4).

Tablo 4: ODTÜ ağaç bankası metinlerinden seçilen sözcüklerin anlam sayıları

Sözcük	Metinlerdeki tümce sayısı	Anlam sayısı
Yan	104	9
Git	189	10
Gör	133	9
Çık	231	15
Al	250	10
Gel	281	12
Yap	328	6
Ol	941	4

6 Sonuç

Türkçe sözcüklerin ortalama anlam sayısı bu alanda üzerinde çok fazla çalışma İngilizce gibi dillere göre çok daha fazladır. Ayrıca tüm gereksinmelere cevap verebilecek Türkçe bir bütüncenin de olmaması çalışmaları daha da güçleştirmektedir.

Kaynaklar

Agirre, E., Ansa, O., Martinez, D., Hovy, E., 2001, Enriching Wordnet Concepts with Topic Signatures, Proc. of the NAACL Workshop on Wordnet and other Lexica Resources: Applications, Extensions And Customizations, Pittsburg, USA, 123-132.

Altan, Z., Orhan Z., 2003, Disambiguation of Turkish Word Senses by Supervised Statistical Methods, International Journal of Computational Intelligence, Vol:1, 16-21

Altan Z. ve Yanık E., 2001 Kelime Anlamlarının İstatistiksel Çıkarımı için Metin örneklerinin İşlenmesi, İstanbul üniversitesi Elektrik& Elektronik Dergisi 1-2, 287-295

Altıntas E., Karsligil E., Coskun,V., New Semantic Similarity Measure Evaluated In Word Sense Disambiguation , 15th Nordic Conf. of Computational Linguistics, 2005

Bilgin O., Çetinoğlu Ö., Oflazer K., 2004, Building a Wordnet for Turkish, Romanian Journal of Information Science and Technology Vol: 7, Num: 1-2, 163-172.

Bird, M., 1996, System Overload. Excess Information Is Clogging The Pipes Of Commerce - And Making People Ill, In Time Magazine, December 9th, 1996, 46-47.

Fellbaum, C., 1998, WordNet: An Electronic Lexical Database, The MIT Pres

Ide N. Véronis J., Word Sense Disambiguation: The State of the Art **,** Computational Linguistics, 1998, 24(1)

Mihalcea R., Tarau P., Figa E., 2004, Pagerank on Semantic Networks with Application Toward Sense Disambiguation, The 20th International Conference on Computational Linguistics**,** 1126-1132**.**

Oflazer, K., Say, B., Tur, D. Z. H., Tur, G., 2003, Building A Turkish Treebank, Invited Chapter In Building and Exploiting Syntactically-Annotated Corpora, Anne Abeille Editor, Kluwer Academic Publishers, 2003

Uszkoreit H., 2000, Language Technology for Knowledge Management, Proceedings of Japanese-German Workshop Comp. Linguistics, Yokohama, 26 May 2000, 1-10.