Towards a Description of Word, an Urdu Corpus-based Approach

Author's Details:*Abrar Hussain Qureshi, A PhD Candidate. *** Dr.Muhmmad Akram**

Abstract.
A great deal of scholarly discussion has centered on the linguistic status of the word. It would be useful to draw attention to two fairly general and constant characteristics of words across a wide range of languages. The first is that a word is typically the smallest element that can be moved around without destroying the grammaticality of the sentence (ignoring any semantic effects) But no means all word are equally mobile in this sense, but with very few exceptions, the smallest mobile units are words. The morphemes constituting a single word have a rigidly a fixed sequential order. The second major characteristic of words is that they are typically the largest units that resist 'interruption' by the insertion of new material between their constituent parts. Even these two traditional views about the phenomenon of word do not cover all the mechanism of word. It was due to this reason that the term "Lexical Item" was introduced. Let us debate in detail that what the inconsistencies were in the traditional concept of "word" and how the "Lexical item" is a useful and neutral hold –all term that captures and helps to over come instabilities in the term "Word".

Introduction:

Traditionally word is a linguistic phenomenon, which is recognized in most languages (Katamba, 2005). The nature of words varies from language to language. The conception of word is determined afresh within the system of every language, and as a result the word as element of speech is language specific, not language universal (Haspelmath, 2002). The various kinds of language have their own broadly similar words, but even there is variation from language to language inside a category. Despite such complications certain features are more or less true for many if not all languages (Balhouq, 1976). A translation equivalent of the English, "word" is found in most languages, e.g. Urdu /lafgz/, Arabic /kalimattun/ French /mot/ German / wort /. Etc. The native speaker of any language is often in common agreement about what constitutes a word and the number of words in a given utterance (Hay, 2000). In traditional grammatical theory, as J. Lyons (1968) says, "The word is a unit par excellence ". It is the basis of the distinction that is frequently drawn between morphology and syntax and it is the principal unit of lexicography. In modern linguistics the term is the subject of extensive debate. There is no adequate universal definition of it that is acceptable to most linguists. More over linguists are not unanimous on the question of regarding the word as a linguistic unit. Seiler (2003) maintains the idea that the word is a linguistics unit. From his point of view

"A major obstacle which has prevented some linguists from seeing anything linguistically relevant in the word is the idea that the word should be a unit. We should rather give up this idea right from the start, for we feel strongly that the word cannot be called a unit in the same sense that the morpheme is so termed. The term unit is it self not well defined nor applied consistently." He stresses this view even further when he says that the word is not a unit but a constituent of a sentence or a clause. In fact most linguists are of opposite opinion on this question and do not accept the claims put forward by Seiler.

Kramsky (1969) attempts a definition that is applicable to words in most of the languages of the world. However, because of the limitation of the definition, he does not call it a universal definition. He states, "The word is the smallest independent unit of language referring to a certain linguistic reality or to a relation of such realities and characterized by certain formal features (acoustic, morphemic) either actually (as an independent component of the context) or potentially (as a unit of the lexical plan)".

Halliday (1961) prefers to use "word" purely as a unit of grammar and to use "lexical item" in lexis and semantics. In his grammatical model of "Scale and category "the word is one of five grammatical units hierarchically organized in the following order: 'sentence, 'clause, 'group, 'word, and 'morpheme

A standard definition of the word is found in a paper written in 1926 by the American linguist Leonard Bloomfield, one of the greatest linguists of twentieth century. According to Bloomfield, "a minimum free form is a word." By this he meant that the word is the smallest meaningful linguistic unit that can be used on its own. It is a form that cannot be divided into any smaller units that can be used independently to convey meaning . For example كتا ب *kitab* (book) is a word. We cannot divide it up into smaller units that can convey meaning alone.

This difference of approaches seems to originate, primarily, from the fact that the "word" is inherently the most ambiguous one. Several criteria are involved in any attempt to work out a definition of it. In what follows we will discuss some of those criteria as well as any definitions that may have been introduced as a result of their application.

The Orthographic Aspect:

Branford (1967) defines the word in terms of alphabetic or syllabic writing system: a visual sign with space around it or the form spelt out on paper, in writing or in print and probably the most powerful image of the 'word' that we have. According to Schmitt (2000)

> "Many people would consider meaning the most important aspect of learning a word, yet there has recently been an increasing awareness that orthographic knowledge, traditionally considered a lower level type of knowledge, is a key component to both vocabulary knowledge and language processing in general. This awareness stems from research that has shown both that the eye fixates on most words in a text rather than skipping over many of them, and from psychological research that has shown the complexity of orthographical decoding".

According to Mc Arthur, many people are not entirely comfortable with a new word until they have seen it securely stated on paper, when they can, as it were, photograph it for future reference. This literate image can be regarded as primary in shaping the ordinary person's view of words, as illustrated by such questions as' how do you spell that word?' Such a word is the product of a number of arbitrary decisions made at different times over thousands of years of struggling towards literacy: an alphabet of a certain kind, running in a chosen direction, with letters of certain shape and size, with conventional spacing between such letters or ways of joining letters and inscribed by various means on various selected surfaces .

In practice this definition encounters serious problems. According to Durrani, some languages have dots instead of space to mark their orthographic words: Oscan, an ancient language of Italy uses dots to mark word boundaries. Consider the text below:

STATUS.PUS.SET.HURTIN.KERRIIIN.VEZKEI.STATIF.

Table 2: Oscan Language used Dots

Some languages of south East Asia use space to mark the sentence boundaries. So the orthographic word definition will totally fail in this scenario.

Grammatically, we find that some orthographic words belong to more than one grammatical class.

N	Concordance	
268	بےخیال میرا ہے نہیں وہ نہیں پاکستان دل :دل پاکستان وہ بھی نہیں ہے آپ کے پاس دل	
269	ہوئے ضرور ہوئے نا کیوں ہوئے تو ہوئے پاگل دل ہیں بہت ہی پیارا سا نغمہ دل تو پاگل ہے صاحب یہ	
270	خواص)و ماع)کاش آم جی شاہ اکرب جی پت دل ہے۔ شدھ شری 801 شری شری پات ساہ جبی شری	
271	یہ اور شعر،یہ یہ کا روئد ے اور گا جائد ے بو پتھر دل سال، دور بو گے ہم چاہنے والوں سے اتنی دیر،	
272	،دیا جواب نے آپ ؟ ہے ہوگیا پتھر دل تھے کہ اس نے بھاگتے ہوئے کہا ، اباجان تمہارا	
273	سننےکے کرنے پیش سنگیت اعلی اور پذیر دل اعلی سنگیت پیش کرنے کم اشتراک سے ہے.حد	
274	کرلے سے سنگیت لوک میں پیرائے پذیر دل سلےقہ سے اس ساز کا استعمال کرتے ہیں۔ اور	
275	سےدارو دوا تو چوٹ کی تن ہے۔ لگی چوٹ پر دل ہیں پست شکست خوردہ۔ سکریٹری بابپ کے	
276	کیدکھانے مسرور کو اپن نہیں۔ ہی غم کچھ پر دل طرح مہمانوں میں #گھوم رہی تھی۔ گویا اس کے	
277	اختلاجاور پوھلنے سانس اور ہے پڑتا زور بہت پر دل کمزور آدمیوں کے لئے بہت مضر #ہے، اس سے	

In the above concordance, the node word دل *dil* (heart) is functioning as verb and noun. Words like گدھا *ghaddaha* (an animal and a foolish man) belong to the class of nouns and adjectives. As to the semantics, we may find that one orthographic form possesses more than one meaning, e.g.چاند *chand* (as moon and beloved son or daughter) as in the following concordance:

کےقدرت مال۔ کو دےکھنے واقعی چہرہ سا چاند وہ حقیقٹ میں نظر آئے۔ سےاہ نقاب میں سے	135	
ص6 29 (161)تھے دنےے لگا چاند اصل جشن انقلاب نے اس رات میں اور بھی چار	136	
عملپر اسی وہ اور ہے وجود کا سورج اور چاند پر پوتے اس پر اس وقت تک قائم رہیں جب تک کہ	137	
پرورسلڑکا کا اس ہوا۔ بےاہ سے بدھ لڑکے کے چاند نسل سے تھے۔ منو کی اےک لڑکی الا تھی۔ جس کا	138	
پروںکے مور صورت،پیاری تیری میں چاند کی وحشت بھری آنکھوں میں تیری نگابیں۔	139	
مضبوطاور محراب صحےح لےکن لگادےتی چاند سے سجاتے اور پھر نبت کاری عمارت کو چار	140	
راسمختلف کے سورج بڑھنے،اور گھٹنے کے چاند پر ادا کرنا اتنا ضروری سمجھا جاتا تھا کہ موسموں،	141	
علمیخالص اور تھا واقف سے سبب کے گہن چاند عبارت نقل کرکے ہے ثابت کےا ہے کہ اگرچہ وہ	142	
راتاور دن تبدےلی کی موسموں گہن سورج اور چاند آسمان اور دنےا کے نظرےہ طبقوں پر مشتمل ہے،	143	
پرندےاور درندے جانور،جنگلی مختلف سورج چاند، ڑلپ ،اےرد ،تخرد ںےم ئاشا نا 2 (14)6	144	
شاملتصوےرےں کی جانوروں ان اور سورج چاند، پرندوں کی تصوےروں کے علاوہ درختوں، پھولوں،	145	
جےسے بھی،(Bound)پابندی اور وغےرہ چاند آزاد بھی ہو سکت ےہیں؛ جےسے قلم، کتاب،	146	
مثالکی استثنا ہے وغےرے۔ گنگا اور سورج چاند، اسما پر اس اصول کا اطلاق نہیں ہوتا؛ جےسے	147	
کےزمانہ اس ۔ بے پھےنکی اتار ٹوپی اپنی نے چاند کے متعلق لکھتے ہیں کہ اس مےنار کو دےکھ کر	148	
اورکردے حدہ علٰےسے جلاداس کوئی جلد کی چاند دی ، کیڑے کی مدح میں کہتے ہیں کہ اگر	149	
سےچاند کا ستاروں ہے۔ نہیں صحےح ہے چاند کےا ںےا ںوراٹس ہو کف لاک ےےگ (Harcourt)	150	
کاکابےں نہ لےکن ہے۔ نہیں تعلق کوئی سے چاند میں اےک چاند ہے صحےح نہیں ہے۔ ستاروں کا	151	
کاماضی ہے کےجےے۔ غور پر گڑھوں کے پر چاند میں جمع شدہ رکارڈ اور ےاد داشتیں ہیں۔	152	
اطرافکے اس سب سےارے اور ستارے چاند، کہ زمےن ساکن اور مرکز کائنات ہے اور سورج،	153	
ہیںبھجتی طرف کی سےاروں دوسرے ےا پر چاند کی ب نسبت سادہ ہیں۔ جب ہم خلائی جہازوں کو	154	
کےٹوٹتے کے ستاروں اور گرہن سورج گرہن،چاند کو سنتے ہیں جس کے لئں ہمارے پاس کان ہیں۔	155	
اورلذتوں کی لمس وشبوں، خَ پھولوں،ستاروں،چاند کی طرف مائل ہوت ہیں۔ چنانچہ پہاڑ، درےا	156	

On the other hand we may find that one orthographic form refers in a certain context to the same object, concept or idea, e.g. جاہل *jahil* and ان پڑھ *unparh* (illiterate) نا لائق *na-laaiq* and بیوقوف *baywaqoof* (foolish) etc.

However, even in written text there are potential theoretical and practical problems with an orthographic definition. For example, if 'bring, brings, bringing, brought' or 'long, length, and lengthen' or, less obviously, 'good, better, and best' are separate words, would we expect to find each word from the set listed separately in a dictionary. If so, why and if not, why not. Are these one word or several. So an orthographic definition is one that is formalistic in the sense of being bound to the form of a word in a particular medium. It is not sensitive to distinctions of meaning or grammatical function. To this extent it is not complete .

The Semantic Aspect:

Another criterion that is used to identify a word is meaning . This criterion is based on the belief that each word has a meaning, and that, in a language, each unit of meaning or segment of meaning could be identified and separated from other units of meaning. Each item thus separated is called a word.

This cannot be used as an independent criterion to define word because the meaning factor used here poses a problem in distinguishing morpheme and word. When this criterion alone is applied to identify a word, it is not possible to distinguish between a word, a phrase and an idiom; even groups of words that combine to give a single meaning will come under this category. Another problem with this criterion is that of word boundary and meaning boundary; if they coincide it is all right; if they do not coincide, this criterion will not work.

135	کے قدرت مال۔ کو دیکھنے واقعی چہرہ سا چاند وہ حقیقت میں نظر آئے۔ سیاہ نقاب میں سے
136	ص6 29 (161)تھے دئیے لگا چاند اصل جشن انقلاب نے اس رات میں اور بھی چار
137	عملپر اسی وہ اور ہے وجود کا سورج اور چاند پر پوتے اس پر اس وقت تک قائم رہیں جب تک کہ
138	پرورسلڑکا کا اس ہوا۔ بےاہ سے بدھ لڑکے کے چاند نسل سے تھے۔ منو کی ایک لڑکی الا تھی۔ جس کا
139	پروںکے مور صورت؛پیاری تیری میں چاند کی وحشت بھری آنکھوں میں تیری نگاہیں۔
140	مضبوطاور محراب صحیح لیکن لگادیتی چاند سے سجاتے اور پھر نبت کاری عمارت کو چار
141	راسمختلف کے سورج بڑھنے،اور گھٹنے کے چاند پر ادا کرنا اتنا ضروری سمجھا جاتا تھا کہ موسموں،
142	علمیخالص اور تھا واقف سے سبب کے گہن چاند عبارت نقل کرکے یہ ثابت کیا ہے کہ اگرچہ وہ
143	راتاور دن تبدیلی کی موسموں گہن سورج اور چاند آسمان اور دنیا کے نظریہ طبقوں پر مشتمل ہے،
144	پرندےاور درندے جانور،جنگلی مختلف سورج چاند،ڈلپ ،ایرد ،تخرد یےم ئاشا نا 2 6(14
145	شاملتصویریں کی جانوروں ان اور سورج چاند، پرندوں کی تصویروں کے علاوہ درختوں، پھولوں،
146	جیسے بھی،(Bound)پابندی اور وغیرہ چاند آزاد بھی ہو سکت ہیں؛ جیسے قلم، کتاب،
147	مثالکی استثنا ہے وغیرہ۔ گنگا اور سورج چاند، اسما پر اس اصول کا اطلاق نہیں ہوتا؛ جیسے
148	کےزمانہ اس ۔ ہے پھینکی اتار ٹوپی اپنی نے چاند کے متعلق لکھتے ہیں کہ اس مینار کو دیکھ کر
149	اورکردے حدہ علیٰسے جلاداس کوئی جلد کی چاند دی ، کیڑے کی مدح میں کہتے ہیں کہ اگر
150	سےچاند کا ستاروں ہے۔ نہیں صحیح ہے چاند کےا ہے م ںورلس ہو ک لگ ہےگ (Harcourt)
151	کاکابہے نہ لیکن ہے۔ نہیں تعلق کوئی سے چاند میں ایک چاند ہے صحیح نہیں ہے۔ ستاروں کا
152	کاماضی ہے کے جیے۔ غور پر گڑھوں کے پر چاند میں جمع شدہ رکارڈ اور یاد داشتیں ہیں۔
153	اطرافکے اس سب سیارے اور ستارے چاند، کہ زمیں ساکن اور مرکز کائنات ہے اور سورج،
154	ہیںبھجتے طرف کی سیاروں دوسرے ےا پر چاند کی بہ نسبت سادہ ہیں۔ جب ہم خلائی جہازوں کو
155	کیٹوٹنے کے ستاروں اور گرہن سورج گرہن،چاند کو سنتے ہیں جس کے لئے ہمارے پاس کان ہیں۔
156	اورلذتوں کی لمس وشبوں، خَ پھولوں،ستاروں،چاند کی طرف مائل ہوت ہیں۔ چنانچہ پہاڑ، دریا

In the above concordance the node word چاند *chand* , at least in a few lines, is not a semantic unit until it is analysized in its context. So sometimes, a meaning is not in single words rather in the whole context.

The psychological Aspect:

The native speaker's consciousness of the existence of the word as a functioning unit of his/her language reveals that there is certain psychological validity in the recognition of word as a language unit. According to E.Sapir,

> "Linguistics experience both expressed in standardized written form and as tested in daily usage indicates overwhelmingly that there is not, as a rule, the slightest difficulty in bringing the word to consciousness as a psychological reality. No more convincing test could be desired than that of the native, quite unaccustomed to the concept of the written word, has, nevertheless no serious difficulty in dictating a text to a linguistic student word by word" .

It is fairly clear, therefore, that the psychological aspect plays a significant role in the identification of words. But it has to be admitted that at present at least there is no way of studying this criterion objectively. Any investigation in this respect will inevitably involve techniques that are subjectively founded.

The phonological Aspect:

The form that emerges in speech and is heard clearly by the attuned listener but not necessarily by any one else. The attuned listener usually means a fluent native adult, but can in many cases mean an adult with the right technical, professional or other knowledge and expectations so as to be able to interpret the area within a flow of sound, which constitutes a word .

According to Durrani (2007) in Hebrew last syllable is regularly stressed so the word boundary is likely to fall after each stressed syllable. In Turkish another phonological criteria as vowel harmony exists. Vowel within a word shares some

quality. So a word boundary is likely to occur when the quality changes. For example suffix meaning "in" appears as "de" or "da" occurs depending upon previous morphemes. 'In the house' is "evde" but 'in the room' is "odada". Suffix 'my' has four forms 'ev-im' for 'my house', 'at-lm' for 'my horse', 'gozum' for 'my eye' and finally 'topum' for 'my gun'. In English each phonological word has exactly one main stress. Consider the following sentence

"The rest of the books'll have to go there"

There are five main stresses falling on (The rest), (of the books'll), (have to), (go), (here). So there are five phonological words.

But it is not always simple to distinguish the word boundary within a flow of sound. For example the Urdu word the phonological word بس "*bus*" has the same orthography with various meanings.

17	میرااور تھا رہا ہو پروگرام کا شاعری شعرو آج بس بہت بہت شکریہ تو
18	جی ہوں آئی لیکر غزل ایک بس کیا سنائینگی
19	مجھےاگر مہینہ کا رمضان کا رب ہے شکر بس بلکل ٹھیک ہیں آپ سنائے جی
20	ہے ہوئی سجی سے وجہ کی ہی تو ہے جناب بہت بہت شکریہ محفل آپ ہی کی تو ہے جناب بہت بہت شکریہ بس آپ بس
21	صاحبڈکٹر ئیں دعا کی سب آپ ہے شکر کا ﷲالبَس شکر ہے آپ کا جی
22	جیاشون ۔ چھے گیا راجی امے کرو ویبا موکلا بس گھنو گھنو آبھار تمے ای میل موکرا وتے کرو چھو
23	بِسہمت سے بات اس تک کہاں آپ ۔ ہوں ٹھیک بس کیا حال ہیں جناب ۔
24	گولیاںنے آدمی داھاری بندوق نزدیک کے اٹاپ بس مل رہی ہیں کہ اتری شہر فیصل شین میں ایک
25	میںجب پروگرام ایک اور گیا ہوتا شوق سے وہاں بس وہاں سے مختلف انٹرنیشنل ٹور کے چلڈرن فیسٹیولس
26	میںاس ہے جو چیز بھی کوئ ہیں کہتے یہی وہ بس ہے لائف میں اس کی اور میری والدہ کی مجھے
27	کہانے آپ جیسا جی نونیدر ہے رہا چل اچھا ببہت بس ٖکل ٹھیک ہوں آپ سنائیے کیسا چل رہا ہے وہاں ۔
28	رہیہو نہیں کوریلسیز فلم بڑی کوئی تو بفتے اس بس اور کچھ رہنی نی بتانے والی بات بالیووڈ کی ۔
29	چابا جی کو ملنے اسی بس شروع میں بھی آپ نے جو پہلا گِےت لگإءا تھا
30	اچھے بہت ہے بات کےٕا واہ واہ ارے ہوگئے پاک بس مےٕں ہے باق ہوگئے دھوٕئ گٕٓے ہم إءسے کے

Its first meaning is "the feeling of fullness, satisfaction or completeness" while its second meaning is as an English borrowed word when it is used in the sense of "passenger vehicle". In spoken language, the distinction of individual words is even more complex: short words are often run together, and long words are often broken up. More over, phonologically, we find that one phonological form may be represented by more than one orthographic form, e.g کاب ، قاب *qalb* (dog and heart).

We often assume the distinctness of phonological words in the stream of speech, as though they were as distinct as graphological words, forgetting that there are no consistent and unambiguous pauses or boundaries in normal speech, and that syllable boundaries do not necessarily honor the grammatical boundaries between words. We also tend to forget that in highly literate societies an ability to appreciate phonological words is influenced by a separate but related capacity to use and appreciate graphological words. Effectively, the realization of "words" in the graphic and phonemic mediums impinges strongly on each other. Many people can use phonological words without being sure of how they are realized in writing, or vice versa. They can write words without being completely sure about how they are pronounced. Another possible definition of a word is that it will not have more than one stressed syllable. Thus کتاب *kitaab* (book), ہا تھ *haath* (hand), ا نسان *insaan* (human), etc. are unambiguous words. Again, however, we should note that some of the forms designated above as not transmitting meaning (e.g. کا *ka,* کی *ki* میں *mey* (postpositions in Urdu)) do not normally receive stress, except when a particular expressive effect is required. As the majority of world languages are not written, the scientific determination of word boundaries becomes difficult and important.

The statistical aspect:
One can simply count words on a page, and state how many there are. In this case one is simply counting items with white space between them, regardless of what they are, but presumably having taken a decision about how to treat hyphenated words and such solid compounds as طو طا چشم *tauta chachem* (selfish) as opposed to other collocational forms such as کا لا ب ا د ل *kale badel* (black clouds). It becomes clear that even in this simple exercise a consistent set of rules is necessary. If then we wish to count only instances of a particular word on a page then effectively we are counting tokens

of a recurring type, and can say that there are many of instances of a word. We have to be careful, however, and must decide whether we are counting graphological or morphological words.

Are آنا *ana* (to come) آیا *aya,*آ *a,* آئی *aee,* آئے *aae,* and آئیں *aeen*” different words or do we count them as tokens of the type آنا *ana* (come)? If we specify our plan and aim, we can count anything we like, and no one will be confused.

N Concordance

ہے ٹھیک سے،وجہ کی کمی کیونکہ کل فیملی کے جو ہوتے ہیں ان سے لے لیتے ہیں آج	114
جسہوگا ضائع وقت پر بےنےوں بارےک موہوم کل ہوگا جہاں اگر گرائمر پر بحث ہوگی اور وےدانت	115
دودن،پورے ہیں۔ سکتے کر آرام ہم پرسوں اور کل اچھا شب۔بخےر (ارےنا کے ہاتھ کو بوسہ دےتا ہے)	116
دودن،پورے ہیں۔ سکتے کر آرام ہم پرسوں اور کل اچھا شب۔بخےر (ارےنا کے ہاتھ کو بوسہ دےتا ہے)	117
ہیپہلے بہت کہ ہے ہوتا معلوم اےسا تصور کا کل کے او خر میں ہندوستان کے تصوف میں صلح	118
پرکو نصف ےعنی %5کے مزدورں صنعتی کل سربراہ کی حےثےت رکھتا ہے اور ہمارے ملک کی	119
کیمال اور ہے ہوتی پر پےمانہ وسےع پےداوار کل نظام کی بنےادی خصوصیےت ےہ ہے کہ آج	120
پھربے پر اسکولوں داری ذمہ کی تعلیم کہ بج#کل و نما کی بنےاد گھر پر ہی پڑ چکی ہوتی ہے۔ آج	121
بیںکرتے استعمال ہی کا ان میں ٹرینڈ ہیں رہے چل کل راگ ہیں مگر جو سات راگ ہیں جو نارمل آج	122
دوسراکل اب ۔ ہوگی ہی کرنی حاصل جیت یہ تو کل کرو یا مرو والی استیتھی ہے ٹیم کے لئے ۔ انہیں	123
جائےکھیلا بیچ کے بالینڈ اور ذیمبابوے میچ دوسرا کل ۔ انہیں کل تو یہ جیت حاصل کرنی ہی ہوگی ۔ اب	124

Inevitably, however, such problems as کل *kal* (yesterday) and کل *kal* (tomorrow) will emerge to add spice to the undertaking, and further subtle rules will be incorporated until what seemed to be a simple exercise becomes surprisingly complex, especially if we become interested in shades of meaning, in words used in a particular sense.

The LEXICAL ITEM:
The study of lexis in Britian began with the work of J.R. Firth between the late 1930s and 1960s. Firth rejected the division of linguistic analysis into watertight compartments such as semantics, phonology, etc. This is reflected in the use of terms like lexico grammatical and lexical semantics. Firth's idea was taken up and developed by a number of British linguists most notably Halliday, McIntosh, Mitchell, and Sinclair. Both Halliday and Sinclair maintain that a lexical theory is needed to complement grammatical theory. And the grammatical description does not account for all the patterns in a language.

A lexical item may be simply defined as that level of linguistics, which deals with the open sets of formal items of language, for example' پہاڑ *pahaar* (mountain)' پانی *pani*' (water)' خوبصورت *khoob-sooret* (beautiful), کتاب *ketaab* (book), etc.

N Concordance

چرمی میں اس تھی۔ شکل نئی ایک کی کتاب سے استعمال نہیں کرا جاسکتا تھا0 کوڈ ایکس	101
لے ٹے سٹیم بالکل یہ کہ جی ہے یہ اول تو کتاب سے سکریاں سنائی دیں اس کے بارے بتائے	102
راشدجسے ہے ترجمہ اردو کا باب ایک ایک کے کتاب ہے ایک اچھے سائنس داں دھنونت کشور گپت کی	103
لکھ کھار طور واضح میں دریافت کی ہندوستان کتاب محبت کی فضا قائم رہ سکیں۔ پنڈت نہروں نے اپنی	104
حصہ بے شتر کا معشرے بندو ہمارے تھا۔ ربا لکھ کتاب پر ایک سریر #حاصل بحث اور ایک حاصل	105
جاتالے را کاٹ میں شکل کی صفحات طرح کی کتاب تھی۔ اس میں چرمی پارچوں کو دور حاضر کی	106
اورمحنت زیادہ میں اس تو گا چاہے دیکھنا کتاب جانے بغیر مصنف یا عنوان کے ذریعے	107
اورمحنت زیادہ میں اس تو گا چاہے دیکھنا کتاب جانے بغیر مصنف یا عنوان کے ذریعے	108
کے ایش نظرے یے ہے میں اردو میں پنجاب کتاب پر قائم ہوئی ہے۔ حافظ محمود شیرانی نے اپنی	109
سفرنامہ ہے مگر ہے سفرنامہ میں دیکھنے کتاب دیکھتے دنیا کہا سے #کہا ہو چکی تھی۔ ہے	110
کے ترجمہ میں فارسی کا ونش بری کتاب نے سری کرشن کی سوانح حیات سے متعلق	111
کچھ کو سامع ن ہمارے آپ میں بارے کے کتاب ڈاکٹر صاحب یہ بتائے گا کہ اس	112
نام سائنسی کا کپاس یا روئی بندی ۔ عنوان)کا کتاب یز یرگنا کیا) ملہ بلہ کلہ# سریف cotton of	113
جلدوں پانچ اب اور تھی میں جدلوں بے س کتاب آخر میں یا وسط میں لکھی گئی ہوگی اصل	114
بڑی بڑی مطابق کے بے ان کے الابصار مسالک کتاب کی کثرت تھی ۔ شہاب۔الدین العمری مصنف	115

Its form is governed by sound and writing, its content by meaning and use. It may also be defined in terms of denotation and connotation that belong to one of the four classes' noun, verb, adjective, and adverb. LEXICAL ITEMS or VOCABULARY ITEMS is a useful and neutral hold –all term that captures and helps to overcome instabilities in the term WORD, especially when it becomes limited by orthography.

Acoording to Cruse (1986) the criterion for a lexical unit was that it should be 'at least one word'. We must, therefore, examine what this entails. A great deal of scholarly discussion has centered on the linguistic status of the word. It would be useful to draw attention to two fairly general and constant characteristics of words across a wide range of languages. The first is that a word is typically the smallest element that can be moved around without destroying the grammaticality of the sentence (ignoring any semantic effects) and all words are equally mobile in this sense, but with very few exceptions, the smallest mobile units are words. The morphemes constituting a single word have a rigidly a fixed sequential order. The second major characteristic of words is that they are typically the largest units that resist 'interruption' by the insertion of new material between their constituent parts

According to Sinclair "A lexical item may, then, be temporarily defined as a formal item (at least one morpheme long) whose pattern of occurrence can be described in terms of uniquely ordered series of other lexical items occurring in its environment" There are two main categories of lexical words: simple words (single units of meaning) and complex words (which can be broken down into smaller units of meaning). Examples of simple words are: ا سلا م *Islam* (name of religion), کتاب *ketaab* (book), د ر و از ہ *dervazah* (door), etc. Examples of complex words are:

بد بو *Bud-bu*-------- Bud+bu (evil smell)

خو بصو ر ت Khoob-soorat----------khoob+soorat (beautiful)

Contrary to lexical item, in this respect, is grammatical item, which is that level of linguistic form at which operates the closed systems. Prepositions, pronouns, determiners, etc. are called grammatical items. Thus پتنگ *Patang* (kite) is the realization in print of single Urdu lexeme پتنگ *patang*. The nouns گد ھا *ghaddaha* and چا ند *chand* represent at least two lexemes each, گد ھا *ghaddha* (a particular animal and a symbol of stupidity' چا ند*Chand* the moon and a symbol of beauty). Most Urdu dictionaries treat گد ھا *ghaddha* (noun) a single headword with two senses. But they are two lexical items; where as derivatives as چا ند نی *Chandini* (moon light) is considered separate lexeme. In Urdu as in other languages lexemes may be single words گد ھا *ghaddah*, چا ند *chaand*, parts of words نا ک *naak* as in خطر نا ک *Khater-naak* (dangerous) group of words (the compound کا لا رنگ *kaala rang* (black color) and the idiom آ سما ن ٹو ٹنا *asmaan tootna* (to get into hot water) or shortened forms پاکستان *Pakistan* (for Punjab, Sind, Baluchistan, Sarhad, Kashmir) etc. The lexical item may also raise special problems for learners because these can be and usually are realized variously by special morphological words. Thus,

رو نا *Rona* (weep)---------- infinitive / mesculine / feminine,

رو *Ro* (weep)----------- imperative / mesculine / feminine,

رو یا *Roya* (He wept)-------past indefinite / mesculine / singular,

رو ئی *Roee* (She wept)-------past indefinite / feminine / singular,

رو ئے *Roae* (They wept)-----past indefinite/mesculine / plural,

رو ئیں *Roeen* (They wept)----past indefinite / feminine / plural etc.

are distinct morphological words, with various realizations in speech and in writing. Most of us would insist, however, that there is a sense in which they are all the same word and we would call that word رو نا "*rona*". This is recognition of a core of content meaning lying behind the ways in which the word is amended to fit into a syntactic flow. Such amendments, of course, are done by process markers such as affixes or in the case of common irregular verbs by suppletion.

Suppletion can often obscure a relation of this kind; as for example the young child or the foreigner does not see the link between جا نا *Jana* (go) and گیا *gay-a* (went). Suppletion establishes for us, however, the need for the lexical item. Because clearly گیا *gay-a* is not an autonomous word in the same way that جا نا *jana* is. The lexical item is an abstraction of some subtlety, and should not be confused with the lexicographic word, because dictionaries are constrained by tradition to list گیا *gaya* as the past of جا نا *jana* and so on. The lexical item, however, can be taken as the prior condition in any paradigm, so that lexical unit جا نا *jana* lies behind جا ، جا نا ، گئیں ، گئ ، گئے ، گیا، *jana, ja, gaya, gaee, gae* and *gaeen* etc. It becomes a matter of personal preference, however, whether one treats گا چکی ہے *ga chooki hae* (has sung), گا تی ہے *gaati hae* (sings), یا گا *ga-ya* (sang) as single words, or clusters of morphological versions of lexical item There are two sub categories of lexical item.

The proper Item:

The proper item serves to name things, and may be as simple as ا قبال *Iqba*l (name) or as complex as دا ر السلام *Dar-ul-Salaam* (house of peace) as transparent as ڈ بل رو ٹی *double roti* (bread) or as opaque as کر سی *kursi* (chair) as immediately relevant in all its constituent parts as the بجلی گھر *Bijli gher* (Grid station) or as constitutionally irrelevant as سفید خون *sufaid khoon* (decline of love) . These names are clearly lexical, in that they take possessive markers and the like, but differ from other content words in that it makes little sense to count their occurrence and to place them in frequency list although this has indeed been done.

The Multifaceted Item:

Lexical items with the same phonological form while with different orthography قاب کاب *qalb* (dog and heart) and کل *kal* (yesterday and tomorrow) when spoken are called multi-faceted words. These raise special problems when one starts to count words. Commonly word counters choose to count all occurrences of کل *kal* as one word. However, they are two different lexical items and need to be entered in a dictionary separately.

THE GRAMMATICAL ITEM:

The grammatical item belongs to the structures-creating system of the language and serves to link things together . Such items (adverbial, preposition, determiners, pronouns, conjunctions etc) have a high frequency and have their especial semantics, as for example, the particles with their semantics of position and direction, space and time. It is normally assumed that the sets of such words cement the lexical words together that carry most, though not all, of the content of message. The بد *bud* in بد مز ا ج *budmijaz* (bad tempered person) for example, has a qualitative similarity and بد *bud* means bad. This suggests that grammatical words have a kind of honorary membership of the class "word" by virtue of the way in which a language is used, or perhaps more cogently by virtue of the way in which we have chosen to think of it, to write it down and so on. Certainly, all grammatical words potentially have the capacity to behave like affixes, as in the خود *khood* of خو د غر ض *khood garz* (selfish) and the نا *na* of نا سمجھ *na-samegh* (foolish).

N Concordance

منظربی جلدی کو اسباب اور مابیت کی یمظن#بد پیدا کر دیا۔ ہے خیال کیا جانے لگا کہ دماغی	16
بنابنا بوٹے بیل نے انہوں کا گاڑی اس تو لعنت۔ بد بنائیں اور ان کے نیچے لکھا چشم بددور، چشم	17
ہوواضح ۔ ہے سکتی کر تربیت کی مردوں اخلاق بد کرنے کے قابل ہے ، ناسمجھ ، کوتاہ ہیں اور	18
کچھوقت پر کہ ہے#بوجاتا ضدی اور مزاج بد رہتا ہے۔ اور کسی طرح نہیں بہلتا۔ وہ اس حد تک	19
کچھوقت پر کہ ہے#بوجاتا ضدی اور مزاج بد رہتا ہے۔ اور کسی طرح نہیں بہلتا۔ وہ اس حد تک	20
سرخمبارک چہرہ کا آپ کر سن ہے فرمائیں۔ دعا بد تو عرض کیا : حضور دشمنوں کے حق میں	21
داروغتھا نہیں موجود وہاں ایلیڈ سے قسمتی بد داروغہ کو حکم دیا کہ ایلیڈ حاضر کیا جائے	22
امیر۔ ہے کرتا محبت امیر کہ ہو بےرغ ،بد خواہ وہ دیندار یا ہے۔دین ہو ، وہ نیک ہو کہ	23
اوروجوبات مختلف کئی کی بضمی بد نہ جائیں تو کم از کم کسی مضبوط چیز کو پکڑ لیں۔	24
بوعلامتیں اور وجوبات مختلف کئی کی بضمی بد کو پکڑ لیں۔ بد بضمی	25
پیٹکو سر اوقات بعض ۔ ہےسکتی بگڑ اور بضمی بد یا بےبنگم طریقے سے بیٹھنے یا فکر کی وجہ سے	26

On occasions, they can also be lexicalized, as in the آبدیدہ ہونا *aab-dida hona* (to weep). The borderline between, to categorize, is hazy and is made hazier still when we find set phrases such as بے پر کی اڑانا *bay per ki urana* (to talk unwisely). بے *bay* is clearly is a grammatical word. Many postpositional and other phrases behave holistically in this way and underline the peculiar semantic-syntactic fluidity of language, between entirely lexical and entirely the grammatical.

A distinction between lexical and non-lexical item is the difference between content words and the functional words . Major word classes (Noun, verbs, adjective and adverbs) are called lexical while minor word classes (Prepositions, determiners, conjunctions and pronouns) are called non-lexical.

Concluding Remarks:

It is fairly clear from the rival views displayed above that there is no general agreement among linguists on the status of word as a linguistic unit. This difference of approaches seems to originate, primarily, from the fact that the term 'word' is inherently the most ambiguous one. As fluent readers of Urdu, we tend to think only in terms of written words, probably because they are easily identified by the fact that they are separated by blank spaces. But words may be defined in different ways from different perspectives, with each perspective picking out a somewhat different object from the others. These ambiguities relate to various levels of language, as well as to the problem of counting words as types. In practice, linguists apply a mixture of all these methods to determine the word boundaries of any given sentence. Even with the careful application of these methods, the exact definition of a word is often still elusive and in all the scholarly attempts to define word, the general features that are identified and included in the definition remained the same, but the weight given to each of these features and the componential organization of these features in the definition of word differed. It is for these reasons that the artificial term "lexical item" may be more useful in the discussion of lexical issues and purposes of L2 vocabulary teaching and learning. It is a general concept that underlies all this diversity. But according to Halliday and Yallop (2007) it is the notion of content word that corresponds to our lexical item.

ACKNOWLEDGEMENT

Abrar Hussain Qureshi
Assistant Professor
Department of English
Govt. College Chichawatni
PhD Scholar
Department of English
BZU Multan
Pakistan
abrarqureshi74f@hotmail.com

Dr. Muhammad Akram
PhD (IUB)
Pakistan
muhammadakram80@yahoo.com

References:

Balhouq, S.A. (1976). *The Place of Lexis in Foreign Language Learning* Unpublished M.A. Thesis submitted to University of Scheffield.

Branford, W. (1967) The Elements of English : An Introduction To the Principles of Study of Language, Routledge and Kegan Paul, London.

Cruse, D.A. (1986) *Lexical Semantics* Oxford University Press.

Halliday, M. A. K. (1961) Categories of the Theory of Grammar. Word 17: 241–292.

Haspelmath, M. (2002). *Understanding morphology*. London: Arnold (co-published by Oxford University Press).

Hay, J. (2000). *Causes and Consequences of Word Structures*, A PhD thesis, Northwestern University

Katamba, F.(2005) *English words*. Routledge Taylor and Francis group London and New York.

Kramsky, J. 1969 The Word as a Linguistic Unit. Mouton. The Hague and Paris.

Lyons, J.(1968) Introduction to Theoretical Linguistics, Cambridge University Press, Cambridge.

Sailer, M. (2003). *Combinatorial Semantics and Idiomatic Expressions*
　　　　in Head-Driven Phrase Structure Grammar. Ph.D thesis, Unpublished

Schmmit, N. (2000) Vocabulary in Language Teaching. Cambridge University Press