

# إحصاءات اللغة العربية في الويب بين التفصيح والتوليد

<https://www.youtube.com/watch?v=IBnCNAN8fT4&t=894s>

د. سلطان ناصر المجيول

جامعة الملك سعود

# التطور اللغوي في العربية القديمة

- الفصح، والأفصح، والضعيف، والمنكر، والمتروك [المزهر في علوم اللغة للسيوطي]
- اختلف بعض علماء العربية قديمًا في تحديد مفهوم الفصح (الاتجاهات):
  1. الفصاحة في كلام ما نُسب على العرب الفصحاء (معيار المكان والزمان): (معجم الصحاح للجوهري)
  2. الفصاحة في كثرة الاستعمال (ثعلب في كتابه: الفصح)
  3. الفصاحة فيما بني على القياس النحوي المطرد (ابن جني في كتابه الخصائص)

هل يُمكن أن يُضاف إلى اللغة ما ليس فيها قياسًا على ما فيها؟



# التفصيح والتوليد في العربية

## التفصيح

الأصوات (التمائل، والإبدال، والتباين، والقلب المكاني)

### الأبنية الصرفية

التداخل بين الأوزان: فَعَلَ وَأَفْعَلَ،

صوغ الجموع على غير القياس: جمع صيغة "فَعَلَ" على "أَفْعَلَ" ما لم تكن فاء الفعل واوًا ولامها مماثلة لعينها: نَجْم:

أَنْجُم | ولكن جُمعت لِجَام على أَلْجُم

التصغير: فُعَيْلٌ. ظهور صيغة "فُعَيْلٌ" بين عوامٍ صقلية: كُتَيْبٌ

زيادة حرف في البنية الأصلية مثل (نواة: نواية) و(حصاة: حصاية)

### التركيب

الخلط بين الأفعال المتعدية واللازمة: (تَأَكَّدَ الأمر) بدل (تَأَكَّدَ من الأمر) | "حرمه من اللعبة" بدل "حرمه اللعبة"

استعمال قط للمستقبل: ما فعلت هذا قطُّ

اشتقاق مصدر على وزن (فِعَالَة): لِيَاقَة، ثم تحولها إلى لِيَاقَة

## التوليد

ظاهرة قديمة وحديثة

- المجاز (الخبث في العروض)
- الاشتقاق (التكميم)
- الترجمة الحرفية
- الاقتراض

# المعاجم العربية القديمة والحديثة

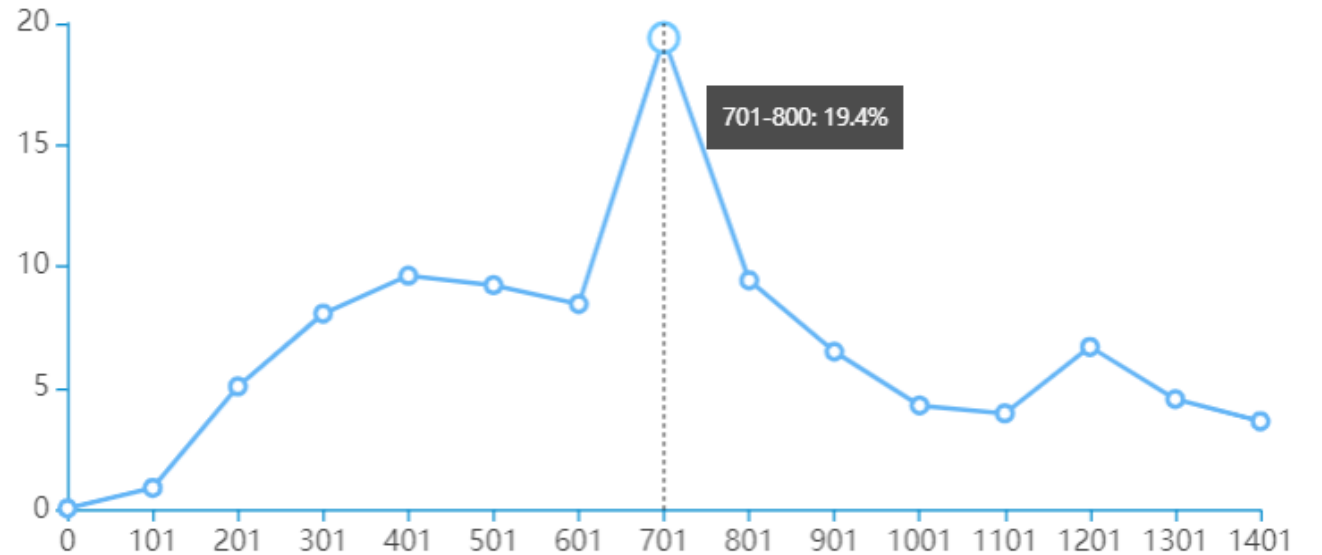
المعاجم المعاصرة	المعاجم الوسطى	المعاجم الأولى	
الرائد/ المعجم الوسيط/ المنجد/ العربية المعاصرة	معجم ما استعجم/ أساس البلاغة/ العباب الزاخر/ القاموس المحيط/ مختار الصحاح/ لسان العرب/ تاج العروس	العين/ البارع في اللغة/ الصحاح في اللغة/ تاج اللغة والصحاح/ المحكم والمحيط/ المخصص/ تهذيب اللغة/ جمهرة اللغة/ الألفاظ/ مجمل اللغة/ مقاييس اللغة	
2,340,322	3,825,200	6,241,678	عدد الكلمات بالتكرارات
154,957	1,380,723		عدد الكلمات ذات المحتوى
79,854	402,964		عدد الكلمات ذات المحتوى غير المتقاطعة
الكلمات الأساس: 8,285 الكلمات الأساس المنفردة في المعاصرة: 677	الكلمات الأساس: 11,784 الكلمات الأساس المنفردة في الأولى والوسطى: 4,177		إحصاءات الكلمات الأساس الرقمية
	الكلمة الأساس المتقاطعة (المشتركة): 7,607		إحصاءات الكلمات الأساس المشتركة

# (الفصيح) مدونة النصوص الإسلامية المفتوحة OpenTI

عدد الكلمات	السنة الهجرية
1,008,000	0-100
14,000,000	101-200
78,000,000	201-300
124,000,000	301-400
148,000,000	401-500
142,000,000	501-600
130,000,000	601-700
298,000,000	701-800
145,000,000	801-900
100,000,000	901-1000
66,000,000	1001-1100
61,000,000	1101-1200
103,000,000	1201-1300
70,000,000	1301-1400
56,000,000	1401-
1,537,000,000	المجموع

## نسبة الكلمات بين السنوات

النسبة المئوية من مليار و 537 مليون



<https://alraqmiyyat.github.io/OpenTI/> (745,390,461) الكلمات الحقيقية:

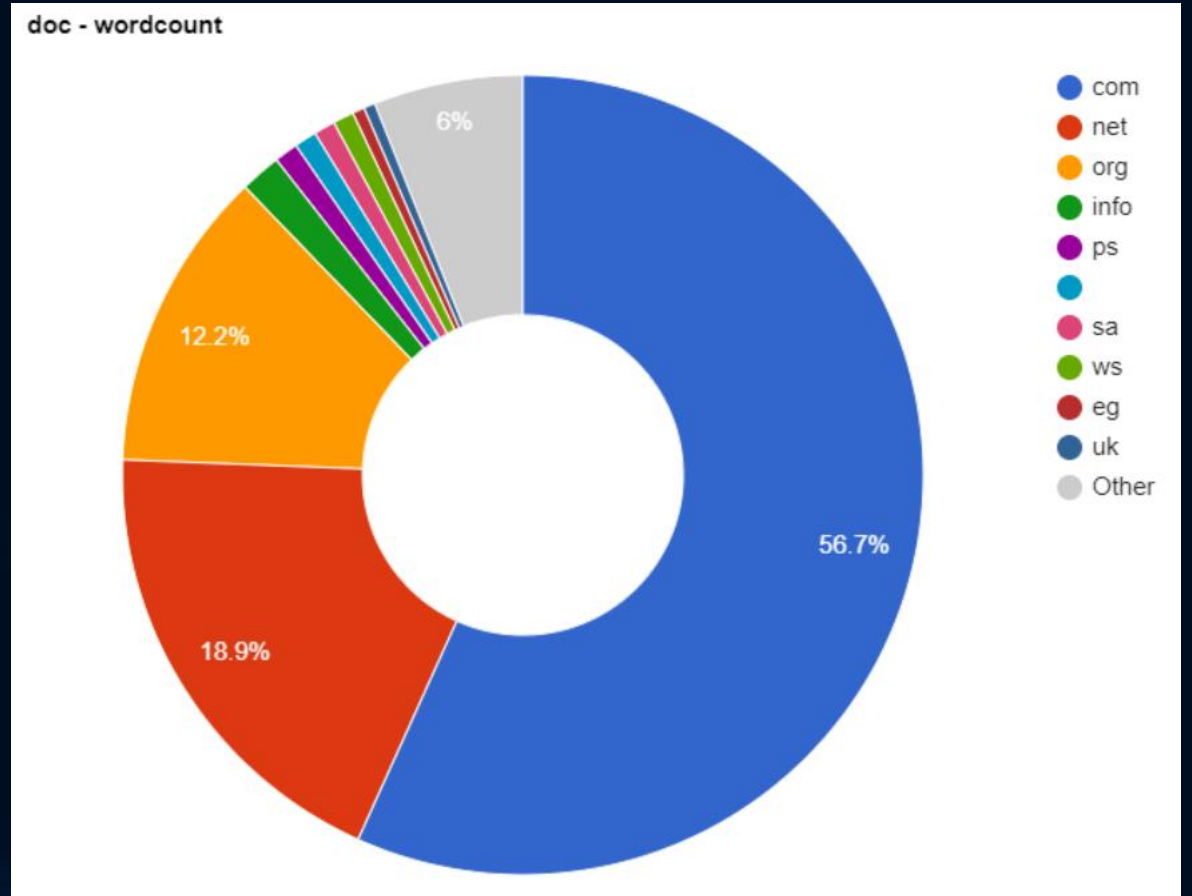
الكلمات من دون تكرار: 1,788,937

# التوليد (المكتوب والمقول) بين المصنوع (أنموذج) والتلقائي (عامية)

تتعدد المداخل بتعدد الصيغ الإملائية على مستويات التوليد: مثل (الأنجلوساكسونيون)

العدد بالتكرارات	الفئات
8,322,097,229	الرموز
7,475,624,779	الكلمات
177,011,603	الجملة
104,623,566	الفقرات
11,543,931	الملفات

العدد بدون تكرار	الفئات
410K	الكلمات المشوّهة
590K	الكلمات غير المشوّهة
120K	التضاعف والمتعدد إملائيًا

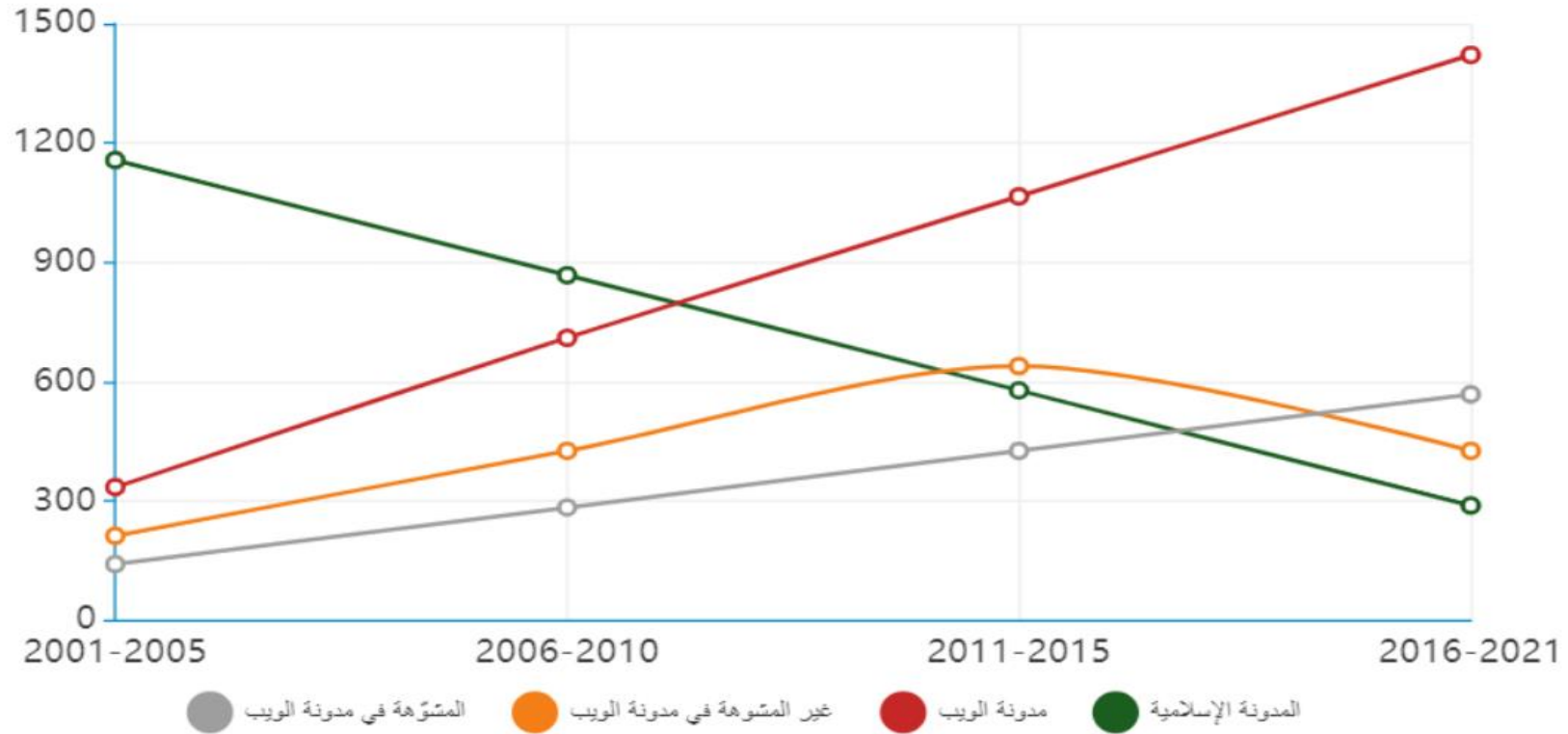


## الوحدات المعجمية بين المدونة الإسلامية ومدونة الويب العربية

التكرار المعيّر في كل مليون	الوحدات المعجمية	المصدر
2892.907	1,788,937	الفصح في الويب
3553.548	2,197,469	التوليد في الويب
2132.128	1,318,481	التوليد غير المشوهة
1421.418	878,987	التوليد المشوهة

## الكلمات بين المدونة الإسلامية ومدونة الويب العربية: القائمة الممعيرة normalized

التكرار الممعيير لكل 10,000 كلمة



## الكلمات بين المدونة الإسلامية ومدونة الويب العربية (المشوّه وغير المشوّه): أحجام التأثير

حجم التأثير بين المدونة الإسلامية ومدونة الويب العربية (بين مجموعتين): تقدير التكرارات النسبية بوصفها متوسطات مع الانحرافات المعيارية وحجم العينة

الفترة 2011-2020	الفترة 2001-2010
اختبار التمهيد (1000)	اختبار التمهيد (1000)
القيمة الاحتمالية 0.750	القيمة الاحتمالية 0.772
زيادة نسبة المتوسط: 4.017	زيادة نسبة المتوسط: 5.766
مجال الثقة 95 بالمئة [54.343, 62.377]	مجال الثقة 95 بالمئة [-97.61, 109.143]

كوهن-د على أنه الفرق بين المتوسطين المُعبر عنهما بوحدات الانحراف المعياري

تفسير  $d$ : أكبر من 0.3 يكون التأثير صغيرًا. أكبر من 0.5 يكون التأثير متوسطًا. أكبر من 0.8 يكون التأثير كبيرًا

# الفصح والتفصيح والتوليد بنوعيه (المشوّه وغير المشوّه): سياسة وتخطيط التوازن

➤ سياسة التوازن والتخطيط له

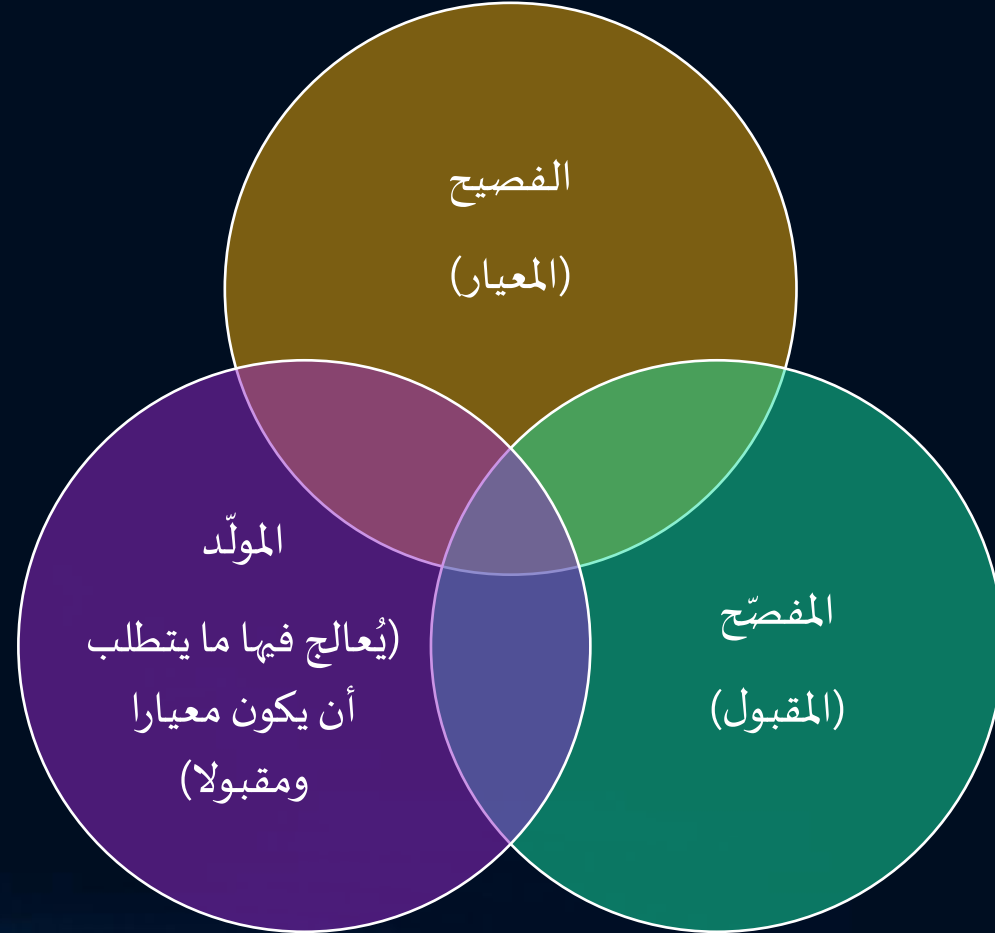
1. لعبة المحصلة الصفرية

2. توازن ناش

3. توازن ناش الفرعي المثالي

➤ كلمات يومية جديدة تبلغ عشرات الآلاف في الاستعمال حسب مدونة الويب: / يهدس/ يهرد/ يتكأء عليه (يشق عليه)/ منهضة/ الثرثور/ التيساع/ تخريط/ إلخ.

➤ افتقار اللغة العربية إلى تسمية الموجودات والمواد الثقافية material-culture الجديدة على الرغم من وجودها في ترجمات الكتيّبات الصناعية



نختم بـ

## التوصيات للسياسات اللغوية والتخطيط للغة العربية في الويب

### الجهات المعنية بالتوصيات

- مجمع الملك سلمان العالمي للغة العربي
- مراكز اللغة العربية في العالم العربي
- الجامعات الإلكترونية
- المؤسسات المعنية بحوسبة اللغة العربية (المعجم، والدلالة، والخطاب)

### أهم التوصيات

- التفريق بين التفصيح الممكن والتفصيح غير الممكن
- تفصيح ما يُمكن تفصيحه من الكلمات المولدة غير المشوهة والبحث عن سبل توحيد الرسم الإملائي
- تفصيح التراكيب في المقولات التركيبية النحوية في لغة الصحافة
- سدّ الغربة أو الفجوة المعجمية في الويب العربي (مشاريع المعاجم والأنطولوجيا): هناك كلمات لا تظهر في معجم المعاني ولا تظهر في المعجم المعاصر لشركة صخر.